

<http://lexikos.journals.ac.za>

Lexikos 26

<http://lexikos.journals.ac.za>

Lexikos 26

Redakteur

Editor

Danie Prinsloo

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 26:2016

AFRILEX SERIES 26:2016



BURO VAN DIE WAT

STELLENBOSCH

<http://lexikos.journals.ac.za>

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

Kopiereg © 2016 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2016

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 978-0-9946528-2-9
ISSN 1684-4904

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

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

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

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

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

Span van Roterende Redakteurs / Team of Rotating Editors

Dr. H.S. Ndinga-Koumba-Binza (RSA en Gaboen/RSA and Gabon)
Prof. D.J. Prinsloo (RSA)
Prof. Elsabé Taljard (RSA)

Adviesraad / Advisory Board

Prof. A. Adamska-Salaciak (Pole/Poland)
Prof. H. Béjoint (Frankryk/France)
Prof. H. Chimhundu (Zimbabwe)
Prof. F. Dolezal (VSA/USA)
Prof. R.H. Gouws (RSA)
Prof. R.R.K. Hartmann (Groot-Brittanje/Great Britain)
Prof. M.H. Heliel (Egipte/Egypt)
Prof. W. Martin (België en Nederland/Belgium and The Netherlands)
Prof. I.A. Mel'čuk (Kanada/Canada)
Prof. A.M.F.J. Moerdijk (Nederland/The Netherlands)
Dr. J. Tent (Australië/Australia)
Prof. J. Van Keymeulen (België/Belgium)
Prof. P.G.J. van Sterkenburg (Nederland/The Netherlands)
Prof. L.S. Vikør (Noorweë/Norway)
Prof. H.E. Wiegand (Duitsland/Germany)

Redaksiekomitee / Editorial Committee

Dr. M.M. Bagwasi (Botswana)
Prof. H.L. Beyer (Namibië/Namibia)
Prof. W.A.M. Carstens (RSA)
Prof. E. Chabata (Zimbabwe)
Prof. C.J. Conradie (RSA)
Prof. A.E. Feinauer (RSA)
Prof. R. Finlayson (RSA)
Dr. S. Hadebe (Zimbabwe)
Prof. I.M. Kosch (RSA)
Dr. P.A. Louw (RSA)
Mnr. K.J. Mashamaite (RSA)
Prof. P.A. Mavoungou (Gaboen/Gabon)
Dr. V.M. Mojela (RSA)
Mnr. M.C. Mphahlele (RSA)
Dr. H.S. Ndinga-Koumba-Binza (RSA en Gaboen/RSA and Gabon)
Dr. D. Nkomo (RSA en Zimbabwe/RSA and Zimbabwe)
Prof. T.J. Otlogetswe (Botswana)
Prof. A.N. Otto (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	ix
Foreword	x
<i>Danie Prinsloo</i>	
'n Woord van AFRILEX	xi
A Few Words from AFRILEX	xii
<i>V.M. Mojela</i>	
Redaksionele doelstellings	xiii
Editorial Objectives	xiv

Artikels / Articles

The Incorporation of Specialised Data in Lexicographical Meaning Explanations: A Discussion Based on Sports and Fitness Terms	1
<i>Heidi Agerbo</i>	
Etc. The Long-Lasting Defining Device: Unravelling the Mystery	36
<i>Safi Eldeen Alzi'abi</i>	
On the Metalexigraphic Genre of Dictionary Reviews, with Specific Reference to <i>LexicoNordica</i> and <i>Lexikos</i>	60
<i>Henning Bergenholtz and Rufus H. Gouws</i>	
Adapting a Historical Dictionary for the Modern Online User: The Case of the <i>Dictionary of South African English on Historical Principles</i> 's Presentation and Navigation Features	82
<i>André du Plessis and Tim van Niekerk</i>	
Op pad na 'n omvattende woordeboekkultuur in die digitale era	103
<i>Rufus H. Gouws</i>	

The Application of the Prototype Theory in Lexicographic Practice: A Proposal of a Model for Lexicographic Treatment of Polysemy <i>Ana Halas</i>	124
Lemmatisation of Fixed Expressions: The Case of Proverbs in Northern Sotho <i>I.M. Kosch</i>	145
What French for Gabonese French Lexicography? <i>Blanche Nyangone Assam, Hugues Steve Ndinga-Koumba-Binza and Virginie Ompoussa</i>	162
Die rol van die US se Departement Afrikaans en Nederlands in die ontwikkeling van die Afrikaanse leksikografie onder Nederlandse invloed: Verklarende standaardwoordeboeke <i>Gerda Odendaal</i>	193
A Critical Analysis of Multilingual Dictionaries <i>D.J. Prinsloo</i>	220
Probleme in Bezug auf die slowenische lexikographische Terminologie <i>Simona Štavbar</i>	241
Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-UVA <i>Sven Tarp and Pedro A. Fuertes-Olivera</i>	273
Illustrative Examples in a Bilingual Decoding Dictionary: An (Un)necessary Component? <i>Alenka Vrbinc and Marjeta Vrbinc</i>	296
Über die Konstruktion einer hierarchischen Textverbundgesamtstruktur. Ein Beitrag zur Theorie der Wörterbuchform <i>Herbert Ernst Wiegand</i>	311

How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions <i>Sascha Wolfer and Carolin Müller-Spitzer</i>	347
Comparing Dictionary-induced Vocabulary Learning and Inferencing in the Context of Reading <i>Di Zou</i>	372
Leksikonotas / Lexiconotes	
Considering the Merits, and Some Demerits, of the <i>Longman English–Japanese Dictionary</i> <i>Defa Ren</i>	391
Learner Features in a New Corpus-based Swahili Dictionary <i>Beata Wójtowicz</i>	402
The Corpora of China English: Implications for an EFL Dictionary for Chinese Learners of English <i>Lixin Xia, Yun Xia, Yihua Zhang and Hilary Nesi</i>	416
Resensies / Reviews	
Peter Gilliver. <i>The Making of the Oxford English Dictionary</i> <i>J.C.M.D. du Plessis</i>	436
Oskar Reichmann. <i>Historische Lexikographie. Ideen, Verwirk- lichungen, Reflexionen an Beispielen des Deutschen, Niederlän- dischen und Englischen</i> <i>Maria Smit</i>	446
Publikasieaankondigings / Publication Announcements	456
Voorskrifte aan Skrywers	457
Instructions to Authors	458

Voorwoord

Die 2016-uitgawe van *Lexikos* verskyn in die jaar waarin AFRILEX sy 21ste verjaardag gevier het tydens die 21ste Internasionale Konferensie van AFRILEX in Tzaneen, Suid-Afrika. Ons is dankbaar vir die entoesiastiese akademiese bydraes tot ons vereniging en sy tydskrif *Lexikos* deur ons lede en alle belangstellendes in die leksikografie regoor die wêreld asook teenoor die Buro van die Woordeboek van die Afrikaanse Taal vir deurlopende leiding en ondersteuning gedurende hierdie hele periode. *Lexikos* 26:2016 is geredigeer deur D.J. Prinsloo ondersteun deur Steve Ndinga-Koumba-Binza en Elsabé Taljard. My opregte dank gaan aan Steve en Elsabé vir hulle bydrae. *Lexikos* 27:2017 sal deur Steve Ndinga-Koumba-Binza geredigeer word.

Ek wil ook my hartlike dank betuig teenoor die ongeveer 45 beoordelaars van regoor die wêreld wat aan vanjaar se keuringsproses deelgeneem het deur dit werklik met mening aan te pak, nie slegs deur die skryf van kritiese en insiggewende verslae nie, maar ook deur outeurs met uitvoerige kommentaar van hulp te wees. Hierdie kommentaar het aansienlik bygedra tot die verbetering van die kwaliteit van die artikels en die handhawing van die akademiese gehalte van hierdie nommer van *Lexikos* as 'n tydskrif van hoë akademiese status.

Wat die tegniese en administratiewe aspekte van die publikasie van die tydskrif betref, gaan opregte dank aan Tanja Hartevelde van die Buro van die WAT, wat deur Hermien van der Westhuizen ondersteun is.

D.J. Prinsloo
Redakteur

Foreword

The 2016 edition of *Lexikos* is published in the year when AFRILEX celebrated its 21st birthday at the 21st International Conference of AFRILEX in Tzaneen, South Africa. We are grateful for the enthusiastic academic contributions to our association and its journal *Lexikos* from our members and everyone interested in lexicography from across the world as well as to the Bureau of the Woordeboek van die Afrikaanse Taal for its continuous guidance and support throughout this entire period. *Lexikos* 26:2016 has been edited by D.J. Prinsloo supported by Steve Ndinga-Koumba-Binza and Elsabé Taljard. My sincere thanks goes to Steve and Elsabé for their contribution. *Lexikos* 27:2017 will be edited by Steve Ndinga-Koumba-Binza.

I also wish to sincerely thank the approximately 45 adjudicators from across the world who participated in this year's review process for really putting their minds to it, not only in writing critical and informative reports, but also for assisting authors through extensive comments. These comments substantially contributed to improving the quality of the articles and maintaining the academic quality of this volume of *Lexikos* as a journal of high academic standing.

Regarding the technical and administrative aspects of the publication of the journal, sincere appreciation goes to Tanja Harteveld of the Bureau of the WAT, who was supported by Hermien van der Westhuizen.

D.J. Prinsloo
Editor

'n Woord van AFRILEX

AFRILEX wil prof. D.J. Prinsloo, die Adjunk-President van AFRILEX, bedank dat hy die belangrike rol van redakteur vir vanjaar se nommer, naamlik *Lexikos* 26, vervul het. Prof. Prinsloo is die Hoof van die Department Afrikatale aan die Universiteit van Pretoria, en 'n uitstekende navorser op die gebied van die linguistiek en die leksikografie, veral betreffende navorsing oor inheemse Afrikatale. Namens AFRILEX wil ons prof. Prinsloo bedank vir 'n knap stuk werk. Ons wil ook vir dr. H.S. (Steve) Ndinga-Koumba-Binza bedank vir en hom gelukwens met sy bereidwilligheid om by die redaksionele span aan te sluit om te help met die volgende nommer van *Lexikos*. Hy is 'n jong navorser en 'n bekende leksikografiese vakkundige. AFRILEX is bly dat 'n persoon van sy kaliber deel vorm van die redaksionele span as redakteur van *Lexikos*.

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 wat betref die kwaliteit van die gepubliseerde artikels. Binne AFRILEX is ons bly om te sien hoe die tydskrif, wat beide kwantiteit en kwaliteit betref, van krag tot krag gaan met die verskyning van elke nommer sodat dit gunstig kan kompeteer met soortgelyke tydskrifte elders in die wêreld. Al hierdie suksesse kan alleenlik toegeskryf word aan die werk van die redakteurs en die ewekniebeoordelaars wat verseker dat die *Lexikos*-manuskripte op 'n vlak geskik vir insluiting in die tydskrif is. Ons wil ook die outeurs van die artikels bedank, veral AFRILEX-lede, wat voortgaan om *Lexikos* van kwaliteitsartikels te voorsien wat jaar na jaar die status van die tydskrif verhoog en handhaaf.

Weer eens 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 baie wat nie die konferensies kan bywoon nie.

Maropeng Victor Mojela
President: AFRILEX

A Few Words from AFRILEX

AFRILEX wants to thank Prof. D.J. Prinsloo, the AFRILEX Deputy President, for performing the important role of being the editor of this year's issue, i.e. *Lexikos* 26. Prof. Prinsloo is Head of the Department of African Languages at the University of Pretoria, and an outstanding scholar in the field of linguistics and lexicography, especially regarding research in African indigenous languages. On behalf of AFRILEX, we want to thank Prof. Prinsloo for a job well done. We also want to thank and congratulate Dr H.S. (Steve) Ndinga-Koumba-Binza for accepting to join the *Lexikos* editorial team to assist in the next issue of *Lexikos*. As a young researcher and well-known lexicographic scholar, AFRILEX is pleased to have a person of his calibre to form part of the editorial team as editor of *Lexikos*.

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

Once more we want to encourage all AFRILEX members 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 many who could not attend the conference.

Maropeng Victor Mojela
President: AFRILEX

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

The Incorporation of Specialised Data in Lexicographical Meaning Explanations: A Discussion Based on Sports and Fitness Terms*

Heidi Agerbo, *Centre for Lexicography, Department of Business Communication, School of Business and Social Sciences, Aarhus University, Aarhus, Denmark (heap@bcom.au.dk)*

Abstract: There is a problematic treatment of field specific words in both general language dictionaries as well as specialised dictionaries, which in this paper is exemplified using the sports and fitness subject fields as the point of departure and focusing on meaning explanations. Particularly the treatment in general dictionaries in many cases results in inadequate solutions that cannot sufficiently accommodate the dictionary users' needs if we consider these from the viewpoint of the function theory, according to which a dictionary is designed to help a specific user type with a punctual information need that has occurred in a specific type of extra-lexicographical user situation. In this study, a total of 44 general and specialised information tools in English and Danish were analysed. The first part of the investigation involved an analysis of the introductory matter to determine the intended user situation(s) and user(s) of each tool. In the second part, a selected number of sports and/or fitness related lemmata in the dictionaries were analysed. These test words and their meaning explanations were in each dictionary compared to the dictionary's mentioned user situation(s) and user(s). Based on this comparison, a number of identified problems with the meaning explanations are discussed, and a set of principles for best practice when producing meaning explanations of specialised terms is suggested. The study revealed that quite a number of these information tools either do not clearly specify their functions or, to some extent, cannot live up to their intended functions. Some additional interesting findings were also made as some of the dictionaries either explicitly or implicitly are designed with operative and interpretative functions, which are functions that have yet not been fully incorporated into lexicographical theory. The overall findings from the study have led to a suggested function-based definition of meaning explanations to be applied in all types of information tools.

Keywords: MEANING EXPLANATION, FUNCTION THEORY, INTRODUCTORY MATTER, USER TYPE, USER SITUATION, GENERAL DICTIONARY, SPECIALISED DICTIONARY, SPORT, FITNESS

* This article was presented as a paper at the Twenty-first Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Xitsonga and Sesotho sa Leboa National Lexicography Units, Tzaneen, South Africa, 4–6 July 2016.

Opsomming: Die opname van gespesialiseerde inligting in leksikografiese betekenisomskrywings. 'n Bespreking gebaseer op sportsoorte en fiksheids-terme.

Vakspesifieke woorde se hantering is problematies in sowel algemene as gespesialiseerde woordeboeke. In hierdie artikel word dit toegelig met die vakgebiede sport en fiksheid as vertrekpunt en word daar gefokus op betekenisomskrywing. In baie gevalle is dit veral die hantering in algemene woordeboeke wat uitloop op onvoldoende oplossings wat nie voldoende voorsiening maak vir die gebruiker se behoeftes as ons dit beskou vanuit die hoek van die funksieteorie nie, waarvolgens 'n woordeboek ontwerp word om 'n spesifieke soort gebruiker te help met 'n presiese inligtingsbehoefte wat spruit uit 'n spesifieke soort buiteleksikografiese gebruikersituasie. Tydens hierdie studie is 'n totaal van 44 algemene en gespesialiseerde inligtingshulpmiddels in Engels en Deens ontleed. Die eerste deel van die ondersoek het ingehou dat 'n ontleding van die inleidende gedeelte gedoen is om die bedoelde gebruikersituasie(s) en gebruikers van elke hulpmiddel te bepaal. In die tweede gedeelte is 'n uitgesoekte aantal lemmas wat met sportsoorte en/of fiksheid te doen het in die woordeboeke ontleed. Hierdie woorde wat deel vorm van die toets en hulle betekenisomskrywings in elke woordeboek is vergelyk met die gebruikersituasie(s) en gebruiker(s) wat in die woordeboek genoem word. Op grond van hierdie vergelyking word 'n aantal probleme bespreek wat met die betekenisomskrywings geïdentifiseer is en 'n stel beginsels vir die beste praktyk word voorgestel vir die saamstel van betekenisomskrywings in die geval van gespesialiseerde terme. Die studie het getoon dat 'n groot aantal van hierdie inligtingshulpmiddels òf nie hulle funksies duidelik spesifiseer nie, òf in sekere mate nie heeltemal aan hulle bedoelde funksies voldoen nie. Bykomende interessante gevolgtrekkings is ook gemaak, aangesien party woordeboeke eksplisiet of implisiet ontwerp is met operatiewe en verklarende funksies, funksies wat nog nie volledig opgeneem is in die leksikografiese teorie nie. Die oorhoofse bevindings uit hierdie studie het gelei tot 'n voorgestelde funksie-gebaseerde definisie van betekenisomskrywings wat toegepas kan word op alle tipes inligtingshulpmiddels.

Sleutelwoorde: BETEKENISOMSKRYWING, FUNKSIETEOORIE, INLEIDENDE GEDEELTE, SOORT GEBRUIKER, GEBRUIKERSITUASIE, ALGEMENE WOORDEBOEK, GESPELIALISEERDE WOORDEBOEK, SPORT, FIKSHEID

1. Introduction

The soul of a dictionary, however, is the quality of its definitions
Morris (1985: 6)

As this quote shows, many lexicographers and dictionary users consider the definition (or meaning explanation, as will be the term applied throughout this paper) the most important part of a dictionary. The focus of the study presented in this paper was to determine the usefulness of the meaning explanations of sports and fitness related terms provided in a number of Danish and English general and specialised dictionaries. The term *usefulness* is here defined on the basis of the function, i.e. whether the explanations satisfy the intended user type who has a specific information need in the intended extra-lexicographical user situation as expressed in the function theory; see e.g. Bergenholtz and Tarp (2003) and Tarp (2008a). Many lexicographers today will agree

with the statement that a dictionary should be produced with a specific user type as the point of departure, cf. Diab (1990: 21-51) and Nesi (2013), who both comment on the development of the user-oriented focus in lexicography. However, proponents of the function theory argue that the genuine purpose of a dictionary first and foremost depends on its function, thus, all data types should be incorporated into and presented in an information tool based on their relevance according to both the intended user situation and intended user (Nielsen and Almind 2011: 154). In the study presented in this paper, the areas of sport and fitness are used to form a small sample for investigating the usefulness of meaning explanations of terms in both general and specialised information tools.

2. Analysis of information tools

The types of information tools chosen for the current investigation can be divided into three types: general dictionaries, multi-field dictionaries (sports dictionaries) and single-field dictionaries (e.g. tennis and martial arts dictionaries). In addition, dictionaries related to sports and fitness, e.g. a dictionary on diet and exercise and a dictionary on sports injuries, were included. Some of the general dictionaries were chosen based on their online availability, and others were randomly chosen from their physical availability from the office shelves at the Centre for Lexicography — these amounted to 17 relevant dictionaries in total¹. A search was made for multi-field and single-field dictionaries at the Danish State and University Library and bibliotek.dk, a website where you search all Danish public libraries simultaneously. This only resulted in 9 additional dictionaries. A search with Google gave access to another 2 online multi-field dictionaries and 2 electronic (static) multi-field dictionaries. As it was hoped to find more information tools on sports and fitness for the analysis, a search was made at amazon.com and amazon.co.uk, which resulted in 14 additional dictionaries. Thus, a total of 44 different types of information tools (i.e. tools that contain collections of structured data and are designed in a certain way to help users solve information-related problems) were selected for the analysis, e.g. dictionaries, lexicons and glossaries².

The test words were randomly selected from English and Danish newspaper articles in sport, a number of which were from the early 1970s and the rest from after 2000, a selection made in order to match words from old newspaper articles with old dictionaries and words from recent newspaper articles with new dictionaries. 10-15 test words were chosen from each selected sports branch, which were selected based on the author's personal experience with and knowledge about these specific sports. The early articles included: boxing, (association) football, swimming and tennis, respectively; and the later articles included: boxing, (association) football, handball, tennis, running and fitness, respectively. The reason for the difference in the selection of sports branches in the old and new newspaper articles is the changing popularity of the different

sports. Running started to become popular in the 1970s, but in the last 5-10 years, it has become more popular than ever, which means that newspapers and magazines often write about running, and modern dictionaries will (or should) contain more running terms. Fitness became very popular around the end of the 20th century, and today the fitness industry is bigger than ever, which has entailed an increased interest for this subject in newspapers and other media. Handball did not appear in any of the searches made in the English newspapers from the 1970s, and therefore articles and tests words in swimming were selected instead. In a few of the dictionaries for which these search terms were not relevant, an analysis was made of the meaning explanations on every 30th page in the dictionary. As is clear from these descriptions, the study is not statistically valid, but instead a qualitative and exemplary study, i.e. it can highlight tendencies. The purpose of the study was not to count which test words occurred in which dictionary, but to analyse how the lemmata were described in different information tools on the same or related topics and compare these descriptions to the stated dictionary functions.

2.1 The introductory matter

The purpose of the analysis of the dictionary introductions — also called prefaces, introductory remarks, introductory pages, introductory guides, initial information for users, and preliminaries — was to identify the intended users and user situations since this kind of information is usually provided in this part of a dictionary, cf. Bergenholtz and Tarp (1994: 174-175). Nielsen (1990: 52) describes a simple macrostructure as one typically containing only a preface and the (alphabetical) arrangement of the lemmata, thus, according to this statement, all dictionaries will most likely contain an introduction.

The different types of users typically identified in specialised lexicography are laymen, semi-experts and experts, who can be characterised according to their field-specific knowledge and also according to their language skills, cf. Bergenholtz and Tarp (1994: 17-18). Of course, when analysing the selected learner dictionaries, it will also be necessary to incorporate this type, the learner, into the user typology, and here it is possible to distinguish between three types: beginner, intermediate and advanced. In addition to this, an extra category should be added: the general user (which is, admittedly, a very broad category). It is problematic to call people who use general dictionaries laymen, as this term is seen in relation to a certain specialised field: the same person can be an expert in chemistry and a layman in linguistics, but you do not consider him an expert until he is associated with chemistry, and you do not call him a layman until he is associated with linguistics. People who use general dictionaries are not evaluated according to a specific specialised field or sub-field, and therefore it would be better to call them general users (see also Béjoint 2015).

All of the general dictionaries have a communicative function, more spe-

cifically, they are produced for reception and production. The investigated specialised dictionaries are either only for reception — none of them can be used for production — only for cognition, or both of these two functions. Especially much of the back matter in sports dictionaries supports a cognitive function. However, as will be discussed in Section 2.4, some of the dictionaries also incorporated data both in the central lists and in the outer texts that support what the function theory calls the operative and interpretative functions.

2.1.1 The user type

Table 1: Overview of the data on user types and user situations

	Is described in the introduction/on the website	Is only described on the book cover	No description	Can be inferred from the dictionary title
User	5, 7, (8), 10, 11, 16, (18), 21, 23, 26, 27, 28, 29, 30, 31, 32, 33, 42, 44	22, 24, 34, 35, 40, 43	1, 2, 3, 4, 6, 9, 12, 13, 14, 15, 17, 19, 20, 25, 36, 37, 38, 39, 41	1, 3, 6, 8
Situation	(4), 5, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 25, 27, 28, (29), 30, 31, 32, 33, 42, (43), 44	24, (35), 37, (40)	1, 2, 3, 6, 14, 19, 26, (34), 36, 38, 39, 41	
The numbers represent the specific dictionaries that were investigated and which are listed in the addendum				

19 (+ 6) of the 44 dictionaries comment on the user in the introductions as shown in Table 1 though the amount of detail provided differs significantly as exemplified in Table 2. Some of the dictionaries do not directly describe their intended user(s) in their introductions, though their titles reveal whom these dictionaries have as their target user, e.g. Cambridge *Advanced Learner's Dictionary*. For those (online) dictionaries that do not mention their intended user, it may be because the producers assume that the users know the dictionaries beforehand — some of them did not even have an "about"-section, and the ones that did, primarily mentioned the incorporated data types, i.e. the content of the dictionaries. Most of the general dictionaries that do comment on their users do this only very superficially, e.g. calling them "well-informed adults" or "advanced learners", cf. Table 2. Some of the specialised dictionaries define their users as sports lovers, fanatics or fans, but the characteristics of these are not clear either. Not surprisingly, the analysis of the intended user types in the dictionaries shows a lack of clarity in the description of dictionary users. Also, some of the dictionaries are targeted towards many highly different user types, e.g. Dictionary 7, which is problematic as it is not possible to help all of them in

the same situations; for example, a high school student and a scholar or a professional will not be satisfied by the same meaning explanation for reception, cf. Nielsen (2008: 181), who writes that "user profiles of experts and semi-experts show that these user groups have factual and linguistic competences different from those of laypersons and therefore have different needs." Another problematic issue was identified in Dictionary 44, in which the authors in the introduction write that "[w]hile there are many textbooks for the sports medicine and exercise health professional, there is a huge demand for information on these topics by the general public. It is to fill this need that we have written ...", but on the back cover (probably written by the publisher or marketing staff) it states that the book is "[e]qually valuable on the bookshelves at home for family use or at the clinics of sport and health professionals", i.e. the identification of the intended user is not the same when comparing the introduction and the back cover of the book.

Table 2: Examples of user descriptions in the investigated dictionaries

Dictionary	Dictionary type G = general S = specialised	User description	User type
7	G	This edition has been prepared with a constant regard for the needs of the high school and college student, the technician, and the periodical reader, as well as of the scholar and professional.	General user Layman Semi-expert Expert
10	G	The well-informed contemporary adult.	General user
11	G	Intermediate and advanced learners of English.	Intermediate learner Advanced Learner
23	S	There is nothing in this book for the expert on any particular sport or game. There are no new facts. All that it contains is, so to speak, common knowledge. However, someone who knows little or nothing about, say, yachting or fox-hunting or <i>shatranj</i> or skat, and who wishes to find out some basic facts, will, I hope, find it serviceable.	Layman
26	S	It will be of particular help to health and fitness trainers, students of PE, coaches, and athletes [...] It is also an essential reference for anyone wishing to improve fitness through exercise and healthy eating.	Layman Semi-expert (Beginner ³) (Advanced ⁴)
27	S	An ideal reference for students of sport and exercise sciences, as well as those with a general interest in health and fitness.	Layman (Semi-expert)
28	S	... so that a reader knowing little or nothing about a sport can understand the term in question.	Layman
33	S	... a must for all tennis-lovers and a valuable reference book for all sporting enthusiasts.	Layman (Semi-expert)
44	S	The general public	Layman

2.1.2 The user situation

The number of the investigated dictionaries that comment on the user situation is higher than the number of dictionaries that comment on the user (some of them comment on both), though the mention of user situation ranges from rather broad to more specific comments such as the dictionary having "reading and writing solutions" (Dictionary 4) and that the dictionary "is intended to help the reader understand a sport when he watches it for the first time. The descriptive section explains how it is played — as distinct from how to play it" (Dictionary 20). A few of the dictionaries do not even contain an introduction, but simply write a few pieces of information on the cover, cf. Table 1; this is especially the case for the single-field specialised dictionaries. This is an interesting observation considering the fact that Bergenholtz and Tarp (1994: 173) describe the introduction as an obligatory dictionary component. Words and phrasings such as "a reference (book)" (Dictionaries 25 and 27), "it will be of particular help" (Dictionary 26) and "it is hoped that this book will help" (Dictionary 25) do not adequately explain what kinds of tool the dictionaries are and what they intend to help the user do. It appears that the terms "reference book" and "handbook" correspond to what the function theory calls cognitive dictionaries, but because the functions in these dictionaries are not clearly described in the introductions, these examples have not been added to the "described" columns in Table 1. A number of both the general and specialised dictionaries comment on their content together with or instead of the function, e.g. Dictionary 4, though these two things should not be confused: different types of data (content) may serve the same or different functions so the dictionaries should aim to define what they want the dictionary to be used for, not only describe what they contain. As expressed by Bergenholtz and Bergenholtz (2011), "a dictionary is a tool designed to be used for one or more specific tasks. Of course, this does not mean that you cannot use it for other tasks, but then it will not be as useful." Thus, if dictionary users should be able to use a dictionary optimally, they should be told what the genuine purpose of this dictionary is.

Of course, the fact that the user situation and user are not mentioned in the introductions does not mean that these two central categories have not been applied in the production of the dictionaries — it may be because publishers do not want to restrict their clientele in the description of their product (Wingate 2002: 39). However, it is rather problematic that not all dictionary producers characterise their dictionaries explicitly or with the necessary scientific rigour since this is what is supposed to help a potential user choose a specific dictionary according to his profile and intended use of the dictionary. Often, the producers simply mention these categories for the sake of appearance, cf. Fuertes-Olivera and Tarp (2014: 55).

The following section presents some of the problems that were identified when a comparison was made of the introductions presented here in Section 2.1 to the meaning explanations in the dictionaries.

2.2 Problems with the meaning explanations

The following is a list of the problems identified in the analysis of the investigated dictionaries. Some of these problems have been identified before in lexicographical literature in connection with general dictionaries, e.g. Lombard (1991), but in the present discussion of these points, the criticism will be connected to the user and user situation as well as to sport and fitness, which to this author's knowledge has not been done previously. A selected number of these points — no. 2, 3, 6, 8, 9 and 10 — regarding the incorporation of specialised data in the investigated dictionaries is elaborated below.

1. **Inaccuracy:** The explanations are incorrect in terms of content; examples from other subject fields can be found in e.g. Bergenholtz and Kaufmann (1996) and Hashimzade et al. (2014)
2. **Over-specification:** The explanations are too narrow in terms of the specialised field; they focus only on the most popular branches of sport in the country of production
3. **Under-specification or Simplification:** The explanations are so simplified that you cannot distinguish the relevant lemma from other related lemmata, cf. the purpose of the *differentia specifica(e)* in a classical definition
4. **Irrelevance:** The explanations contain non-relevant data, i.e. data that do not support the given dictionary function
5. **Circularity:** The explanations incorporate the lemma or a derivation, which means that the user will most likely need to make an additional lookup or is left with no help; this is called a derivational definition, cf. Wingate (2002: 35)
6. **Unsuitable formulations:** The explanations are not explained in a way that serves the intended user in the intended user situation
7. **Synonym definitions:** The explanations contain only a synonym that functions as a cross-reference to another lemma or other lemmata, which results in unnecessary additional look-ups, or that is assumed to be known by the user; see also Lew (2013: 298)
8. **Non-lemmatisation:** Non-lemmatisation especially of analytical compounds and especially of terms in running and fitness training
9. **Scientific uncertainties:** Non-expression of scientific uncertainties or of recent scientific research results
10. **Diatechnical markers or Subject labels:** Inconsistency in or lack of marking of specialised fields in the meaning explanations, cf. Rull (2003) and Van der Merwe and Fuertes-Olivera (2014)

11. **Dependency:** The definitions depend on preceding meaning explanations, i.e. they are not autonomous and cannot be understood independently, cf. Lew (2013: 299)

Point 2: Over-specification

In Agerbo (2015), the term *red card* is discussed based on its description in three English and two Danish general online dictionaries. In the current study, this has been expanded to 17 dictionaries (the rest of the analysed dictionaries did not contain this lemma), some of them general dictionaries, others sports (multi-field or single-field) dictionaries. The result from an expanded study of the same term provides the same conclusion as the smaller study:

Almost all the dictionaries that have lemmatised *red card* associate this term with its use in football, which is problematic because the term has different meanings in different sports. Besides football, Dictionaries 3 and 43 also mention rugby, and Dictionary 4 also mentions handball in which (it is implicitly argued) the red card has the same function as in football. Dictionaries 2, 9 and 21 take a more general approach, either stating that the explanation is in sports or mentioning football as an example of a sport in which the red card is used. Dictionary 28 does not even mention the relevant sport (there is no diatechnical marker and no sport is explicitly mentioned, cf. Point 10) but based on the explanation of the term, it appears that this dictionary also describes the use of the word in football. In Dictionary 33, it is demonstrated that the word carries different meanings in different sports, but even in this dictionary, the treatment is too narrow in terms of the specialised fields since only field hockey, soccer, volleyball and wrestling are mentioned; as explained in Agerbo (2015: 151), the card also has different meanings in for example badminton and race walking. In summary, most of the dictionaries, specialised as well as general, have only selected the use of the word in a single branch of sport, thus neglecting other sports branches. The same is the case for e.g. the lemma *heavyweight* for which 10 (+ 1) of 22 dictionaries only mention boxing, not weightlifting, wrestling or horse shows in which this term is also applied, and these 10 dictionaries are both general and specialised dictionaries: G = 1, 3, 4, (6), 8, 11, 41 and S = 22, 25, 30, 33. Dictionary 6 is in parenthesis because the simple definition only includes boxing, whereas the full definition includes boxing and wrestling.

In a general dictionary, it can be difficult to show all meanings of a sports term as it is applied in the different sports, and general users are most likely not interested in all of them unless they have an information need related to a cognitive situation. In dictionaries for learners of English such as Dictionaries 3, 5 and 8, the typical approach of selecting the most frequent meanings (the use of the term in the most frequent/popular sports) results in the most useful explanations for both production and reception as these users will most likely come across the word in its most frequent use, i.e. in football. However, in dictionaries for native speakers such as Dictionaries 2 and 9, it would have been

better (also) to present the user with explanations of the infrequent meanings of the term (e.g. in English culture, this could be in badminton, handball and/or race walking) as most native speakers will most likely be familiar with the frequent uses, whereas coming across the infrequent uses in a text (i.e. in a reception situation) will result in an information need and entail a lookup situation.

Some of the specialised dictionaries partially narrow down the sports fields they cover in the introductions, e.g. Dictionaries 19 and 23, of which the latter states that the area of attention is, "American; popular; on the collegiate or professional level; modern", but none of the investigated dictionaries apart from Dictionary 25 provides a complete list of covered sports (there is a list in Dictionary 18, but it is not complete). This means that the user cannot know whether a sport has been intentionally deselected or has been overlooked. For example, in the Danish general and sports dictionaries, e.g. Dictionary 14, the explanation of *handball* covers only the use of the word in the Danish invented sport, in which two teams each with seven players try to score goals, whereas in a number of the English general and sports dictionaries, e.g. Dictionary 6, only the American game with the same name, in which two or four players try to hit a ball against a wall, is mentioned.

Table 3: Red card

Explanations of the lemma <i>red card</i>		
1	G	in football , a small, red card that is shown by the referee (= the official who is responsible for making certain that the rules are followed) to a player who has not obeyed a rule and who is therefore not allowed to continue playing
2	G	(<i>sport</i>) a card of a red colour displayed by a referee to indicate that a player has been sent off
3	G	In football or rugby , if a player is shown the red card, the referee holds up a red card to indicate that the player must leave the pitch for breaking the rules.
4	G	a red card held up by the REFEREE in a football match , to show that a player has done something against the rules and will not be allowed to play for the rest of the game [→ yellow card]
5	G	a card shown to a football player for a serious offence against the rules, ordering them to leave the field. A player is shown a yellow card for a minor offence and if they are shown another yellow card for a second offence, they are then automatically shown a red card too.
6	G	<i>soccer</i> : a red card that a referee holds in the air to indicate that a player who has broken the rules of the game will not be allowed to continue playing
8	G	(in football (soccer)) a card shown by the referee to a player who has broken the rules of the game and is not allowed to play for the rest of the game
9	G	(In soccer and some other games) a red card shown by the referee to a player who is being sent off the field
14	G	(translation) card which a football or handball referee present when he dismisses a player

21	S	A red card shown by the referee (as in international soccer) to indicate that a player is being sent off.
22	S	<i>Soccer</i> : The red card, about the size of an ordinary playing card that is shown by the referee to indicate a player is being sent off the field for a violation of the rules.
23	S	ASSOCIATION FOOTBALL This is shown by a referee to indicate that a player is being sent off
28	S	A playing card-sized card that a referee holds up to signal a player's removal from the game; the player's team must play the rest of the game shorthanded; presented for violent behavior or multiple rule infractions (two yellow cards = one red card)
30	S	(<i>association football</i>) a red-colored card shown by the referee to a player who is being sent off, either because he has committed a serious foul or because he has already been shown two yellow cards
33	S	<ul style="list-style-type: none"> • Red card (Field Hockey) • Red card (Soccer) • Red card (Volleyball) • Red card (Wrestling) <p>Field hockey ... Soccer: A player is dismissed from the field of play if the referee shows him a red card. This happens either because the player has received two yellow cards in the same match, or because the referee judges that he has committed a single offense that warrants a sending off in its own right. When a player has been sent off, his team must play the rest of the game with 10 players (unless further players are dismissed). Volleyball ... Wrestling ...</p>
41	G	<i>soccer</i> a red card that a referee holds in the air to indicate that a player who has broken the rules of the game will not be allowed to continue playing — compare YELLOW CARD
43	S	(Under the label <i>soccer</i>): Expulsion (red card): Players are expelled if they commit a serious foul, are violent, use abusive, offensive, or insulting language, or receive a second yellow card during the game. (Under the label <i>rugby</i>): A red card is handed out in the case of especially rough play or dangerous behaviour, or if a player has repeatedly contravened the rules. It results in the player's immediate expulsion.
Bold formatting is added to highlight the external (diatechnical markers) or internal field specific marking		

Point 3: Under-specification or Simplification

For the lemma *goalkeeper* (focussing on the association football meaning), many of the explanations are so simplified that you cannot distinguish the relevant lemma from other related lemmata, i.e. this is a case of under-specification. The analysed explanations can be divided into the following three groups (though Dictionary 43 is an exception as its explanation involves the following components: the goalkeeper is the last line of defense + he stops shots + he has certain privileges + he directs the defense):

1. A person whose job it is to guard the goal, defend the goal, protect the goal, or prevent the ball from going into the goal (3, 4, 5, 6, 7, 8, 9, 10, 11, 22, 30, 41, 42)
2. + a player (who stands) in the goal or a player who stands in front of the goal (1, 2, 13, 14, 15)
3. + a player who has certain privileges ((12), 18, 19, 21, 28, 33)

It is remarkable that none of the general dictionaries include data on the special privileges, which is actually the main characteristic that distinguishes this player from the other (defending) players. It is the task of all defending players to stand somewhere in front of the goal, guard the goal and try to prevent the opponents from scoring, not only the goalkeeper's job, and for example in case of a corner kick, some of the defending players also stand in the goal next to the goalposts to help the goalkeeper, which means that both groups 1 and 2 above are not particularly good at describing the meaning of *goalkeeper* for neither communicative nor cognition purposes. In addition to these criteria, a goalkeeper must always wear clothing that is different from the other team players' clothes, making it possible to distinguish this player from all the other players, but this criterion is only mentioned in Dictionary 18.

Of the analysed dictionaries, it is only some of the specialised dictionaries whose explanations can be used for reception for general users, not any of the explanations in the general dictionaries. If a meaning explanation should be incorporated in a general dictionary for reception, it would have to include at least the three points mentioned above though this would still result in a rather broad explanation. Another solution would be to add specifications about the goalkeeper in each sport and make them accessible via links, resembling the approach in Dictionary 33. If the meaning explanation were to be incorporated in a production dictionary, it would not be necessary to distinguish between different sports, and only a rather short explanation would be needed, e.g. "the player who is assigned certain privileges to protect the goal," because the user is interested in data types that can assist him in text production; in this situation, the user (typically) already knows the meaning of the word. For cognition, the explanations provided in Dictionaries 12, 18 and 21 are useful, and this also corresponds somewhat to the description of their function(s) in their introductions, e.g. to add to the sum of knowledge for a sports addict.

Table 4: Goalkeeper

Explanations of the lemma <i>goalkeeper</i>		
1	G	(in some sports) the player who stands in the team's goal to try to stop the other team from scoring
2	G	(<i>sport</i>) a player in the goal whose duty is to prevent the ball, puck, etc, from entering or crossing it
3	G	A goalkeeper is the player in a sports team whose job is to guard the goal.
4	G	the player in a sports team whose job is to try to stop the ball going into the goal [= goaltender <i>American English</i>]

5	G	the player whose job is to stop the ball going into the goal in games such as football
6	G	a player who defends the goal in various games (especially soccer)
7	G	a player who defends the goal in various games (as hockey, lacrosse, soccer)
8	G	(in football (soccer), hockey, etc.) a player whose job is to stop the ball from going into his or her own team's goal
9	G	A player in soccer or field hockey whose special role is to stop the ball from entering the goal.
10	G	A player assigned to protect the goal in various sports.
11	G	In games such as football and hockey, the goalkeeper is the player who guards his or her own team's goal.
12	S	(translation) player who guards → the goal. (Bandy) ... (Football) As the last person in the defense, he has the most responsible role as a mistake made by him cannot be amended. He has a special role 1) because in his own penalty area he may use his hands and 2) because none is allowed to push him in this area unless he is <i>holding</i> the ball (i.e. not fisting it) or intentionally tries to hold back the opponent. When the keeper is holding the ball none may try to kick it away from him. Requirements to a good goalkeeper are esp. 1) the ability to read the game, 2) quick reflexes, 3) good with his hands, 4) reliability in the game. To be a <i>very good</i> goalkeeper, he must also be able to leave the goal at the right moments. A keeper <i>must</i> 1) use his hands instead of his legs when possible, 2) always try to catch the ball and only fist it when there is a risk of being pushed to the ground, 3) quickly get the ball away from his field — preferable to one of the wings, 4) at corner kicks position himself at the furthestmost goal post, 5) at shots from the wing position himself at the nearest goal post. The keeper is only permitted to leave the goal when he is sure that he will catch the ball first; he <i>must do it</i> when an opponent has been able to dribble away from the last defender in the field. The keeper must cooperate with his backs and direct them — and other players — when necessary. Of the usual technical skills required by a field player, the keeper must be good at kicking, especially to a "dead" ball so he can kick his own goal kicks. The keeper must be good at catching and throwing as well as fisting. (Handball) ... (Hockey) ... (Ice hockey) ... (Water polo) ...
13	G	(translation) the player in football and similar games who is placed in the goal (in order to defend it)
14	G	(translation) player in a ball game who is placed in the goal and whose job it is to prevent the opponents from scoring, e.g. by catching the ball
15	G	(translation) a player in e.g. a football or ice hockey team who stands in front of the goal to prevent a score
18	S	A player who guards the goal in certain goal games, notably field hockey, ice hockey, lacrosse, soccer and water polo. The goalkeeper usually keeps closely within the goal area and virtually never takes part in offensive maneuvers. He is, as a rule, accorded privileges denied his teammates. Abbreviation: G. Other specifications: <i>Field hockey:</i> ... <i>Ice hockey:</i> ... <i>Lacrosse:</i> ... <i>Soccer:</i> Similar duties to those listed. The goalkeeper, unlike other players, can touch and throw the ball when he is within the penalty area. He can likewise carry it, but cannot take more than 4 steps without bouncing it. He has <i>no</i> special privilege when he is outside the penalty area (which see). His equipment is no heavier than that of other players (he wears no shinguards), but he should wear colors distinguishing him from other players. He may go anywhere in the field, but only the player designated by a team to the referee as its goalkeeper may enjoy the special goalkeeper's privileges listed above. <i>Water polo:</i> ...
19	S	(field hockey) ... (ice hockey) ... (soccer) the player who stands between the posts and is allowed the use of his hands within his own penalty area.

21	S	1. The defensive player in various goal games who normally plays in front of the goal he is defending to keep the ball or puck from going in for a score. The goalkeeper is the team's last line of defense and he often is allowed to use special equipment and is accorded certain privileges while in his normal position that are denied his teammates. The goalkeeper in soccer is the only player who is permitted to play the ball with his hands but only while he is within the penalty area ...
22	S	The player assigned to protect the goal.
28	S	The player positioned directly in front of the goal who tries to prevent shots from getting into the net behind him; the only player allowed to use his hands and arms, though only within the penalty area.
30	S	(<i>general</i>) the player who defends the goal in a game such as association football
32	S	Field hockey ... Ice hockey: ...
33	S	Each team has a designated goalkeeper, whose role is to stop the opposing team from scoring a goal. The goalkeeper is the only player who is allowed to handle the ball, but this is permissible only inside his own penalty area.
41	G	a player who defends the goal in various games (especially soccer) — compare GOALTENDER
42	G	the soccer or hockey player assigned to protect the goal
43	S	The only player allowed to use his hands, he must stop shots from the opposing team. He is his team's last line of defense and directs his teammates when the other team is attacking.

Point 6: Unsuitable formulations

Usually, explanations are targeted towards a specific intended user type. This was clearly the case in the learner's dictionaries in which many applied a restricted vocabulary and simple syntax (for a criticism of restricted vocabularies, see De Schryver and Prinsloo (2011: 8)). However, not all the investigated dictionaries managed to match the language to the intended user.

Many anatomical and physiological terms that once were foreign to the general dictionary user have today become part of the general vocabulary due to the increased popularity of fitness and sport, a process known as de-terminologisation, see e.g. Meyer and Mackintosh (2000). This is for example happening with words such as *ATP*, *cardiovascular*, *deltoid muscle* (or *deltoids*), *hypertrophy* and *plyometrics*. Such words should therefore be incorporated into both general and specialised dictionaries. Even though these may be seen as specialised words (at least at the initial incorporation in general dictionaries), the meaning explanations of these should be formulated in a way that the intended user easily understands them. An example of this is presented by Bergenholtz and Kaufmann (1997: 115-117), who suggest different explanations of *gene* for laymen and semi-experts, respectively. In Dictionary 24, a specialised dictionary, the very first lemma in the central list, *abdomen*, is explained as "[t]he part of the body in mammals that lies between the thorax and the pelvis and encloses the viscera, or internal organs of the body (commonly called the intestines); the belly." According to the dictionary book jacket, the dictionary "uses concise explanations to make the world of exercise understandable to everyone." In terms of vocabulary, the first part of the explanation is not suitable for

everyone ("everyone" is also rather a bold statement to make). Dictionary 9, a general dictionary, presumably targeted at general users who are native speakers of English, presents the same problem, "[t]he part of the body of a vertebrate containing the digestive and reproductive organs; the belly" (*belly*, however, is explained as "[t]he front part of the human trunk below the ribs, containing the stomach and bowels", which is better suited for the English speaking general user). Dictionary 1, also a general dictionary, manages much better to formulate its content to its target user (advanced learner), "the lower part of a person's or animal's body, containing the stomach, bowels, and other organs, or the end of an insect's body" though it could be discussed whether this is one or actually two explanations.

If we also compare Dictionaries 1 and 9 in their explanations of *ATP*, the explanation in Dictionary 9 is better suited for a semi-expert than for the intended non-expert, "[a] compound consisting of an adenosine molecule bonded to three phosphate groups, present in all living tissue. The breakage of one phosphate linkage (to form adenosine diphosphate, ADP) provides energy for physiological processes such as muscular contraction" whereas the explanation in Dictionary 1 (for advanced learners) is directed at a learner or even a general user, "an important chemical in the cells of living organisms that store energy and releases it when it is needed."

Point 8: Non-lemmatisation

Some of the specialised dictionaries contained a number of lemmatised compounds, e.g. Dictionary 18 contained 45 lemmata with the word *running*, e.g. *running broad jump* and *running half gainer*, many of them used in fancy diving. However, many of the sports dictionaries did not include such compounds (at least not in the case of the test compounds) and neither did most of the general dictionaries. This is not a surprise as the general practice is not to lemmatise analytical lemmata, i.e. lemmata whose meanings can be inferred (or are assumed to be inferable) based on the single words constituting the compound. However, for reception, explanations of compounds such as *barefoot running* and *protein drink*, which may seem easily interpretable, should be provided, and neither for laymen nor semi-experts should the explanations simply be: 'running barefooted' or 'running without footwear' and 'a drink containing proteins', cf. Bergenholtz and Agerbo (2014a), who suggest the following explanation of *protein drink* for reception in a general dictionary:

drink that consists of protein rich ingredients or protein powder mixed with water or milk; typically used by sportsmen to maximize the effect of a workout, by patients who have gone through a surgery, or people on a diet in which case they need to increase their consumption of proteins and decrease their consumption of sugar

The simple explanations mentioned above may be useful in a production dictionary, but they do not help any of the users when they have an information

need related to reception — a more detailed explanation is needed, cf. an analysis of corpus data, which for the lemma *protein drink* showed that people often discuss (1) when this drink is used and (2) for what purpose. In a sports dictionary, only the comment about sportsmen in the explanation above is relevant, and the rest after the semi-colon should therefore be excluded from the explanation. In a reception or cognitive sports dictionary for semi-experts, the explanation might also contain data on how the protein drink affects the body physiologically, but this is not relevant in a layman explanation. Thus, the amount of detail provided will depend on both the user and user situation.

None of the analysed dictionaries have lemmatised the compound *barefoot running*, but neither in this case can the user actually infer its meaning in a reception situation, e.g. that barefoot running could also be done in minimal shoes and that they are designed for the foot to land on the surface in a specific way. A suggestion for a general dictionary could be:

special way of running that involves running barefooted or in minimal shoes, which have a very thin sole, making it possible to run as naturally as possible, i.e. as if you were not wearing shoes, and which results in you landing on flat feet with your weight on the anterior part of your feet

Not only (analytical) compounds were often disregarded, but also terms in running and fitness. Though the present study was not a statistical study, it was somewhat surprising to find that a number of the general dictionaries, for which their producers claim that they are regularly updated, did not contain such terms even though running and fitness is today very popular, which is why they should be included in modern dictionaries.

Point 9: Scientific uncertainties are not expressed

The word *sidestik* (Eng. *side stitch* — not lemmatised in the English dictionaries, cf. Point 8), appeared in Dictionary 13, a Danish general dictionary, in which the explanation is (translation), "pricking pain in the right side of the body, e.g. due to excessive strain." As for the first part: side stitches often occur on the right side, but they can also occur on the left, i.e. it is incorrect only to write that it is the right side in which it occurs (cf. Point 1 in the list). As for the second part: no scientific research has yet been able to identify the causes of side stitches; there are many theories and suggestions as to why side stitches occur and how to prevent them from occurring, but so far no scientific evidence has been able to establish any of this as a fact. This means that pointing at "excessive strain" in the explanation is a scientific uncertainty, not a fact, which the user is not made aware of. In Dictionary 27, a specialised dictionary on food and exercise, the author sometimes comments on scientific uncertainties, e.g. for the lemma *muscle growth*, it is explained that this is mainly due to an increase in the muscle fibres, but that "a growing body of evidence" suggests

that it could also be due to an increase in the number of fibres, though this has not been demonstrated conclusively. The same is expressed in Dictionary 44 for the lemma *hypertrophy* in which it says that hypertrophy "may also be due to an increase in the number of muscle fibres" ("may" expresses uncertainty). Such uncertainties are not only needed in specialised dictionaries, but should also be expressed in some general dictionaries that include such terms. Dictionary 9, a general dictionary, states that hypertrophy is "[t]he enlargement of an organ or tissue from the increase in size of its cells", i.e. this dictionary only states verified facts and ignores the scientific uncertainty, which is also problematic. As mentioned by Leroyer (forthcoming), explanations should be updated whenever progress in science and technology modifies meaning and may jeopardise understanding. For a word such as *stretching*, it should be made clear that there is no agreement as to how much it helps or how much stretching one should do, and for a term such as *compression sock* (one of the test words that was not lemmatised in any of the analysed dictionaries), the uncertainties concerning its function should also be highlighted, at least for cognition, but possibly also for reception if this explanation comments on the function of this type of sock as in the following suggested meaning explanation:

sock made of cotton, polyamide and spandex that covers the lower part of the leg and increases the pressure on the peroneus during physical activity, e.g. running or bicycling, which improves the blood circulation and thus entails better performance, improved endurance and decreased recovery time; in recent years, compression socks have become very popular in sports, but scientific research has yet not been able to verify the performance-enhancing effect of these socks when using them for physical activities

Admittedly, whether such comments on scientific research is useful for all users in a reception situation could be questioned, as a user looking up the term after having read it in a text might not be interested in these data. But, simply stating the assumed causes, functions etc. of the item as a fact is not helping the user either. For a general user in a production situation, such comments should not be provided. An interested layman and a semi-expert would want to be provided with these data both for reception and cognition. Gouws and Tarp (forthcoming) explain that information overload must be avoided, and therefore the lexicographer will have to pay attention to the intended user and user situation when making decisions about what to include and exclude from the meaning explanation. It is not a question of whether the meaning components are semantic or encyclopaedic, and there is not one correct solution, but many possible ones of which the lexicographer must choose the most useful one for the given function, cf. Nielsen (2010).

Point 10: Diatechnical markers or Subject labels

In Dictionary 28, a sports dictionary, no diatechnical markers were used at all, which means that for lemmata used in several sports branches, the user is left

with no help to determine the context, thus its usefulness is limited even though the user knows that the general field is sport. In a number of the other specialised dictionaries, e.g. Dictionary 21, each meaning of a term is marked with a diatechnical marker, e.g. *field hockey* and *soccer*; in some cases, a general explanation is provided before the more specific explanations, thus helping laymen, who are not interested in reading about the meanings in all the different sports. In many of the general dictionaries, the use of diatechnical markers is unsystematic, e.g. in Dictionary 14, a Danish general dictionary (see Rull (2003) for a similar analysis of the application of diatechnical markers in Spanish general dictionaries). On its website, the producers write that (a) a meaning is marked with a field marker if the most typical use of this word is in this specific field and if it is less known in general language, and that (b) if the word is monosemous and used in a specialised field, this is not marked, but its relation to the field will be mentioned directly or indirectly in the explanation. As can be seen in Table 5, this approach has resulted in highly different solutions to the presentation of data in the case of the sports field; only the diatechnical marker SPORT is applied, thus all subfields and specific sports, e.g. *ball games* and *tennis*, will have to be mentioned using other strategies, which the dictionary user might have difficulties understanding correctly. For example, what does "and similar sports" involve (no. 3), why is "in ball games" written before the explanation for *skytte*, but inside the explanation for *tackling* (no. 7), and is the comment at the end of the explanation of *back* a separate explanation (no. 17)? As argued by Van der Merwe and Fuertes-Olivera (2014: 84) in their analysis of South African wine dictionaries, subject labels are very useful, both for communication-oriented functions (though especially in bilingual dictionaries) and also in cognitive-oriented dictionaries. However, they have to be applied in a way that helps the user understand the data he is reading instead of confusing him.

Table 5: Diatechnical markers in Dictionary 14 (Den Danske Ordbog)

Type	Marking of the field(s) in or around the explanation	Examples (translations)	Terms* (translations)
1	All the relevant sports branches are indirectly integrated in the explanation	card shown by a football or handball referee when he sends off a player	<i>red card</i> , line player, hook, jab, ring corner, uppercut, marathon run
2	The general field, sport, is indirectly integrated in the explanation	the fixed time that a sports game lasts	<i>game time</i> (pol1), boxing, ironman(pol2a), weightlifting, top seeded
3	Examples of the relevant sports branches are mentioned in the explanation	a way of playing in e.g. ice hockey and football in which a player uses a board or a co-player to quickly send the puck or ball past an opponent	<i>(do a) one-two</i> , interval training
		medium heavy weight class in boxing and similar sports — e.g. in amateur boxing 71-75 kg	<i>middleweight</i>

4	The relevant sports branches are mentioned in the explanation	person who manages fights in boxing, wrestling and martial arts and who makes sure that the rules are followed	<i>referee</i> , knockout, double fault, tennis player, match ball
5	The relevant sport is mentioned before the explanation	in football : a kick to the ball taken from one of the four corners, and which is awarded to the attacking team when the ball crosses the back line after last being touched by a player from the defending team	<i>corner kick</i>
6	A subfield is mentioned before the explanation	in ball games : player who participates in the attack in one side of the field	<i>wing</i>
7	A diatechnical marker is mentioned before the explanation + A subfield is mentioned	SPORT in ball games : player who (with luck) kicks or throws the ball towards or in goal	<i>shooter</i>
		SPORT tackling an opponent in ball games	<i>tackle</i>
8	A diatechnical marker is mentioned before the explanation	SPORT area close to one of the sides of the football field where coaches, audience and others are placed	<i>side line</i> (pol1a), <i>wing</i> (pol1c), <i>midfield</i> (pol2), <i>line</i> (pol2b), <i>line</i> (pol2c), <i>ace</i>
9	A diatechnical marker is mentioned before the explanation + An additional narrowing, the typical sport, is mentioned after the explanation	SPORT area between the goalmouths in a field where the midfielders primarily work, and from where the attack is initiated — particularly in football	<i>midfield</i> (pol1)
10	Examples of relevant sports are mentioned after the explanation	special kind of throw thrown directly towards the goal which a team is awarded after a foul has been made by the opponents — occurs e.g. in handball and basketball	<i>penalty throw</i>
11	A diatechnical marker is mentioned before the explanation + Indirect examples of sports branches are integrated in the explanation at the end	SPORT line that marks the long side on e.g. a football field or tennis court	<i>side line</i> (pol1)
12	A general diatechnical marker is mentioned before the explanation + Examples of relevant sports branches are integrated in the explanation at the end	SPORT shooting or hitting a ball hard before it touches the ground in e.g. football or tennis	<i>volley</i>
13	A diatechnical marker is mentioned before the explanation + The specific sport is mentioned indirectly in the explanation	SPORT each of the three or more rounds lasting three or two minutes that a boxing match is divided into	<i>round</i>
14	A general diatechnical marker is mentioned before the explanation + The specific sport is mentioned in the explanation	SPORT part of a set in tennis where only one of the players has the right to serve · is won by winning at least four duels; a player must win minimum six games before he has won the set	<i>game</i> (pol4a)
15	A diatechnical marker is mentioned before the explanation + The relevant sports branches are mentioned after the explanation	SPORT bar with a number of weight plates secured to it — used in weight lifting and weight training	<i>barbell</i>

16a	Integrated in the explanation at the end are some examples of relevant sports branches	passing the ball across the field in e.g. football	<i>diagonal pass</i> , clay court, serve (pol1)
17	At the end are some examples of relevant sports branches and these are used to comment further on the core explanation	player in certain ball games who primarily has defensive tasks in one side of the field — also has an important offensive role in e.g. handball	<i>back</i>
18	An additional narrowing in the form of the most typical sport is mentioned after the explanation	ball that is played or jumps back into play after a pass, a shoot or a duel — especially in football	<i>second ball</i>
19	No marking of the general field nor of more specific sports branches	the time a player spends on the field during a game	<i>game time</i> (pol1a), list of top scorers, running shoe, forehand, serve (pol1a), ultra-running
*44 of the investigated 72 terms were lemmatised in the online edition, but only 35 were lemmatised in the paper version. The examples provided in column 3 are of the italicised terms in column 4.			

2.3 Main guidelines for writing meaning explanations of specialised terms

- (1) *The data must be lexicographically relevant*
Lombard: ~ completeness
- (2) *The explanation must be easily understandable for the intended user type*
Lombard: ~ clarity
- (3) *The data must be correct and reliable*
Lombard: accuracy
- (4) *The meaning explanation must be independent*
Lombard: independency
- (5) *The incorporation of polysemes should be based on the user and user situation, not frequency*
- (6) *Scientific uncertainties should be presented in explanations as uncertainties, not as fact, and they should be mentioned whenever they are relevant for the given function*
- (7) *The relevant sports field(s) should be marked with field labels in a way that supports the dictionary function and that makes it possible to quickly and easily access the relevant explanation*

Lombard (1991) comments on some of the same points for writing good meaning explanations, though not all are similar and not all the conclusions are the same. The first guideline, relevance, means that the incorporated semantic elements must be selected based on their relevance according to the function.

Nothing more (over-specification) and nothing less (under-specification) than what is relevant for the intended user situation and user should be incorporated. Lombard agrees with this, but he considers encyclopaedic or extra-linguistic information non-essential whereas relevance is not based on a distinction between semantic and encyclopaedic data. The second guideline, clarity, is also mentioned by Lombard, but in the current paper it is considered from a functional perspective, i.e. that the amount of detail in the explanation, the sentence structure, etc. should depend on the user, which means that the explanation should not always be simple and without detail. The third guideline is not worth discussing as this is something all lexicographers would agree with. However, it should be emphasised that correct and reliable is not the same as "the truth". The lexicographer does not present the truth, but produces a tool that among other things contains explanations that are incorporated in order to help a specific user type with a specific type of need in a specific type of user situation. The fourth guideline is something that can be dealt with in an e-tool by turning some of the words in the explanations into links so the user can easily be taken to the relevant article or by presenting additional explanations as pop-up text when the mouse is moved over the relevant word. But lexicographers should avoid both article internal and article external dependence as this will always mean extra work in the look-up situation for the dictionary user. The fifth guideline entails that the selection of polysemes in a dictionary should not only be based on the most frequent meanings of a term, but what explanations different users may be assumed to be looking for in different situations. The sixth guideline is particularly relevant for specialised terms. In Dictionary 13, a term such as *sort hul* (Eng. *black hole*) is explained and the dictionary mentions that this phenomenon is believed to exist, but that this has yet not been proven. The same approach should be applied for sports terms such as *compression sock*, *side stitch* and *barefoot running*. The seventh guideline means that whenever it is possible and useful to describe the specialised field (especially in a general dictionary), this should be done with a diatechnical marker. The number of markers could be expanded to include all the sports selected for the given dictionary instead of the general marker SPORT. Also, the marker FITNESS should be added to modern dictionaries.

A couple of extra guidelines could be added to the above list: (8) the empirical basis used for the production of explanations should be selected according to the dictionary function and the type of lemma being described, and (9) the semantic elements must be placed in an order that supports the dictionary function. The eighth guideline is not discussed above in Section 2.2, but is derived from descriptions of the empirical basis in the introductions of the different dictionaries. Most general language dictionaries are today corpus-based, but when it comes to explanations of specialised terms, other sources such as experts and other written texts are more useful for deriving semantic components for the explanations, cf. Bergenholtz and Agerbo (2014b). A corpus-based approach, e.g. applied in Dictionaries 4, 5, 9 and 13, will not always

provide sufficient data for explanations of specialised terms as the corpora are not compiled with specialised texts, cf. Bergenholtz and Kaufmann (1996). This guideline concerning the empirical basis is indirectly connected to the meaning explanation. The ninth and final guideline is not discussed above either but was derived when looking at the explanations presented in the next section, which are connected to interpretative and operative functions: for some lemmata, the selection and order of presentation of semantic components will differ across functions, e.g. the explanations of *press-up*. It is not for all lemmata that a difference will occur in the order of the components, and some components that are considered more essential than others will always occur initially.

These guidelines highlight that more attention should be given to sports terms especially in general information tools, and also in specialised tools. These dictionaries are useful for some users in some situations, but no clear lexicographical theory appears to have been applied in the production of these dictionaries. The general dictionaries are mainly driven by a linguistic approach, and many of the specialised dictionaries are written by for example sports journalists with no knowledge about lexicographical theory.

2.4 Additional findings

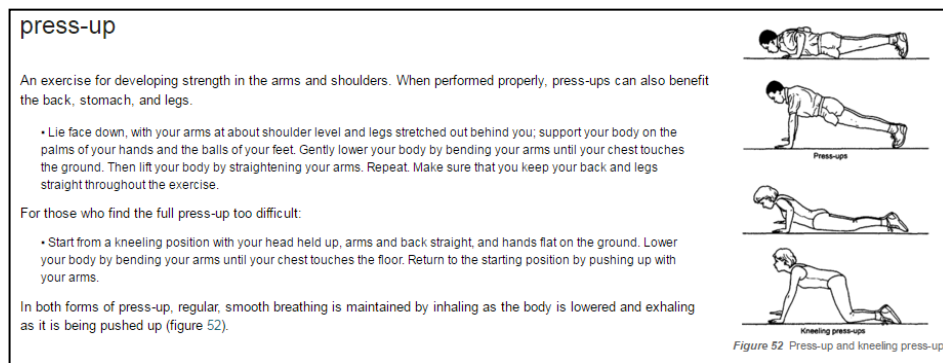
A couple of intriguing discoveries were made in the analyses of Dictionaries 12, 18, 21, 25, 27, 36 and 44: Danish Sports Lexicon (1944/1945), The Dictionary of Sports (1949), Webster's Sports Dictionary (1976), What's What in Sports — the Visual Glossary of the Sports World (1984), Food and Fitness: Dictionary of Diet and Exercise (1997/2003), The Visual Dictionary of Baseball (2001) and The Encyclopedia of Exercise, Sport and Health (2004). These demonstrated the incorporation of explanations that do not aim to satisfy a user's need in communicative and cognitive situations, but which are incorporated into dictionaries with explicit or implicit operative and/or interpretative functions, which are functions that have only been mentioned in a small number of articles in the lexicographic literature, e.g. Tarp (2008b), Tarp (2008c), Bergenholtz, Bothma and Gouws (2015) and Agerbo (2015), but which have yet not been fully integrated into lexicographical theory.

2.4.1 The operative function

In Dictionaries 12, 18, 27 and 36, the meaning explanations do not only comment on the actual meanings of the lemmata, but they also instruct the user in carrying out the exercise, movement or something else to which the lemma refers. For example, in Dictionary 12, the lemma *Hovedspil* (Eng. *heading*) involves an explanation of the part(s) of the head with which you can hit the ball; a description of how you should approach the ball; it explains that you

should keep your eyes on the ball until you hit it; that you should head the ball when defending and attacking; and that you should head the ball downwards when trying to score a goal. In Dictionary 18, the description of *crawl* is much more instructive than what you typically find in dictionaries. In this case, the way of breathing, turning of the head, movement of the arms and movement of the legs is much more detailed than if the word were to be described for reception. Actually, in its introduction, Dictionary 12 states that it, among other things, aims to be an *instructive* reference work, i.e. it not only wants to explain to a user what something is, but also how to do it. Dictionary 18, however, does not mention instructions as part of its function. In Dictionary 27, the introduction states that "the book provides the information you need to make sensible decisions about the food you eat and the activities you undertake." It does not explicitly mention instructions for these activities, but if we look at e.g. the lemma *press-up*, this article does not only contain an explanation, it also provides a separate exercise instruction, both presented as a written and a visual explanation. Dictionary 36 does not contain an introduction.

Figure 1: *Press-up* in Dictionary 27



What has been identified here is an *operative function* (i.e. instructive), which only one of the three dictionaries explicitly mentions in the introduction as one of its functions. A dictionary with an operative function is designed to help a user who has an information need that needs to be solved so he can carry out a mental or physical act. This function is one that we are familiar with from handbooks, manuals, how-to books and similar reference works, but which has yet not been discussed fully in a lexicographical context (Tarp 2008c: 126). However, though the explanation of *crawl* is useful for operation, it could also be argued that the explanation is written to serve a user in a cognitive situation: If the dictionary is designed to help a user looking for information out of curiosity, the dictionary has a cognitive function, but if it is designed to help a user

perform an act, the dictionary has an operative function. This, however, has not been specified in the introduction.

In the introduction, Dictionary 44 mentions that it among other things wants to explain how to do something, e.g. how to lose weight, i.e. one of its functions is the operative function. For the lemma *overweight*, the user is informed that in order to lose weight, s/he will have to participate regularly in physical activity and eat a healthy diet. Though neither this explanation nor the one of *crawl* are formulated with imperatives (which is the typical linguistic form of instructive texts, cf. Nielsen (2006: 15-16)) as in the case of *press-up*, these could still work as instructions because the purpose of these explanations is to help people find information about *how* to do something.

In terms of the relevant users, the operative function entails an expanded user typology, cf. Section 2.1, since for operation, the focus is also on skills, cf. Tarp (2008a: 153-159). Thus, we cannot simply talk about layman, semi-expert and expert, but should incorporate the following types of athletes: beginner, possibly also intermediate, and advanced. However, users in an operative situation are not only characterised by their skills, but (potentially) also according to their specialised (technical) knowledge. A person may for example be skilled in weightlifting, but not have much knowledge about the body's reactions to weightlifting, or he may be both skilled and have knowledge about the field. In terms of the data types wanted for this situation and the characteristics of these users, no answers have yet been provided in meta-lexicography. But for example, the explanation of the lemma *press-up* suggests that an article in an operative dictionary should not only consist of an instructive explanation, but possibly also other data types such as "variants" and "common mistakes", presumably especially for beginners⁵.

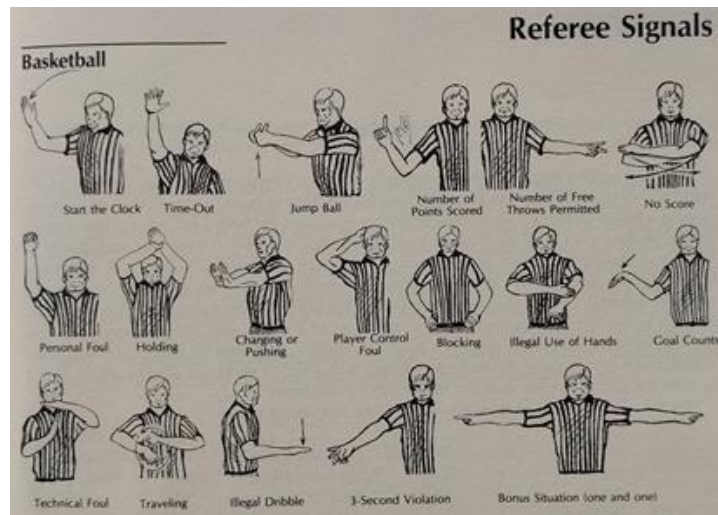
2.4.2 The interpretative function

Both Dictionaries 18 and 21 incorporate as part of the back matter descriptions of referee signals in different sports (e.g. basketball) in the form of illustrations. These non-linguistic signs support what in the function theory has been called the *interpretative function*, cf. Tarp (2008b: 185) who writes that "[r]ecently, it has been discussed whether there is a fourth main user situation, the interpretive one, where the user needs to interpret signals and symbols in the surrounding world, but it is still too early to conclude anything in this respect." Though these visual data are not part of the central list, but are outer texts, the dictionaries appear to have a secondary interpretative function. In Dictionary 18, the images are supplemented by text explaining what the referee is doing and also what this movement means. As in the case of Point 2 with the discussion of the lemma *red card*, the lexicographers have also here only incorporated the referee signals in the most popular sports. Other less popular sports such as handball and volleyball have not been included.

Figure 2: Referee signals in Dictionary 18

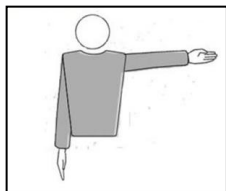


Figure 3: Referee signals in Dictionary 21



Some interesting questions related to research of the interpretative function are for example how such an explanation should be formulated and how the search options in an e-tool with an interpretative function should work. For example, in the above illustrations from Dictionary 21, for the signal *time-out*, a cross-reference to the lemma *time-out* in the central word list could have been made, thus giving the user a more thorough description, and which would turn this dictionary component into an integrated dictionary component, cf. Nielsen (2006: 8). In an e-tool, separate articles could be made of each signal so the user could more easily search for a certain signal e.g. using a photo in the search field or applying keywords in the search field. If the same signal is used across different sports, then the signal could be called polysemous (though not everyone would agree with this use of the term *polysemous*), and different descriptions of this signal should be provided. A suggested article for an interpretative information tool is provided below. In this suggestion, the meaning explanation is preceded by a description of the sign (hand signal), though this part of the article is not compulsory as some non-linguistic signs are almost indescribable.

Volleyball



Hand signal: Arm extended towards the serving team

Explanation: The serving team is requested to put the ball in play

The above-mentioned referee or umpire signals are as mentioned attached in the dictionaries as outer texts, but this is not the case for Dictionary 25, in which these different types of signals are presented in the articles for each different sport (the lemma list only consists of branches of sport), e.g. in baseball, basketball, ice hockey and lacrosse. In Dictionary 21, the lemma *flag* is divided into 5 polysemes, of which the first corresponds to the use in racing, "Any of a series of flags of different colors used to signal the competitors on the course in motor racing and yacht racing"; this text is on the opposite page supplemented by images of the different flags used in yacht racing, each flag supported by a text that briefly describes what the flag looks like and how it should be interpreted. Thus, it is possible to find data that support an interpretative function both in the central lists and in the outer texts. However, as in Section 2.4.1 where the function of the dictionary could either be cognitive or operative, it could also be argued that in the case of referee signals, the function could instead be operative; e.g. if a basketball referee is interested in learning how to signal *time-out*, he will search in an operative dictionary and be provided with somewhat the same data.

For most of the mentioned dictionaries, the analysed data were most likely not included with the aim of creating interpretative and operative dictionaries,

but the data point to the fact that even though these functions have not been discussed in meta-lexicography, they are actually applied in practical lexicography.

3. A functional approach to meaning explanations

It is not a surprise that it is possible to find different meaning explanations in different dictionaries, and some of the investigated explanations seem to have been produced according to a specific user type. However, many have not been produced properly according to the user situation, though this was for example suggested by Nielsen (2011, 2013); as shown in Table 1, in many cases, the user situation is not even specified in the introduction.

Many contributions on lexicographical meaning explanations comment on the form of the meaning explanation, e.g. that it should be formulated using a sentence definition, analytical definition or folk definitions (see e.g. MacFarquhar and Richards (1983), Stock (1988) and Wingate (2002)), and in terms of the content, they argue that the word must neither be over- or underspecified and that it should contain semantic data, not encyclopaedic data. The discussion concerning encyclopaedic and semantic data is not new, cf. e.g. Haiman (1980), Wierzbicka (1992), Bergenholtz and Kaufmann (1996) and Bergenholtz and Agerbo (2014a), of which the latter argue that in lexicography, it is not relevant to formulate a clear-cut boundary between semantic and encyclopaedic knowledge because the goal for a lexicographer is to identify and select meaning items as well as identify, select and distribute semantic components that can solve the needs of the potential dictionary user in a specific type of extra-lexicographical user situation. Thus, when we decide what and how much data to incorporate into a meaning explanation, it is the function and not the semantic-encyclopaedic relation that is the crucial factor. This is demonstrated with some general suggestions below (this is not an exhaustive list):

1. An interested layman wants to know more than s/he already knows about a certain word or theme for a school paper and therefore looks this up in a cognitive dictionary (= **cognition**).
Requires: A much more elaborate meaning explanation than for reception. This could be solved with a lexical note that elaborates on the meaning explanation of the lemma from the reception dictionary, i.e. the lexicographer can show these two together in the cognitive dictionary instead of writing a different meaning explanation.
2. A layman is reading a text and looks up a word in order to understand its meaning in this specific text (= **reception**). There are two situations: A) The reader has seen or heard the word before, but is not quite sure what it means, or the reader thinks s/he can deduce the meaning from the con-

text, but s/he wants the meaning verified; B) The reader has never come across the word before and cannot deduce its meaning from the context.

Requires: A) A short explanation that verifies what the user thinks the word means. This should only be a short sentence and no details are needed.

B) A rather elaborate meaning explanation; the reader should be able to understand the meaning of the word in the given context. A general meaning explanation is not very useful — factual and cultural details are important.

3. A semi-expert is reading a text and looks up a word in order to understand its meaning in this specific text (= **reception**).

Requires: A much more detailed meaning explanation than the one provided for the layman. A semi-expert has specialised knowledge and is therefore interested in the finer details and is able to understand these.

4. A layman or semi-expert is writing a text and wants to check the inflection of a certain word so s/he uses a spelling and grammar dictionary (= **production**). This dictionary should contain meaning explanations in case of homonymy as the user needs to be able to distinguish between the homonyms if the dictionary applies grammatical homonymy.

Requires: Only a synonym if the person knows something about the relevant theme, but sometimes relevant synonyms do not exist; in the latter case, a simple meaning explanation will suffice, e.g. a short phrase.

5. A layman or semi-expert is writing a text and wants to use a synonym for a certain word so s/he uses a synonym dictionary (= **production**). There are two situations: A) The word is monosemous; B) The word is polysemous.

Requires: A) No explanation is needed.

B) A short sentence that makes it clear what the difference is between the two or more meanings of the word.

6. A layman (A) cannot remember or (B) does not know the word s/he is looking for when writing a text, but s/he knows its meaning (= **production**). For this purpose, a dictionary can be made in which the user can search for the word using Boolean operators and key search terms.

Requires: A meaning explanation that can help him/her decide on the right word (if it exists) in the result list.

(A) A short explanation is enough as the user already knows the word.

(B) The explanation should be as elaborate as for reception since the user is not familiar with this word.

7. A person is working on a translation and looks up an unknown word from the source text in a bilingual dictionary in order to find an equivalent (= **translation**). There are two situations: A) The person is not an advanced L2-speaker and does not know much about the relevant theme; B) The person is an advanced L2-speaker and knows a lot about the relevant theme:
Requires: A) A rather detailed meaning explanation so the translator knows exactly which equivalent to use if there are more than one equivalent. The explanation should be supplied in both L1 and L2.
B) Only equivalents should be supplied as the person has sufficient knowledge to distinguish between the equivalents without the help of meaning explanations.

8. A beginner is not sure how to do a certain physical exercise and therefore looks the word up in a dictionary in order to get an instruction on how to carry out this exercise (= **operation**).
Requires: The explanation should briefly state what kind of exercise this is, and then it should describe in detail, step by step, what the person should do. This explanation should also include variations of and common mistakes connected to this exercise.

9. A person is in a situation where s/he does not understand what a street sign means and therefore does not know how to act in the situation (= **interpretation**). Therefore, s/he looks the sign up in a dictionary. There are two ways of searching: A) using search terms; B) searching with a photo of the sign.
Requires: A detailed description of the non-linguistic sign as the user's search words will be matched with the words in the explanation. The explanation must also include a description of how the person should act according to this sign as this is what is needed to solve the problem in the situation. In addition to a written explanation, the article should also contain a photo or an illustration, as the user in situation B will make a search with a photo, hoping to find an image in the database that matches it.

This overview shows (1) that many different explanations can and should be made according to the dictionary function. If we look at Dictionaries 6 and 42, these two general dictionaries actually present different explanations — but they do not explain the reason for providing different explanations, nor is it optimal that different types of explanations are shown in the same article, see e.g. the lemma *aerobics* in both of these online dictionaries. Their distinction made between simple/primary and full definitions is also sometimes non-

existent, resulting in two similar explanations, e.g. *pull-up* in Dictionary 42. The overview also shows (2) that no distinction can or should be made between semantic and encyclopaedic knowledge in the selection and distribution of data — such a distinction should not be the factor determining the selection of semantic components. The lexicographer must always consider what types of data are relevant according to the dictionary function, not whether these are semantic or encyclopaedic. From this, we can extract the following function-based definition of the term lexicographical meaning explanation:

Lexicographical meaning explanation: *A description in the form of one or more words, a sound, a video and/or an image of a word, phrase, idiom, collocation or some other language construction or of a non-linguistic sign, which is presented in a dictionary article, and which has been formulated, drawn or recorded to help satisfy the need of a specific user type in a specific extra-lexicographical user situation. The selection of semantic components and their organisation in the explanation is also based on the user and user situation.*

If we look at the explanations above in the case of the interpretative and operative functions, these appear to be different than the other types of explanations. In the case of the operative function, the article of *press-up* in Dictionary 27 and the article of *crawl* in Dictionary 18 contain a short introductory explanation of the lemma and then a description of the moves. In the case of the interpretative function, Figure 2 shows that each signal is connected to an explanation (or a description) of the signal and an explanation of its meaning. Thus, we are actually working with three different types of overall lexicographical explanations:

1. Meaning explanation
2. Instructive explanation (or operative explanation)
3. Non-linguistic sign explanation (or interpretative explanation)

Thus, all major functions — communicative, cognitive, operative and interpretative — apply a certain type of meaning explanation (as shown above, there are many different types of meaning explanations), and in addition, the operative dictionary applies an instructive explanation and the interpretative dictionary sometimes incorporates a non-linguistic sign explanation. I.e. (8) and (9) should be revised:

- 8'. A beginner is not sure how to do a certain physical exercise and therefore looks the word up in a dictionary in order to get an instruction on how to carry out this exercise (= **operation**).

Requires: A short *meaning explanation* in the form of a sentence (not as elaborate as for reception) + *an instructive explanation*.

- 9'. A person is in a situation where he does not understand what a sign in the street means and therefore does not know how to act in the situation (= **interpretation**).

Requires: In some cases a *non-linguistic sign explanation* + a *meaning explanation* telling what the sign indicates and how one should react to it. Of course, the sign should also be represented non-verbally.

These definitions of lexicographical explanations are relevant both for general and specialised dictionaries as the basis for these concepts is not the content of a dictionary, but the function. In addition, this functional approach to meaning explanations is much more flexible and dynamic as sought after by Nielsen (2011), i.e. a meaning explanation should not be made as a one-size-fits-all, but according to the given function. A number of different explanations of the same lemma could thus be included in the same database and extracted for the different dictionaries produced from this database.

4. Concluding remarks

The study presented in this paper has highlighted the often occurring inadequacy of meaning explanations of terms both in general and specialised information tools. This was demonstrated by comparing the stated function of each dictionary to a number of selected explanations in the dictionary. Based on these findings, it is suggested how specialised words or multi-word constructions in future dictionaries should be treated. As the approach applied in this study has been deduction, a next step would be to test these suggestions on dictionary users.

Endnotes

1. In Denmark, there are today two competing online dictionaries of contemporary Danish: Den Danske Ordbog and Den Danske Netordbog. Because the author of this paper works as an editor on the latter, this dictionary has been discarded in the study.
2. Though throughout this paper the term *dictionary* is often applied, it could be replaced by the broader term *information tool*.
3. See Section 2.4.1 for further comments on this new user type termed *beginner*.
4. See Section 2.4.1 for further comments on this new user type termed *advanced*.
5. The operative function in a sports setting is currently being investigated by this author by filming different sports teams during training and fitness classes with the aim of identifying the information and questions exchanged between trainers/instructors and their members, converting this into data types. The same approach was followed by Patrick Leroyer (personal communication) in the production of OENOLEX Burgundy, a French monolingual online wine dictionary; see Leroyer and Høy (2013).

References

- Agerbo, H. 2015. How to Describe Sports Terms in Information Tools for Laymen. *Estudios de Lexicografía*: 137-160.
- Béjoint, H. 2015. Dictionaries for General Users. Durkin, P. (Ed.). 2015. *The Oxford Handbook of Lexicography*: 7-24. Oxford: Oxford University Press.
- Bergenholtz, H. and H. Agerbo. 2014a. Extraction, Selection and Distribution of Meaning Elements for Monolingual Information Tools. *Lexicographica* 30: 488-510.
- Bergenholtz, H. and H. Agerbo. 2014b. Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries. *Hermes. Journal of Language and Communication in Business* 52: 125-139.
- Bergenholtz, H. and I. Bergenholtz. 2011. A Dictionary is a Tool, a Good Dictionary is a Mono-functional Tool. Fuertes-Olivera, P. and H. Bergenholtz (Eds.) 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 187-207. London/New York: Continuum.
- Bergenholtz, H., T.J.D. Bothma and R.H. Gouws. 2015. Phases and Steps in the Access to Data in Information Tools. *Lexikos* 25: 1-30.
- Bergenholtz, H. and U. Kaufmann. 1996. Enzyklopädische Informationen in Wörterbüchern. Weber, N. (Ed.). 1996. *Semantik, Lexikographie und Computeranwendungen*: 167-180. Tübingen: Max Niemeyer.
- Bergenholtz, H. and U. Kaufmann. 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. *Hermes. Journal of Linguistics* 18: 91-125.
- Bergenholtz, H. and S. Tarp (Eds.). 1994. *Manual i fagleksikografi: udarbejdelse af fagordbøger — problemer og løsningsforslag*. Herning: Systime.
- Bergenholtz, H. and S. Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes. Journal of Language and Communication Studies* 31: 171-196.
- Diab, T. 1990. *Pedagogical Lexicography. A Case Study of Arab Nurses as Dictionary Users*. Tübingen: Max Niemeyer.
- De Schryver, G.-M. and D.J. Prinsloo. 2011. Do Dictionaries Define on the Level of Their Target Users? A Case Study for Three Dutch Dictionaries. *International Journal of Lexicography* 24(1): 5-28.
- Fuertes-Olivera, P.A. and S. Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter.
- Gouws, R.H. and S. Tarp. Forthcoming. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 29.
- Haiman, J. 1980. Dictionaries and Encyclopedias. *Lingua* 50(4): 329-357.
- Hashimzade, N., G.A. Myles and G.D. Myles. 2014. Can Authority be Sustained while Balancing Accessibility and Formality. *Hermes. Journal of Language and Business Communication* 52: 11-24.
- Leroyer, P. Forthcoming. Dictionaries for Text Reception. Fuertes-Olivera, P.A. (Ed.). Forthcoming. *The Routledge Handbook of Lexicography*. Oxfordshire: Routledge.
- Leroyer, P. and A. Høy. 2013. Vinsmagningsordbogen Oenolex Bourgogne. En milepæl i pragmatisk fagleksikografi. *Nordiske Studier i Leksikografi* 12: 287-302.
- Lew, R. 2013. Identifying, Ordering and Defining Senses. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 284-302. London: Bloomsbury.
- Lombard, F.J. 1991. Die aard en aanbieding van die leksikografiese definisie. *Lexikos* 1: 158-182.

- MacFarquhar, P.D. and J.C. Richards.** 1983. On Dictionaries and Definitions. *RELC Journal* 14(1): 111-124.
- Meyer, I. and K. Mackintosh.** 2000. When Terms Move into our Everyday Lives: An Overview of De-terminologization. *Terminology* 6(1): 111-138.
- Morris, W. (Ed.).** 1985. *The American Heritage Dictionary*. Boston: Houghton Mifflin.
- Nesi, H.** 2013. Researching Users and Uses of Dictionaries. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 62-74. London: Bloomsbury.
- Nielsen, S.** 1990. Lexicographical Macrostructures. *Hermes. Journal of Language and Communication in Business* 4: 49-66.
- Nielsen, S.** 2006. A Functional Approach to User Guides. *Dictionaries: Journal of the Dictionary Society of North America* 27: 1-20.
- Nielsen, S.** 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.
- Nielsen, S.** 2010. The Relevance of Lexicographic Functions. Aminova, A. and N. Fattakhova (Eds.) 2010. *Comparative Philology and Multilingualism. Materials of the International Scientific Conference*: 132-135. Kazan: Tatar Academy of Science.
- Nielsen, S.** 2011. Function- and User-related Definitions in Online Dictionaries. Kartashkova, F. (Ed.). 2011. *Ivanovskaya leksikograficheskaya shkola: traditsii i innovatsii*: 197-219. Ivanovo: Ivanovo State University.
- Nielsen, S.** 2013. The Future of Dictionaries, Dictionaries of the Future. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*. London: Bloomsbury.
- Nielsen, S. and R. Almind.** 2011. From Data to Dictionary. Fuertes-Olivera, P. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 141-167. London/New York: Continuum.
- Rull, A.N.** 2003. La terminología del deporte en los diccionarios generales del español. *Revista de Lexicografía* 9: 57-95.
- Stock, P.** 1988. The Structure and Function of Definitions. Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9-14 September 1986*: 81-90. Tübingen: Francke Verlag.
- Tarp, S.** 2008a. *Lexicography in the Borderland Between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Aarhus: Aarhus School of Business, Centre for Lexicography.
- Tarp, S.** 2008b. Revival of a Dusty Old Profession. *Hermes. Journal of Language and Communication Studies* 41: 175-188.
- Tarp, S.** 2008c. The Third Leg of Two-legged Lexicography. *Hermes. Journal of Language and Communication Studies* 40: 117-131.
- Van der Merwe, M.F. and P.A. Fuertes-Olivera.** 2014. The Influence of the User Needs Paradigm in Specialised Lexicography: Some Reflections in Connection with Two South African Wine Dictionaries. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos* 27: 77-96.
- Wierzbicka, A.** 1992. What are the Uses of Theoretical Lexicography? *Dictionaries: Journal of the Dictionary Society of North America* 14: 44-78.
- Wingate, U.** 2002. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. Tübingen: Max Niemeyer.

Addendum

English general dictionaries

1. Cambridge Advanced Learner's Dictionary online
2. Collins English Dictionary online
3. Collins COBUILD Advanced Dictionary of English online
4. Longman Dictionary of Contemporary English online
5. Macmillan Dictionary online
41. Merriam Webster Learner's Dictionary online
6. Merriam-Webster online
7. Webster's Third New International Dictionary paper, 1986
8. Oxford Learner's Dictionaries online
9. Oxford Dictionaries online
10. The American Heritage Dictionary paper, 1985
11. Harrap's Essential English Dictionary paper, 1995
42. Vocabulary.com online

Danish dictionaries: one sports dictionary, four general dictionaries

12. Dansk Sportsleksikon 1-2 paper, 1944–1945
13. Ordbog over det danske sprog online
14. Den Danske Ordbog 1-6 online + paper, 2004–2005
15. Politikens Nudansk ordbog med etymologi paper, 2010
16. Gyldendals dansk-dansk ordbog paper, 1992

English sports dictionaries

17. Encyclopedia of sports paper, 1944
18. The Dictionary of sports paper, 1949
19. Dictionary of Sports paper, 1961
20. The Oxford companion to sports & games paper, 1975
21. Webster's Sports dictionary paper, 1976
22. The Complete Sports Dictionary paper, 1979
23. The Macmillan dictionary of sport and games paper, 1980
24. (The facts on file) Dictionary of Fitness paper, 1984
25. What's What in Sports paper, 1984
26. A Dictionary of Sports Injuries and Disorders paper, 1996
27. Food and Fitness: Dictionary of Diet and Exercise online + paper, 1997
43. Sports: The complete visual reference paper, 2000
44. The encyclopedia of exercise, sport and health paper, 2004
28. Academic Dictionary of Sports paper, 2005
29. Dictionary of Sport and Exercise Science paper, 2006

-
- | | |
|--|------------------------------|
| 30. Dictionary of Sports and Games terminology | paper, 2010 |
| 31. A Dictionary of Sports Studies | online + paper,
2010/2016 |
| 32. Visual Dictionary (Merriam-Webster) | online |
| 33. Sportsdefinitions.com | online |

English subfield dictionaries

- | | |
|--|-------------|
| 34. The concise dictionary of tennis | paper, 1978 |
| 35. A dictionary of the Martial Arts | paper, 1991 |
| 36. Visual Dictionary of baseball | paper, 2001 |
| 37. Running Encyclopedia | paper, 2002 |
| 38. Now 110% complete football lexicon | paper, 2006 |
| 39. Cricket Lexicon | paper, 2006 |
| 40. Bendelow & Kidd's Dictionary of Football | paper, 2015 |

Etc. The Long-Lasting Defining Device: Unravelling the Mysteryⁱ

Safi Eldeen Alzi'abi, *English Department, Jerash University, Jerash, Jordan (alziabi@gmail.com)*

Abstract: In its abbreviated form, 'etc.' is a lexicographic device that dates back to the early 15th century. It is used on a large scale in monolingual dictionaries for native speakers and EFL learners to serve a wide range of linguistic patterns. Unfortunately, there seems to be little research on the way this linguistic unit has been used, despite the fact literature abounds with details about dictionary making.

This descriptive analytical study reveals the way 'etc.' is used in EFL learner's dictionaries and brings to light some unknown evidence regarding its frequency of occurrence. The bald statistics prepared on the use of 'etc.' in the macrostructure and microstructure of a cross-section of four learner's dictionaries show that it has been widely used, but in many cases it seems dispensable. Analysis of a large body of definitions shows that 'etc.' appeared in the definitions of headwords of all word classes and sometimes more than once in many senses of polysemous lexical items in the range of one to five lexemes preceding 'etc.'

This widespread use of 'etc.' may place additional pressure on dictionary users in their desperate attempt to comprehend the definitions. Maybe it is high time that 'etc.' is considered outworn or obsolete and is replaced with something that lends itself to this age of modern technology, where space is no longer a major concern for dictionary makers.

Keywords: LEARNER'S DICTIONARIES, DEFINING TECHNIQUES, STYLE GUIDES, PATTERNS OF 'ETC.', 'ETC.' IN ELECTRONIC DICTIONARIES, DEFINITIONS WITH 'ETC.', FREQUENCY OF OCCURRENCE OF 'ETC.', ITEMS BEFORE 'ETC.'

Opsomming: 'Etc.' Die lank bestaande definiëringshulpmiddel: Die ontrefeling van die misterie. In sy afgekorte vorm is 'etc.' 'n leksikografiese hulpmiddel wat terugdateer tot vroeg in die 15de eeu. Dit word op groot skaal in eentalige woordeboeke vir moedertaalsprekers en EVT-leerders gebruik om 'n wye verskeidenheid taalpatrone te dien. Ongelukkig blyk daar min navorsing te wees oor die manier waarop hierdie taalkundige eenheid gebruik is, ten spyte daarvan dat daar 'n oorvloed literatuur is met besonderhede van woordeboekmaak.

Hierdie beskrywende analitiese studie dui die manier waarop 'etc.' gebruik word in EVT-aanleerderwoordeboeke aan en lê onbekende bewyse bloot oor die frekwensie waarmee dit voorkom. Die onverwerkte statistiek rakende die gebruik van 'etc.' in die makro- en mikrostruktuur van 'n deursnee van vier aanleerderwoordeboeke toon dat dit wyd gebruik is, maar dat dit in baie gevalle weggelaat kan word. Die ontleding van 'n groot groep definisies toon dat 'etc.' voorgekom het in die definisies van trefwoorde van alle woordklasse en soms meer as een keer in baie betekenisse van polisemiese leksikale items in die bestek van een tot vyf lekseme wat 'etc.' voorafgaan het.

Hierdie algemene gebruik van 'etc.' mag bykomende druk plaas op woordeboekgebruikers in hulle desperate poging om die definisies te verstaan. Miskien is dit hoog tyd dat 'etc.' as ver-

ouderd en uitgedien beskou word en met iets vervang word wat homself leen tot hierdie tydperk van die moderne tegnologie, waar ruimte nie langer 'n groot oorweging is vir woordeboekmakers nie.

Sleutelwoorde: AANLEERDERSWOORDEBOEKE, DEFINIËRINGSTEGNIEKE, STYL-GIDSE, 'ETC.'-PATRONE, 'ETC.' IN ELEKTRONIESE WOORDEBOEKE, DEFINISIES MET 'ETC.', VOORKOMSFREKWENSIE VAN 'ETC.', ITEMS VOOR 'ETC.'

1. Introduction

Dictionary definitions illustrate meanings by describing the reference relations of lexical items (Jackson 2002: 70). The explicit accounts of the meaning of lexemes inform dictionary users about the way the lexemes are used in some contexts within a language community. But the process of defining lexemes is very delicate and requires consummate professional skill. Lexicographers are therefore under considerable pressure to develop elegant ways to define word; the definitions have to be both intelligible and easily accessible in order not to mislead the user.

To describe and disambiguate the variant senses of headwords, lexicographers use several defining styles (see Geeraerts 2003 and Osselton 2007). The one most commonly used is the 'analytical' or 'descriptive' style (Ayto 1983: 90 and Zgusta 1971: 258). Generally, it uses a 'genus', the superordinate concept, the broad semantic category to which the lexeme belongs, and the 'differentia', the characteristics and the additional distinctive features typical of the genus; the format is often an incomplete sentence (Moerdijk 2003). The *Oxford Advanced Learner's Dictionary*, 8th edition, (OALD), for example, defines a *hydrangea* as "a bush with white, pink or blue flowers that grow closely together in the shape of a large ball". Here, *bush* is the genus and the differentia, "with white, pink or blue flowers", post-modifies it and distinguishes a *hydrangea* from a(n) 'broom', 'blackthorn' or 'azalea'. Similarly, *bush* is defined by another genus, *plant*, and it is differentiated from other members of the plant category by such distinguishing features as "growing quickly with several hard stems".

The analytical style is appropriate for defining a great many lexemes of most parts of speech, mainly concrete ones. However, lexicographers sometimes face problems with some lexemes such as abstract items (cf. Dziemianko and Lew 2006), which are not readily defined analytically. They therefore resort to defining by means of synonyms (see Van der Meer 2004). This is characterised by the inclusion of a single synonym, a set of synonyms or a synonymous phrase. Consider the definitions of *worsen* in *Longman Dictionary of Contemporary English*, 6th edition, (LDOCE) and OALD, respectively.

worsen: to become worse or make something worse

worsen: to become or make sth worse than before

A third defining technique delineates the function or use of the lexeme and this fits 'grammatical' words such as auxiliary verbs, conjunctions, determiners,

prepositions and pronouns (see Atkins and Rundell 2008 and Jackson 2002). The following definition from *Cambridge Advanced Learner's Dictionary*, third edition, (CALD) indicates the function of 'and':

and: used to join two words, phrases, parts of sentences or related statements together ...

Other defining styles are in use (see Lew 2013; Lew and Dziemianko 2006a, 2006b and 2012 and Rundell 2006) and each proves an effective defining strategy for a particular set of lexemes. Generally, all defining techniques seem to commonly make use of some devices to sensibly describe the different meanings of lexemes including parentheses, a largely used style in earlier editions of learner's dictionaries, and 'etc.' Hanks (1987: 116) asserts that "Over the years a number of conventions have grown up governing the ways in which dictionaries traditionally explain meaning ... For example *of, or pertaining to, any of, various, etc., and esp.* are extremely common in dictionaries ...".

According to Atkins and Rundell (2008: 437), one prime function of parentheses was "to indicate a word's 'selectional restrictions' (a verb's usual range of subjects or objects, or an adjective's typical complements) ...". Consider the examples below.

bijou: (of a building or a garden) small but attractive and fashionable **OALD**

break: to (cause something to) separate suddenly or violently into two or more pieces, or to (cause something to) stop working by being damaged **CALD**

In the first example, the parentheses specify the normal range of nouns modified by the adjective. These suggest that *bijou* is used with nouns such as 'house', 'shop', 'restaurant', 'apartment', 'cafes', 'bars', which are hyponyms of the superordinate 'building', and so on. In the second example, they indicate that *break* is an ergative verb in this sense. If the words in parentheses are suppressed, *break* is intransitive but if the parentheses themselves were removed, it would be a transitive verb (see also Atkins and Rundell 2008). Apparently, there may be a wide range of subjects or objects which occur with a certain sense of any verb to the extent that lexicographers cannot account for the full contextual variability in the definition. They therefore tend to list some of them and use 'etc.' to stand for the remainder, mostly under the pretext of being economical.

The linguistic unit 'etc.' is used extensively in monolingual dictionaries both in the definitions and examples. It is the intention of this study to reveal the way 'etc.' is used in EFL learner's dictionaries and provide some reliable statistics on the frequency of this usageⁱⁱ. The focus here is on dictionaries for learners because native speakers may find no difficulty handling definitions with 'etc.' and are able to infer the sense from the list preceding 'etc.', while learners may struggle to make the necessary connections. The study also outlines the different patterns 'etc.' represents in the definitions. Then light is shed on some relevant definitions of 'etc.' to discuss some of the reasons underlying

the inclusion of this device. The study proper then follows, and the final section considers some pedagogical issues and recommendations.

2. What 'etc.' means

Etymologically speaking, this is a Latin term which is made up of *et*, "and", and *cetera*, "the rest", or more literally, "the other things". It is commonly abbreviated to 'etc.' but according to Peters (2004), the full spelling, 'et cetera', though rarely used, would also do. Let us see what dictionaries tell us about 'etc.'

Almost all dictionaries demonstrate that 'etc.' is primarily used at the end of a list to indicate that there are more things, objects, people, events, situations or items which could have been mentioned but were omitted for brevity. Along the same lines, LDOCE states that 'etc.' is used when the writer wants to tell the reader that there are many other examples or things of the same kind or class. This is rather misleading because the items left in the list are not always of the same kind or semantic class. The above definition seems a common practice in all Longman lexicographic products. The *Longman Language Activator*, for instance, indicates that 'etc.' means "and others of a similar kind ...". Although this is true on the face of it, in many cases it is still at odds with what exists in dictionaries, where in reality the remaining items are not always of a similar kind (see Alzi'abi 1995). One advantage of the *Activator's* definition is that it justifies the reasons dictionary makers substitute 'etc.' for the remaining items on the list.

Recently, dictionaries seem to slightly modify the definition of 'etc.' but this has been in the direction of more ambiguity. Consider the following two definitions by Collins Cobuild dictionary products (COBUILD) (see references).

- etc:** ... indicates that there are other items, events or situations which you could mention if you had the time and space ... (1987)
- etc:** is used at the end of a list to indicate that you have mentioned only some of the items involved and have not given a full list (2012)

Undeniably, the first definition is much clearer than the second; the use of 'event' and 'situation' shows the range of cases 'etc.' may cover rather than leaving them open to deduction.

3. Why use 'etc.'

It is logical to ask whether lexicographers implement any coherent policy in their application of 'etc.' The dictionary definitions themselves provide the answer to this question. A full examination of some entries shows that no conscious policies have been established regarding the inclusion of 'etc.' (see also Alzi'abi 1995 and 2016). Lexicographers need to speak for themselves as to whether they have any definite underlying mechanism for using this device.

One plausible reason, particularly in paper dictionaries, is economy of

space. Truly, the long list of lexemes covered by 'etc.' is space-consuming but the use of 'etc.' exacerbates users' problem in inferring the lexemes which fill this gap. This is not the sole reason, however. Having scrutinised a large number of definitions, one can discern that many of those containing 'etc.' were not written with the above reason in mind; rather the items covered by 'etc.' were simply too hard to provide or possibly the list of possibilities is too exhaustive and varied to allow for feasible inclusion (see examples in 4.3). In other words, the editors might find it hard to think of more items. This of course has been implicitly suggested in some definitions of 'etc.' Sometimes, it is possible that some writers use it because they know not of any more items to add. It is really a serious matter which could unduly affect the whole look-up process. This being the case, dictionary users will be completely frustrated and it undoubtedly adversely affects their comprehension and production skills. Some definitions with 'etc.' require skilled semanticists to provide other items to replace 'etc.' (see for example the definition of *zest* in LDOCE and OALD). One writerⁱⁱⁱ humorously remarks that 'etc.' is employed when writers run out of items and cannot think of anything else to add to the named ones rather than limiting the range of items provided.

One important issue deduced from the definitions of 'etc.' is that it is used when a fuller list of items is considered to be unnecessary. This at times may be true and many examples exist to substantiate this claim (see 4.3). The question however is: How can editors tell whether the fuller list is (un)necessary, specifically when the items covered by 'etc.' are collocates of another item in the definition? A really vexing matter here is that in many a case 'etc.' stands for no more than two other words, which could be easily included. A few more pages might be tolerable when the resulting definitions are clearer and unambiguous.

A noteworthy point here is that some definitions of 'etc.' indicate that the items before 'etc.' belong to the same semantic field or possibly have similar semantic features. On examining some random entries, it is obvious that some definitions have genus words or superordinates. CALD defines *articulation* as "the way in which you express your feelings and ideas, etc.", where the words 'feelings' and 'ideas', superordinates, belong to the same semantic field, under which many hyponyms can be subsumed. Likewise, OALD definitions of *cannibalise* as "to take the parts of a machine, vehicle, etc. and use them to repair or build another" and *commandeer* as "to take control of a building, a vehicle, etc. for military purposes during a war, or by force for your own use", use the genus or superordinate word 'vehicle', under which several hyponyms can be subsumed. Assumedly, editors expect users to provide superordinate items and co-hyponyms of these items unaided (Alzi'abi 2016). This claim has been asserted by Michael Rundell (p.c.), editor-in-chief of *Macmillan Advanced Learner's Dictionary* (MEDAL 2nd edition). Their Style Guide recommends using 'etc.' only sparingly; that is to say, when the editors expect users to be capable of working out from their own real-world knowledge the exact words which would replace 'etc.' Unfortunately, not all editors strictly adhere to such

guidelines (see below). Other Style Guides might have similar guidance but unfortunately these were virtually inaccessible due to being commercially sensitive materials.

It may be true that dictionary makers use 'etc.' to represent some items when the criterion of classifying these items under a genus proved impossible. Alzi'abi (2016) argues that only a small number of definitions seem to conform to the above image of using 'etc.'; many are composite structures of both genus words and single items. When all the items belong to the same semantic field, dictionary users are less likely to be confused because they think of items along the lines of those included and vice versa (consider the definitions of the noun *access* and the verb *sabotage* in LDOCE and *segue* in OALD).

4. The study

4.1 Aim

The present study addresses the following issues:

1. To explore the way 'etc.' is used in EFL learner's dictionaries and identify the various functions it serves.
2. To estimate the number of headwords generally defined by means of 'etc.' in these dictionaries.
3. To compute the total occurrences of 'etc.' in the definitions of individual headwords as these differ from the above, where some definitions contain more than one 'etc.'
4. To check the average number of items which usually precede 'etc.' in the definitions.

4.2 Method

To be able to look at the way 'etc.' is employed in learner's dictionaries and identify the different patterns and structures for which 'etc.' stands, it was necessary to intensely scrutinise and carefully analyse the definitions, specifically the lexemes supplied by dictionary makers before 'etc.' Four dictionaries on DVD-ROM were closely examined: CALD, LDOCE, MEDAL and OALD. The fifth big learner's dictionary, COBUILD, was excluded as it did not apply 'etc.' in its definitions. Earlier versions, particularly the ones in print, used it very sparingly but since 1995 COBUILD contained none. This may reflect COBUILD editors' policy of frequently abandoning traditional techniques of lexicography (see Barnbrook and Sinclair 1995; Hausmann and Gorbahn 1989; Moon 2007 and Sinclair 1987).

As to 'etc.'-patterns, a rigorous examination of all entries containing this linguistic unit revealed that it had served several grammatical patterns and structures, including verbs followed by a list of adverbs or nouns plus 'etc.', or

possibly adverbs followed by a number of adjectives plus 'etc.' and so on. However, only those commonly used functions in all four dictionaries were listed along with examples. Some other structures were also identified in the corpus but these were ignored as they prevailed only in one, two or three dictionaries. It is highly likely that some other patterns might have emerged had the entries in the remaining letter stretches been examined. A noteworthy point is that the examples cited below were random selections from the ones extracted. Moreover, for brevity, only examples from LDOCE are used, unless otherwise stated.

To give an approximate estimate of the use of 'etc.' all the headwords which utilise it in the definitions were identified using the Advanced Search facilities available, which at times offered some invaluable help. This task was not so easy in LDOCE. Exact figures could not be reached because its Advanced Search was difficult to use and the queries results were awkward to handle. (The figures obtained were therefore prone to errors and should be interpreted with this in mind). A calculation was made of all headwords whose definitions contained 'etc.', which of course did not give the actual occurrences of 'etc.'

Having come up with a basic idea about the number of headwords defined by means of 'etc.' in the four dictionaries, it has been tedious to work out the precise number of times — hundreds, perhaps thousands — 'etc.' has been used in all entries. To get a rough estimate for the number of times 'etc.' occurred in any dictionary, the entries of only two sections of the wordlist were considered as temporarily representative samples. The headwords beginning with the letters 'a' and 's' were randomly chosen. They however seemed to be among the largest sections of the wordlist in any English language dictionary, constituting approximately 20% of the dictionary's average size.

In order to collect evidence regarding the double or treble appearance of 'etc.' in definitions, a thorough examination was carried out of the definitions of the various senses of the headword in the two letter stretches. The figures of the exact number of definitions containing more than one sense defined by means of 'etc.' and the exact number of 'etc.'-occurrences were all computed. Comparing and contrasting these figures help determine which dictionary made much use of 'etc.' in any of the above cases.

To draw a comprehensive picture of the content of the 'etc.' segment in the definitions, i.e. the number of item(s) preceding 'etc.', the 3768 occurrences of 'etc.' in total (1007 in the 'a' section and 2761 in 's' section) were extracted, broken down and further lemmatised and manually tagged to get the total cases where 'etc.' was preceded by one, two, three, four and even five items. This helped establish the total number of each of these categories.

Before moving on to the next section, it is stressed the figures reported below do not involve the cases 'etc.' was used as an extension to lists in the examples and illustrative sentences.

4.3 Results and discussion

Below is an account of the results of a quantitative analysis of dictionary definitions to find out more about the use of 'etc.'

4.3.1 *Etc.* in dictionaries

The linguistic unit 'etc.' seems to be used similarly in all dictionaries. One may think that it is used in the definitions of the highly-specialised or technical vocabulary, where possibly specialist users may know all the words covered by 'etc.' A cursory glance at any of the dictionaries reveals that 'etc.' is employed in the definitions of all types of vocabulary items of all parts of speech and sometimes used more than once in the same definition. Although dictionaries vary tremendously as to the number of lexemes whose definitions contain 'etc.', this device seems to occur at a rather alarming frequency (see Tables 1 and 2 below). A quick browse through paper dictionaries reveals that the average use of 'etc.' ranges between three and seven occurrences per page but it is not easy to explain the disparity which exists between dictionaries for native and non-native speakers as to the number of times they apply 'etc.' in the definitions. Dictionaries for native speakers use 'etc.' on a larger scale. Randomly picked out, *Chambers Pocket Dictionary* (1992) showed that page 777 (reverse→revive), randomly selected, had 20 occurrences of 'etc.', a quite large figure compared to those in learner's dictionaries. Most likely, dictionary compilers assume that native speakers are more able to infer the remaining possibilities covered by 'etc.' because of their own broader linguistic knowledge.

This device, i.e. 'etc.' is commonly used to represent a list of noun collocates (both in the subject and object positions), adverb collocates of verbs or adjective collocates of nouns, e.g.^{iv}

snowballs: if a problem, a plan, an activity, etc .snowballs, it quickly becomes much bigger, more serious, more important, etc

Moreover, it is applied in exemplification cases following expressions like "such as" and "for example". Consider these examples.

shopfitting: the process of preparing the inside of a shop by putting in equipment such as lights, shelves, etc. **MEDAL**

short-circuit: to prevent sth such as a career, plan, process, etc. from being successful
MEDAL

slapstick: the type of humour that is based on simple actions, for example people hitting each other, falling down, etc. **OALD**

string sb along: to allow sb to believe sth that is not true, for example that you love them, intend to help them, etc. **OALD**

4.3.1.1 Patterns of 'etc.' use

A rigorous examination of entries containing 'etc.' revealed that 'it fulfilled

several grammatical patterns and structures. Categorised below are the most commonly used functions and patterns along with illustrative examples. The bracketed plural suffix is added to the word classes for which 'etc.' stands.

It is worth reiterating that the patterns delineated below are those spotted most frequently in the two letter stretches under consideration. It is highly likely that other patterns would have emerged if the entries of the remaining sections had been examined. The examples cited are random selections from the corpus and, as indicated above, are all derived from LDOCE.

4.3.1.1.1 Verbs/verb phrases

Here 'etc.' stands for verbs which might be used with nouns or adverbs. In addition, it stands for some verb phrases or possibly whole clauses as shown below.

V(s)+Adv(s) or Adv(s)+V: The linguistic unit 'etc.' stands for a number of adverbs which qualify or modify the verb, either preceding the verb or following it, or in a number of cases, for verbs which occur with a particular adverb.

admit: ... freely, openly, frankly, etc. admit

V+NP(s): This pattern has 'etc.' to represent a list of nouns which are the 'objects' of the defined verb, particularly when the verb is transitive.

superimpose: combine two systems, ideas, opinions, etc.

V(s)+PP(s): In this pattern, 'etc.' stands for prepositional phrases following particular verbs or possibly a list of verbs which are followed by prepositional phrases.

accede: to agree to a demand, proposal, etc.

Linking Verb+Adj(s): Here 'etc.' covers a list of adjectives used predicatively, adjectives following intensive or linking verbs.

seem: ... seem important/right/strange, etc. to somebody

Verb phrases: In such structures, 'etc.' accounts for some verbs or verb phrases. Sometimes, it covers a whole clause.

self-taught: having learned a skill or subject by reading about it, practising it, etc.

4.3.1.1.2 Nouns

Here 'etc.' stands mostly for a number of nouns which may be used with particular verbs or adjectives.

N(s)+V: These patterns have 'etc.' to stand for a list of nouns serving as the subject of the following verb, mostly when the verb is used intransitively.

abound: if a place, situation, etc. abounds with things ...

N+Prep+N(s): The linguistic unit 'etc.' is employed to represent some nouns in a prepositional phrase following a certain noun, mostly of the form N+of+N.

aspect: one part of a situation, idea, plan, etc. that has many parts

List of nouns/numbers: In this pattern 'etc.' covers a number of nouns on the list usually preceded by some signal words like "such as", "for example" and "for instance".

acute: acute senses such as hearing, taste, touch, etc. // **strike** if a clock strikes one, two, six etc., its bell makes a sound once, twice, six times, etc.

4.3.1.1.3 Adjectives/adjective phrases

Here 'etc.' stands for a number of adjectives which either modify some nouns or are modified by some adverbs. The most common pattern is adj+noun(s), where 'etc.' represents nouns qualified by adjectives.

Adj(s)+N(s): This pattern has 'etc.' to represent some nouns modified by a particular adjective or some adjectives which modify particular nouns.

subtle: ... subtle taste/flavour/smell, etc.

Adj(s)+Prep+N(s): In such a structure 'etc.' covers some nouns which are objects of prepositions following particular adjectives or some adjectives followed by prepositional phrases.

shadow: ... afraid/frightened/scared, etc. of your own shadow

Adj(s)+Prep+V(s): The linguistic unit 'etc.' here represents some adjectives which are followed by a preposition, mostly 'to' or 'for', plus verbs or verbs which usually follow some adjectives.

suitable: ... suitable to use/be shown, etc. // **safe:** safe to use/drink/eat, etc.

List of adjectives: This pattern has 'etc.' to stand for a number of adjectives used in a particular context and are used either attributively, i.e. before a noun or predicatively.

superior: better, more powerful, more effective, etc. than a similar person

In addition, 'etc.' in some structures represents some adjectives in similes, usually preceded by 'as' but this was more prevalent in sections other than the two under scrutiny. LDOCE, for example, has instances such as (**anything 7:** as

important/clear/big, etc. as anything *informal* extremely important, clear etc.). OALD presents (**seraphic**: as beautiful, pure, etc. as an angel and **anything**: as happy, quick, etc. as anything [informal] very happy, quick, etc.). Similarly MEDAL has (**subhuman**: not as intelligent, kind, etc. as you think humans should be).

4.3.1.1.4 Adverbs/adverb phrases

In these patterns, 'etc.' stands for a number of adverbs which may modify some adjectives, i.e. Adv(s)+Adj(s).

aware: politically/socially/environmentally, etc. aware

Analysis of the definitions has shown other patterns but these were not spotted in all dictionaries.

4.3.1.1.5 Prepositional phrases

In some cases, 'etc.' represents a list of prepositional phrases or nouns, adjectives and gerunds as propositional objects as well as prepositions or particles.

abstract: containing the most important ideas or points from a speech, article, etc.

4.3.1.1.6 Wh-Words

Here, 'etc.' represents a list of 'wh-words' following some verbs or used in noun phrases, occasionally in exclamatory sentences. However, these have not been traced in the entries within the two-letter stretches under scrutiny, viz. 'a' and 's' in some dictionaries. Many examples of such cases however existed in the remaining sections of the wordlist, i.e. other letter stretches.

ascertain: ascertain whether, what, how, etc; **assess**: assess what, how, etc.; **say**: say who, what, how, etc.; **see**: ... see why, what, how, etc.; **suggest**: suggest how, where, what, etc.

4.3.1.1.7 Idioms

Here 'etc.' is used in the definition of idioms and serves to represent several similar structures to the ones outlined above including verb phrases, noun phrases, adjective phrases and so on. No examples were cited in OALD, nonetheless.

VPs- shit: ... beat/kick, etc. the shit out of

NPs- shake: ... shake sb's confidence/beliefs, etc.

APs- shift: be on the late/early/night, etc. shift

4.3.2 Number of items defined by 'etc.'

CALD has some 2150 lexical items whose explanations have 'etc.' These cover 1877 definitions, 200 phrases and 73 idioms. Filtering the queries further has shown the definitions belong to 939 nouns, 339 verbs, 236 adjectives, 16 adverbs and 8 prepositions. LDOCE comprises over 3500 nouns, 2000 adjectives, 1800 verbs, 250 adverbs, 60 prepositions and 60 pronouns whose explanations have 'etc.' Some of the occurrences of 'etc.' have been spotted in the definitions of the synonyms of the headword under THESAURUS. MEDAL has some 2042 items defined by means of 'etc.' including 237 phrases and idioms; they cover 808 nouns, 274 verbs, 300 adjectives, 51 adverbs, 24 prepositions, 18 pronouns, etc. As to OALD, the search results have been hard to tackle. The Oxford Dictionary Department (p.c.) report that OALD 9th edition has 7195 occurrences of 'etc.' in 5389 headwords. The figures suggest that multiple uses of 'etc.' occur in some entries. Additional queries have revealed that about 30% of these items were nouns and a slightly smaller proportion was for verbs whereas adjectives amounted to about 12%. Interestingly, matching entries from OALD 8th edition with those in the 9th edition has revealed additional headwords with 'etc.' in OALD 9 are mostly derivatives of some headwords (e.g. secretiveness, securely, simplistically, etc.). These are usually appended to the base forms in the 8th edition without any explanation. Generally, the figures above demonstrate beyond doubt that MEDAL relies the least on 'etc.' and LDOCE the most. Interestingly, MEDAL presents definitions with closed lists but the use of phrases such as "for example" and "for instance" saves the definitions from being over-restricted.

One might be really surprised at the large number of senses defined by means of 'etc.' in learner's dictionaries. The gravity of these instances can best be realised by knowing the total number of headwords in the dictionaries under scrutiny, which is not an easy task. The dictionaries advertise they have included a large number of words, phrases and meanings amounting to (230,000) in LDOCE and (185,000) in OALD compared to far fewer ones in CALD (68.119), according to the Cambridge Dictionaries Online Team (p.c.), and between 45,000 and 46,000 headwords in MEDAL, as reported by Michael Rundell (p.c.).

Table 1: Headwords defined with 'etc.' and actual occurrences of 'etc.' in the definitions

		CALD	LDOCE	MEDAL	OALD ^v
Letter 'a'	headwords defined with 'etc.'	73	291	132	220
	occurrences of 'etc.' in definitions	84	455	207	261
Letter 's'	headwords defined with 'etc.'	234	624	196	723
	occurrences of 'etc.' in definitions	254	1415	230	862

A thorough examination of entries in the two letter stretches in all dictionaries has shown the approximate figures (table 1) for the number of headwords whose definitions contained 'etc.' and the number of times 'etc.' arose in the definitions of the various senses of the items identified.

LDOCE and OALD included far more definitions with 'etc.' than the other two dictionaries, i.e. CALD and MEDAL in these particular letter stretches. Obviously, this is due to the comparatively larger number of headwords in these two letter stretches in both dictionaries. Knowing the approximate number of headwords in the two sections would display a clearer picture of 'etc.' usage in the four dictionaries. One possibly needs to have an idea about the ratio of the entries with 'etc.' to the total number of entries in the two sections. Table 2 below displays the total headwords in these letter stretches in all dictionaries^{vi}, which enables us to reckon the ratio of definitions with 'etc.' to the headword list.

Table 2: No. of headwords in the two-letter stretches in all four dictionaries

	CALD	LDOCE	MEDAL	OALD
Letter 'a'	4300	4000	3000	3500
Letter 's'	6500	9000	6464	7000

LDOCE has the largest number of headwords, followed by CALD and OALD comes third. It has to be noted however that a good number of the headwords in CALD are compounds which also appear in other sections of the wordlist. Evidently, the data in table 1 and 2 show that MEDAL has the smallest number of headwords and it uses 'etc.' much less frequently than the remaining dictionaries; this confirms its compilers' policy to keep the use of 'etc.' to a minimum. Generally, the rate of the use of 'etc.' ranges from about 5% to 10% of the total number of headwords in the four dictionaries.

4.3.3 Frequency of occurrence of 'etc.' in individual entries

Mostly, 'etc.' is used once in the definitions but sometimes it occurs more than once. Occasionally, more than one sense of the same headword is defined by means of 'etc.'; in such cases, it appears in two, three up to twenty senses. This illustrates how the number of headwords whose definitions contain 'etc.' is different from that of the total occurrences of 'etc.' in the definitions themselves (see table 1). To expound, here is an example where 'etc.' is used twice in the same definition.

abreast: walk/ride, etc. abreast to walk, ride, etc. next to each other, all facing the same way- two/three/four, etc. abreast ...

The statistics in table 3 below conclusively shows that LDOCE tends to use 'etc.' in one definition twice as much as all the other dictionaries; OALD comes second. Fortunately, not too many cases exist.

Table 3: Recurrence of 'etc.' in the same definition

	CALD	LDOCE	MEDAL	OALD
Letter 'a'	6	45	13	12
Letter 's'	17	117	7	46

The double use of 'etc.' in certain definitions is one of two cases. First, 'etc.' is duplicated with same preceding items as illustrated below.

- spanking:** when you say that sth is spanking new, etc. you are emphasizing that it is very new, etc. **OALD**
- synthesis:** the act of combining separate ideas, beliefs, styles, etc.; a mixture or combination of ideas, beliefs, styles, etc. **OALD**

The second occurrence of 'etc.' causes no extra trouble; 'etc.' is understood to refer to the same items in both positions. Second, 'etc.' is repeated in a different context in the remainder of the definition as illustrated below.

- sight:** a device that you look through to aim a gun, etc. or to look at sth through a telescope, etc. **OALD**
- syndicate:** to sell an article, a photograph, a television program, etc. to several different newspapers, etc. **OALD**

As indicated above, 'etc.' may appear more than twice in the same definition. The following definitions each has three occurrences of 'etc.':

- adrift:** two points/five seconds, etc. adrift (of somebody) two points, five seconds, etc. behind someone in a competition, race, etc.
- send:** **RADIO/COMPUTER, ETC.** to make a message, electronic signal, etc. go somewhere, using radio equipment, computers, etc.

Likewise, 'etc.' appears once or twice in the definitions of more than one sense of particular headwords. Examples:

- adopt ... 2 adopt an approach/policy/attitude, etc.** to start to deal with or think about something in a particular way **4** to formally approve a proposal, **AMENDMENT**, etc. ... **5** to choose a new name, country, custom, etc., ...

There are more and more instances of 'etc.' occurring in the explanations of more than one sense of the headword. LDOCE for example has many cases where three, four, five, six and more senses of the same headword are defined by means of 'etc.' To illustrate, three senses of each of *abstract* (n), *adopt*, *adrift*, *apart*, *appear*, *seed*, *self*, *sequence*, *seal*, *seek*, *sexism*, *sheer*, *shift* (v), *superficial* amongst others are defined by means of 'etc.' The words *art*, *another*, *ask*, *scramble*, *scrape*, *sail*, *seam*, *supply* and others have four senses each defined with 'etc.' Five senses of each of *at*, *association*, *sell*, *success* and *sum* are defined with 'etc.' The words *sad*, *scope* and *section* have some six senses each defined with 'etc.' whereas seven senses of each of *set*, *second* and *scope* are defined with 'etc.' The verbs

see and *shake* have some ten senses defined by means of 'etc.' whereas 12 senses of each of *say* and *season* and 13 senses of *send* have been defined by means of 'etc.'

Table 4 lists the number of headwords whose two or more of their senses contained 'etc.' The figure is not very high in any of the four dictionaries. It should be clarified once again the figures below represent the number of headwords in which two or more of their senses have 'etc.' rather than the actual occurrences of 'etc.'; in some headwords, 'etc.' might have occurred in twenty senses and might sometimes have appeared twice in one sense. Clearly, LDOCE includes the largest number of cases; OALD comes second and CALD last, at least in the two letter stretches under examination.

Table 4: No. of headwords whose 2+ of their senses contain 'etc.'

	CALD	LDOCE	MEDAL	OALD
Letter 'a'	0	61	19	24
Letter 's'	4	203	18	104

However, the four dictionaries seldom coincide in their use of 'etc.' in the same lexical item. There are cases where one lexeme appears with 'etc.' in one dictionary, e.g. *arriviste* in OALD; two dictionaries, e.g. *assault* in LDOCE and OALD; three dictionaries, e.g. *aspect* in CALD, LDOCE and OALD and in all four dictionaries, e.g. *armhole* and *author* (v). Sometimes these dictionaries use 'etc.' with the same headword but each with a different sense or part of speech as in the case of *scene* in LDOCE, MEDAL and OALD.

Table 5 below helps us establish an accurate picture of the degree to which the four dictionaries tally in their use of 'etc.' in the same headwords, by having all the headwords cross-tabulated.

Table 5: No. of 'etc.' concurrences which tally in all four dictionaries*

	Letter 'a'				Letter 's'			
	CALD	LDOCE	MEDAL	OALD	CALD	LDOCE	MEDAL	OALD
CALD	X	16	8	22	X	52	22	96
LDOCE	X	X	30	90	X	X	59	248
MEDAL	X	X	X	47	X	X	X	86
	74	291	132	220	234	624	196	723

*Figures in the last row are the number of headwords containing 'etc.' (see table 1)

The figures above reveal an obvious glaring disparity within lexicographers' policies on using 'etc.' The dictionaries are so utterly disparate that they concur only in very few cases. The two dictionaries which coincide most often are LDOCE and OALD, which is possibly attributable to the fact they contain more definitions with 'etc.' than the other two. The dictionaries which coincide the

least are CALD and MEDAL. They agree only in cases common to all four dictionaries. The 22 headwords whose definitions contain 'etc.' in all dictionaries are: *absorb, adherence, ambition, armhole, at, audience and audition* under 'a' and *scope, score, send, set, show, side, sideboard, sign, smoke, something, send, spoiler, star, support and sweep* under 's'. Not surprisingly, many of these exist in the *New General Service List* of the most frequent 2000 words^{vii} (see Browne 2014). It goes without saying, the senses of the above lexemes which are defined with 'etc.' may vary from one dictionary to another and this holds true for their word classes. This again supports the contention that most lexical items can be succinctly defined without 'etc.'

Some may believe that the above is a slightly spurious argument since the comparison involved here is of headwords which had been defined using different styles and defining techniques. A more valid comparison is possibly to be drawn between dictionaries using the same defining style with the same headwords, an issue which is impossible to settle at least for the time being. One cannot actually get all dictionaries to use the same defining techniques for the same headwords as it is one of the ways that they define their competitive advantage.

4.3.4 Number of items before 'etc.'

Analysis has shown that the items preceding 'etc.' range from one to four. But in very limited cases, there existed five or even six items, mainly in LDOCE and MEDAL. Table 6 below lists the frequency occurrence of the items before 'etc.'

Table 6: No. of cases for each item-type before 'etc.'

		1 item	2 items	3 items	4 items	5 items	6 items
Letter 'a'	CALD	5	38	35	6	0	0
	LDOCE	20	253	172	8	2	0
	MEDAL	1	107	91	8	0	1
	OALD	50	149	59	3	0	0
Letter 's'	CALD	24	149	72	9	0	0
	LDOCE	70	801	506	38	0	0
	MEDAL	1	112	111	5	1	0
	OALD	196	496	157	13	0	0

The figures vary in all four dictionaries. The largest figure is for the two-item type, 56% of cases, i.e. 2105 in all four dictionaries, followed by the three-item type which constituted about 32% of cases, i.e. 1203 occurrences. OALD seems to use one item before 'etc.' more often than all other dictionaries. The examples below show instances of one, two, three and four items before 'etc.'

(a) **One item**

angry: [literary] an angry wound, etc. is painful and red and looks infected

- (b) **Two items**
abrogate: to officially end a legal agreement, practice, etc.
- (c) **Three items**
above: be above suspicion/reproach/criticism, etc.
- (d) **Four items**
accompaniment: piano, orchestra, organ, guitar, etc. accompaniment

Quite a number of the one-item cases before 'etc.' in LDOCE arise when it duplicates in the remainder of the definition (**alternate:** if something happens on alternate days, weeks, etc., it happens on one day, etc. and ...). As indicated above, some instances with five items before 'etc.' occurred in LDOCE as follows.

art: the arts [plural] art, music, theatre, film, literature, etc. all considered together

Similarly MEDAL, though not very often, provided some definitions with five items before 'etc.':

set: to put someone or something in a position :set someone/something in/into/on/down/back, etc.

Besides, it contained definitions with six items preceding 'etc.' as in the following definition of the noun 'act':

act: [countable] a single thing that someone does: act of violence/vandalism/terrorism/kindness/bravery/desperation, etc.

5. Pedagogical implications

From the above findings, one can claim that the use of 'etc.' is categorised into two types. The first type conforms to what has been suggested in the MEDAL Style Guide. That is to say, 'etc.' can be used when the items which substitute for it are easily deduced by users. For example, at *accommodate* (if a room, building, etc. accommodates a particular number of people ...), the remaining items seem to be easy to supply as they belong to the semantic field 'building', regardless of whether these are residential, educational, industrial, military or parking structures. Likewise, the definition of *dishwasher* in CALD, MEDAL and OALD as (a machine for washing plates, cups, etc) represents an acceptable use of 'etc.', since a user may easily tell what other equipment can be put in the dishwasher, e.g. 'dishes', 'glasses', 'forks', 'knives', 'pans' and possibly 'spatulas' and 'whisks'. Most dictionary users know the different types of structures which provide some room for people or the utensils which can be loaded into a dishwasher. So listing two things of the same class of items would enable dictionary users to infer the remaining names of spaces which *accommodate*. This seems to be a possible "pass" on the use of 'etc.' and the con-

clusion drawn here is that 'etc.' is necessary and therefore acceptable when it represents obvious items.

Along the same lines, it can be argued that some definitions with 'etc.' can at times be reworded to dispense with 'etc.' Evidence backing up this claim emerges from definitions of the counterpart items without 'etc.' in some dictionaries. Taking some more care of the definitions which include 'etc.' can help clarify the meaning and either account for all collocates and items or make them easier to infer from context. Compare for example the different definitions of *artistry*:

great skill in creating or performing something, such as in writing, music, sport, etc.

CALD

the creative skill of an artist, writer, actor, or musician **COBUILD**

skill in a particular artistic activity **LDOCE**

the skill and imagination typical of an artist, writer, musician, or actor **MEDAL**

the skill of an artist **OALD**

Only CALD defines *artistry* with 'etc.'; the other three do not but this is only slightly different from the rest. MEDAL and COBUILD offer the same information save the point relating to sports. The other two definitions remain reasonably clear. Apparently, 'artistic activity' in LDOCE, superordinate, covers 'writing', 'acting', 'sporting', 'playing music' and others. A similar example is the word *amulet* which is defined with 'etc.' in all dictionaries except CALD. In the CALD definition (an object worn because it is believed to protect against evil, disease or unhappiness) the word 'unhappiness' is used in the same way as 'building' above and covers concepts such as *bitterness, cheerlessness, devastation, disaffection, discontent, dismay, distress, gloom, grief, guilt* and others. The COBUILD definition (a small object that you wear or carry because you think it will bring you good luck and protect you from evil or injury) may be well ahead of all others in being concise and clear.

In the second type, the gap left by 'etc.' is almost impossible to fill because the definition has an elusive quality and the user is in doubt as to the range of items which could fill the gap. Sometimes although the items before 'etc.' are in the same class, the user may find it difficult to provide the unnamed items, either because the ones before 'etc.' may be unfamiliar or there are no obvious clues as to how to fill the gap. Many of the instances examined showed that the items preceding 'etc.' had nothing in common at all. Consider the following examples:

adherence: when someone acts strictly according to rules, beliefs, etc. **LDOCE**

succeed: to be successful in your job, earning money, power, respect, etc. **OALD**

The two items before 'etc.' in the first example do not belong to the same semantic field. This would perplex the user; nothing exists to enable them tell whether 'imagination', 'wish' or 'thinking' can be used instead of 'etc.' Likewise, they cannot tell whether someone acts in accordance with a parent's 'advice',

boss's 'instructions' or friend's 'recommendation' because these are neither rules nor beliefs. The same holds true for the second example, despite containing four items before 'etc.' To demonstrate further, here are more problematic examples.

sitting: a period spent by a model who is being painted, photographed, etc. CALD
slash-and-burn: getting rid of a lot of people, systems, etc. in an organization CALD
spawn: the eggs of fish, frogs, etc. CALD

What else can be used instead of 'etc.' in the first example: 'polished', 'cleaned', 'beautified', etc.? Can words such as 'documents', 'furniture', 'cars' be added to replace 'etc.' in the second example? What useful clues exist to help dictionary users decide? As for the last example, the task of dictionary users may be easier if the items which can be used instead of 'etc.' all belonged to the same semantic class. Dictionary users may need to know whether 'snakes', 'toads', 'salamanders', 'newts', 'ducks', 'birds', 'lizards', etc., which all lay eggs, could be included here. Essentially, they are confronted with one of two problems. First, the items which fill the gap do not belong to the same semantic field or share any semantic features, apart from being living creatures. In this case it would be highly likely that users fail to correctly provide the rest of the items to replace 'etc.' Second, no information exists as to restriction on the range of reference of the defined word. They may therefore give up the whole process of guessing the unnamed items.

As items before 'etc.' vary in reference and semantic fields, they may be quite confusing; it is essential that the relationship between the items commonly preceding 'etc.' is intently studied in order not to baffle the users. OALD definition of **segue** (to move smoothly from one song, idea, activity, condition, etc. to another) uses the superordinates 'song', 'idea', 'activity' and 'condition' which diverge considerably; this may confuse users and possibly inhibit their ability to offer any items to replace 'etc.' The same can be said about **sabotage** (to secretly damage or destroy equipment, vehicles, etc. that belong to an enemy or opponent, so that they cannot be used) where 'equipment' and 'vehicles' are unrelated. It is true that users can give some hyponyms of these superordinates, e.g. 'machines', 'railway lines', 'pipelines' or 'bridges' but how could they tell whether this meaning covers public or private 'property' or both? How is this different from 'vandalism'? Is it just the time at which it takes place that contributes to the distinction between the two? The list of similar examples is in fact too long and one can keep talking about them for ages. All in all, such cases of 'etc.'-use need to be entirely abandoned so that definitions are more productive and user-friendly than ever.

In some cases where 'etc.' appears twice in the definitions, the items before 'etc.' may differ and 'etc.' itself does not refer to the same items meant in the first mention (see definitions of *sight* and *syndicate* in 4.3.3), which of course may 'add fuel to the fire' and obfuscates matters still further; the process of guessing the items which substitute for 'etc.' is thus more demanding. In most

cases where 'etc.' is used more than once, the items before 'etc.' are repeated either in the same way or just some of them recur. However, in very few cases, more items are given in the second occurrence of 'etc.' as follows:

star: MAIN PERSON IN A FILM/PLAY, ETC. the person who has the main part, or one of them, in a film, play, show, etc. **LDOCE**

However, in some definitions 'etc.' seems to stand for vital information, whereas in many cases it does not; it is just redundant and seldom adds to the proper use of the defined word as appears in this definition.

object: when someone is extremely unhappy, poor, frightened, etc. **CALD**

Basically, knowing more human qualities does not add to the correct use of the word 'object'. Similarly one wonders whether adding other adjectives may contribute to the clarity of the meaning of *object*. To indicate that the person is 'without hope' may be more useful than 'etc.' This and the like definitions suggest that lexicographers can sometimes dispense with 'etc.'

Dictionary writers can capitalise on those items before 'etc.' to help dictionary users as the number of items before 'etc.' may play a role in illustrating the ones to be inferred. Apparently, the more items given before 'etc.' the easier would be the task of the users; these items are expected to provide more clues to help users guess the words which stand for 'etc.' (Alzi'abi 2016). This however is not always true as there are cases where providing new items to substitute for 'etc.' remains really hard (see Nielsen 2008 for reasons why dictionary consultation may be unproductive). When only one item is listed before 'etc.', it may be an uphill task to provide correct lexemes to replace 'etc.', because no obvious clues exist to help users infer the remaining items from analogy (see also Nielsen 2008). Consider the following examples.

armoured car: a car that has special protection from bullets, etc., used especially by important people **LDOCE**

abstract: (technical) to make a written summary of a book, etc. **OALD**

spill sth out: to tell sb all about a problem, etc. very quickly; to come out quickly **OALD**

safeguard: a law, rule, plan, etc. that protects someone or sth from harm or problems **MEDAL**

One wonders whether a foreign learner with a limited world knowledge can, for example, tell what else other than 'bullets' from which an armoured car may provide special protection. Are 'missiles', 'RPG rockets' 'mortar bombs' included? Again, how can this learner tell whether a 'speech', 'short story', 'poem', 'report', 'communique', 'dictionary' or even 'film' might be abstracted? Interestingly, the last example from MEDAL does not conform to the policy of the editors which aimed at using 'etc.' in cases where the words which could follow were clear enough to be provided by the users themselves. Maybe users might be able to add words such as 'report', 'story', 'version' of something but it

is doubtful whether they can tell what could not be used. Similar instances are far too many to cover here.

One further point to add at this juncture is the fact that dictionary users tend to stop at 'etc.' and focus on the information preceding; they ignore the part of the definition which follows (Alzi'abi 2016); the definition may contain vital information regarding the correct use of the defined item. For instance, when users stop at 'etc.' in OALD definition of **appearance** (an act of appearing in public, especially as a performer, politician, etc., or in court) and disregard the remainder (or in court), they miss valuable information about word usage. Dictionary editors who rely much on the use of 'etc.' have to take care of the information which appears before 'etc.' in the definition. They need to place 'etc.' at the end of the definition and make sure that more successful definitions are offered to avoid any ambiguities.

On the whole, most editors of dictionaries still insist on the use of 'etc.' Peters (2004: 190), for example, believes that using this device is much more efficient than the "wordy translations" which would replace it. It cannot be suggested, therefore, that this deeply entrenched device should be abandoned altogether at one stroke. But more care is required on their part in the use of this device where necessary.

This study was not limitation-free. One limitation was failure to access the full electronic raw data of the dictionaries, their Advanced Search facilities as well as their Style Guides to handle all definitions more easily and present a clearer and more comprehensive picture of the use of 'etc.' Had this data been accessible, more accurate figures, reliable results and hard evidence regarding the use of 'etc.' could have been obtained, let alone more valid conclusions. Although the two letter stretches scrutinised were randomly selected, some may still think that this could be an important limitation of this study as this sampling might not be optimal. Selecting the two-letter stretches, particularly the untypical 'a' letter, may be considered a bad decision for some researchers. According to some (e.g. Bukowska 2010; Cormier 2008 and Miyoshi 2007), the two-letter stretches may not be representative of the dictionary wordlist and a more systematic sampling technique should have been used.

6. Conclusion and recommendations

This research has thoroughly examined the use of 'etc.' in two letter stretches in four learner's dictionaries. Analysis has shown that it is used extensively in the definitions of all types of lexical items. It appears in the definitions of almost all word classes representing several linguistic patterns. Many definitions contain one occurrence of 'etc.' but there are some which have two or three occurrences. It was occasionally used many times in the definitions of headwords with multiple senses. In most cases lexicographers tend to use 'etc.' after two or three items, but sometimes either one or four items precede it.

The examination of a large number of definitions in the four dictionaries

has shown that lexicographers are not of one mind as to the general use of 'etc.' LDOCE and OALD have been overusing this device, unlike CALD and MEDAL. This disparity in the use of 'etc.' reveals a state of considerable confusion encompassing the way 'etc.' is used. One still cannot tell why one definition has one occurrence of 'etc.' and another two or more, and why one definition has one or two items before 'etc.' and another three or four. The same holds true for the individual lexemes which are defined by means of 'etc.' in all dictionaries.

Strunk and White (1972) and Walston (2001) recommend that 'etc.' is not to be used even when the items preceding it belong to the same semantic field especially in cases where the would-be readers might be unfamiliar with items which stand for 'etc.' in a particular context. However, in the light of the above findings, lexicographers — particularly those compiling learner's dictionaries — should be encouraged to dispense with 'etc.' most of the time. They can enumerate the items which replace 'etc.', particularly when these items are not many. In case this proves impossible, they may alternatively reserve the use of 'etc.' to cases when the information is hard to cover, but they have to provide stronger pointers as to what 'etc.' stands for.

Based on the above preliminary findings, it is strongly recommended that some research be conducted to ascertain whether learners fare better when the items before 'etc.' belong to the same semantic field and whether increasing the number of items before 'etc.' correlates to the user being able to infer more lexemes to fill the 'etc.-gap' in definitions.

Endnotes

- i. The theme of this article emerged from a small section about 'etc.' in dictionaries in Alzi'abi (2016).
- ii. It is not the intention of this research to study the history of this linguistic unit.
- iii. See www.urbandictionary.com/define.php?term=ETC.
- iv. It should be reiterated that all examples were drawn from LDOCE unless otherwise indicated.
- v. Although the statistics here concern OALD 8 but a quick scan of the two letter stretches in OALD 9 showed that the latter had only a few more items defined with 'etc.'
- vi. These figures can never be precise unless provided by the publishers of the dictionaries.
- vii. See <http://www.newgeneralservicelist.org/>.

References

A. Dictionaries (All on DVD-ROM except those asterisked)

- Carroll, K. (Ed.). 2012. *Collins COBUILD Advanced Dictionary of English*. Seventh Edition. Glasgow: HarperCollins. (COBUILD7)
- Deuter, M. et al. (Eds.). 2016. *Oxford Advanced Learner's Dictionary*. Eighth revised edition. Oxford: Oxford University Press. (OALD9)

- Higgleton, E., H. Sargent and A. Seaton (Eds.).** 1992. *Chambers Pocket Dictionary*. Edinburgh: Chambers.
- Mayor, M. (Ed.).** 2009. *Longman Dictionary of Contemporary English*. Fifth edition. Harlow: Pearson Education. (LDOCE5)
- Rundell, M. (Ed.).** 2007. *Macmillan English Dictionary for Advanced Learners*. Second Edition. Oxford: Macmillan. (MEDAL2)
- Sinclair, J. (Ed.).** 1987. *Collins COBUILD English Language Dictionary*. London/Glasgow: HarperCollins. (COBUILD)
- Summers, D.** 2002. *The Longman Language Activator*. Second Edition. Harlow: Pearson Education.
- Turnbull, J. et al. (Eds.).** 2010. *Oxford Advanced Learner's Dictionary of Current English*. Eighth edition. Oxford: Oxford University Press. (OALD8)
- Walter, E.** 2008. *Cambridge Advanced Learner's Dictionary*. Third edition. Cambridge: Cambridge University Press. (CALD3)

B. Other literature

- Alzi'abi, S.E.** 1995. *Missing Words and Missing Definitions: NL Arabic Speakers' Use of EFL Dictionaries*. Unpublished Ph.D. Thesis. Swansea: University of Wales.
- Alzi'abi, S.E.** 2016. Help or Hindrance: "Etc" in Learner's Dictionaries? A Study of Verb-noun Collocations. *European Journal of Applied Linguistics* 4(2): 1-27.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Ayto, J.R.** 1983. On Specifying Meaning: A Semantic Analysis and Dictionary Definitions. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 89-98. London/New York: Academic Press.
- Barnbrook, G. and J. Sinclair.** 1995. Parsing Cobuild Entries. Sinclair, J., M. Hoelter and C. Peters (Eds.). 1995. *The Languages of Definition: The Formalization of Dictionary Definitions for Natural Language Processing*: 13-58. Luxembourg: Office for Official Publications of the Communities.
- Bukowska, A.A.** 2010. Sampling Techniques in Metalexigraphic Research. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress*: 1258-1269. Ljouwert: Afûk.
- Browne, C.** 2014. A New General Service List: The Better Mousetrap We've Been Looking For? *Vocabulary Learning and Instruction* 3(1): 1-10.
- Cormier, M.** 2008. Usage Labels in the Royal Dictionary (1699) by Abel Boyer. *International Journal of Lexicography* 21(2): 153-171.
- Dziemianko, A. and R. Lew.** 2006. When You are Explaining the Meaning of a Word: The Effect of Abstract Noun Definition Format on Syntactic Class Identification. Corino E., C. Marelllo and C. Onesti (Eds.). 2006. *Proceedings of the 12th EURALEX International Congress, Torino, Italia, September 6-9, 2006*: 857-863. Alessandria: Edizioni Dell'Orso.
- Geeraerts, D.** 2003. Meaning and Definition. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 83-93. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.

- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing Computing and the Development of the Collins COBUILD English Language Dictionary*: 116-136. London/Glasgow: Collins ELT.
- Hausmann, F. and A. Gorbahn.** 1989. COBUILD and LDOCE II: A Comparative Review. *International Journal of Lexicography* 2(1): 44-56.
- Jackson, H.** 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Lew, R.** 2013. Identifying, Ordering and Defining Senses. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 284-302. London: Bloomsbury.
- Lew, R. and A. Dziemińko.** 2006a. A New Type of Folk-inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information. *International Journal of Lexicography* 19(3): 225-242.
- Lew, R. and A. Dziemińko.** 2006b. Non-standard Dictionary Definitions: What They Cannot Tell Native Speakers of Polish. *Cadernos de Tradução* 18: 275-294.
- Lew, R. and A. Dziemińko.** 2012. Single-clause *when*-definitions: Take Three. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7-11 August 2012, Oslo*: 997-1002. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Miyoshi, K.** 2007. *Johnson's and Webster's Verbal Examples: With Special Reference to Exemplifying Usage in Dictionary Entries*. Tübingen: Max Niemeyer.
- Moerdijk, F.** 2003. The Codification of Semantic Information. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 273-296. Amsterdam: John Benjamins.
- Moon, R.** 2007. Sinclair, Lexicography, and the Cobuild Project: The Application of Theory. *International Journal of Corpus Linguistics* 12(2): 159-181.
- Nielsen, S.** 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.
- Osselton, N.E.** 2007. Innovation and Continuity in English Learners' Dictionaries: The Single-clause *when*-definition. *International Journal of Lexicography* 20(4): 393-399.
- Peters, P.** 2004. *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.
- Rundell, M.** 2006. More than One Way to Skin a Cat: Why Full-sentence Definitions have not been Universally Adopted. Corino E., C. Marellò and C. Onesti (Eds.). 2006. *Proceedings of the 12th EURALEX International Congress, Torino, Italia, September 6-9, 2006*: 323-337. Alessandria: Edizioni Dell'Orso.
- Sinclair, J. (Ed.).** 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing Computing and the Development of the Collins COBUILD English Language Dictionary*. London: Collins ELT.
- Strunk, W. and E. White.** 1972. *The Elements of Style*. Second Edition. New York: Macmillan.
- Van der Meer, G.** 2004. On Defining: Polysemy, Core Meanings, and "Great Simplicity". Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Vol. 2: 807-815. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- Walston, R.** 2001. Coffee Talk. Today's topic: Etc., etc., etc.? <http://www.Columbiaseminary.edu/coffeetalk/035.html> (accessed 9 May. 2015).
- Zgusta, L.** 1971. *Manual of Lexicography*. Janua Linguarum. Series Maior 39. The Hague: Mouton.

On the Metalexigraphic Genre of Dictionary Reviews, with Specific Reference to *LexicoNordica* and *Lexikos*

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

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

Abstract: Wiegand (1984) introduced dictionary criticism as a formal component of a general theory of lexicography. Since then many scholars have focused on various aspects of dictionary criticism. In this article a distinction is made between two major types of dictionary criticism, i.e. the review of metalexigraphical literature and the review of dictionaries. In the review of dictionaries different types can be identified, i.e. a short discussion without a critical evaluation, a more comprehensive review, a review article, reaction to a review or a review article, a peer review and the evaluation of a dictionary as a part of a bigger metalexigraphical contribution. This article gives a brief discussion of the scope of dictionary criticism and looks at some criteria for dictionary reviews before focusing on aspects of dictionary reviews in the Scandinavian journal of lexicography *LexicoNordica* and the South African lexicography journal *Lexikos*. A comparison is made between the reviews in these two journals, identifying striking similarities and differences. It is also shown that the reviews in *Lexikos* frequently refer to topics that are not as relevant to *LexicoNordica*. This is due to the fact that dictionaries reflect something of the linguistic and cultural environment where they are published.

Proposals are made to enhance the quality of reviews in *Lexikos* and to help ensure a stronger theoretical approach to reviews as a genre that elevates the metalexigraphic discourse and improves the dictionary culture of a given community.

Keywords: CHARACTERISTICS OF A GOOD REVIEW, DICTIONARY CRITICISM, DICTIONARY CULTURE, DICTIONARY REVIEWS, GENUINE PURPOSE, METALEXICOGRAPHIC PUBLICATIONS, QUALITY ASSURANCE, REVIEW ARTICLES, REVIEWS, TARGET READERS

Opsomming: Oor die metaleksikografiese genre van woordeboekresensies, met spesifieke verwysing na *LexicoNordica* en *Lexikos*. Wiegand (1984) het woordeboekkritiek as formele komponent van 'n algemene leksikografieteorie voorgestel. Sederdien het baie navorsers aandag aan verskillende aspekte van woordeboekkritiek gegee. In hierdie artikel word 'n onderskeid gemaak tussen twee hoof tipes woordeboekkritiek, naamlik resensies van metaleksikografiese literatuur en resensies van woordeboeke. Verskillende tipes woordeboekresensies

word onderskei, naamlik 'n kort bespreking sonder 'n kritiese evaluering, 'n uitvoeriger resensie, 'n resensie-artikel, reaksie op 'n resensie of resensie-artikel, 'n ewekniere sensie en die evaluering van 'n woordeboek as deel van 'n groter metaleksikografiese bydrae. Hierdie artikel kyk kortliks na die bestek van woordeboekkritiek en na verskillende kriteria vir woordeboekresensies voordat daar gefokus word op aspekte van woordeboekresensies in die Skandinawiese leksikografietydskrif *LexicoNordica* en die Suid-Afrikaanse leksikografietydskrif *Lexikos*. 'n Vergelyking word getref tussen resensies in hierdie twee tydskrifte met verwysing na opvallende ooreenkomste en verskille. Daar word ook aangetoon dat resensies in *Lexikos* dikwels na onderwerpe verwys wat nie so relevant is vir *LexicoNordica* nie. Dit spruit daaruit voort dat woordeboeke iets weerspieël van die linguïstiese en kulturele omgewing waar hulle gepubliseer word.

Voorstelle word gemaak ter verbetering van die gehalte van resensies in *Lexikos* wat kan help om 'n sterker teoretiese benadering te verseker tot resensies as 'n genre wat die metaleksikografiese diskoers verhoog asook die woordeboekkultuur van 'n gegewe samelewing verbeter.

Sleutelwoorde: GEHALTEVERSEKERING, KENMERKE VAN 'N GOEIE RESENSIE, METALEKSIKOGRAFIESE PUBLIKASIES, RESENSIE-ARTIKELS, RESENSIES, TEIKENLESERS, WERK-LIKE DOEL, WOORDEBOEKKRITIEK, WOORDEBOEKKULTUUR, WOORDEBOEKRESENSIES

1. Introduction

In the development of lexicography as an academic discipline different theoretical approaches have come to the fore, focusing on a variety of features and components relevant to lexicographic theory. One of the significant early contributors in this regard has been the German metalexigrapher Herbert Ernst Wiegand. Wiegand (1984: 15) already introduced the different subsections of metalexigraphy, as being (1) the history of lexicography, (2) a general theory of lexicography, (3) research on dictionary use and (4) dictionary criticism. Wiegand (1989: 262) uses the term *Wörterbuchforschung* (dictionary research) and allocates the following four research areas as subfields of the broader research field: (1) research into dictionary use, (2) critical dictionary research, (3) historical dictionary research and (4) systematic dictionary research. In his further research Wiegand has dealt with some aspects of the concept of critical dictionary research, cf. Wiegand (1993), but not nearly in as much detail as he has devoted to some of the other subfields. Wiegand (1998) indicated that critical dictionary research would be a topic in the envisaged second volume of his comprehensive book *Wörterbuchforschung* but the second volume has not been published. In the first volume (Wiegand 1998: 264) he does make the apt remark that scientific dictionary criticism is the mental/intellectual weapon in conflicts between commercial lexicography and dictionary research.

In the metalexigraphic literature discussions that have been devoted to dictionary criticism have often been fairly brief but there are also more in depth discussions of dictionary criticism that can be found in, among others, Ripfel (1989), Nakamoto (1994), Bogaards (1996), Hartmann (1996), Chan and Taylor (2001), Bergenholtz (2003), Nielsen (2003), Swanepoel (2008), Engelberg and Lemnitzer (2009), Svensén (2009) and Schierholz (2015). Engelberg and Lem-

nitzer (2009) offer a thorough discussion in which they focus on criticism of both printed and e-dictionaries. In spite of all the publications dealing with dictionary criticism, the topic has still not been dealt with as comprehensively as one would have wanted it to be done. Even a publication like the three volume international encyclopedia of lexicography (Hausmann et al. 1989–1991) offers scant attention to dictionary criticism, albeit that one contributor, i.e. Osselton (1989), does focus on dictionary criticism in his chapter. Also the supplementary fourth volume of this publication (Gouws et al. 2013) includes but a single chapter that explicitly deals with dictionary criticism, i.e. Swanepoel (2013).

Although Wiegand (1984; 1989) does not analyse this concept and the scope of the term *dictionary criticism* in detail some of the above-mentioned contributions, e.g. Engelberg and Lemnitzer (2009) and Swanepoel (2008) identify and discuss various subsections of this concept.

Renewed recent interest in dictionary criticism resulted in two conferences in April and June 2016 with dictionary criticism as theme, hosted respectively by the University of Silesia (Katowice, Poland) and the Friedrich Alexander University of Erlangen-Nürnberg (Erlangen, Germany) under the auspices of the consortium of EMLex, i.e. the European Masters in Lexicography, a prominent new role player on the international lexicographic scene. Proceedings of these conferences will be published towards the end of 2016 in the book series *Lexicographica Series Maior*.

This article gives a brief discussion of the scope of dictionary criticism and looks at some criteria for dictionary reviews before focusing on aspects of dictionary reviews in the Scandinavian journal of lexicography *LexicoNordica* and the South African lexicography journal *Lexikos*. A comparison is made between the reviews in these two journals, identifying striking similarities and differences.

2. The scope of dictionary criticism

When dealing with dictionary criticism as one of the components of the broader fields of dictionary research and metalexicography it is important to have a clear understanding of the scope of this research field. Looking at the metalexicographic discussion it is clear that the scope of the concept of dictionary criticism does not always go beyond the mere review of dictionaries. The current article assumes that dictionary criticism refers to a much wider lexicographic activity that includes reviews of both metalexicographic literature and dictionaries. In both these categories a distinction is made between (1) a short description without a critical evaluation, (2) a review directed at a publication as a whole, (3) a review article that offers a more comprehensive and scientifically in depth discussion, (4) response by the author of the metalexicographic work or the dictionary to a review or a review article, (5) peer reviews, e.g. of metalexicographic publications and conference abstracts, and (6) the evaluation of a dictionary or aspects of one or more dictionaries as part of a bigger metalexicographical publication.

Category (1) is often only a brief announcement of a new dictionary, often found in a non-scientific publication like a newspaper but also in scientific journals as a lexicographic news snippet, with a very restricted discussion of the new product, whereas category (2) gives a more comprehensive discussion, either in a scientific or a non-scientific publication, where the reviewer has the opportunity to give a well-motivated assessment of the reviewed dictionary or metalexigraphic publication. In dictionary criticisms belonging to category (3) the review article offers an in depth discussion of either the dictionary or metalexigraphic publication as a whole or one or more topics selected from the relevant publication for a comprehensive and critical discussion. An example of such a criticism is Swanepoel (2014) with its focus on the distinction between criteria for the description and evaluation of the design features of dictionaries, and the content, structure and operationalization of lexicographic evaluation criteria with regard to the *Woordeboek van die Afrikaanse Taal*, in his review article of volume 14 of this dictionary.

Contributions in category (4) are not found that often but this category represents a component of dictionary criticism that should be expanded because it leads to a critical and bidirectional discussion of lexicographic works. Authors of reviewed publications should get the opportunity to respond to reviews of their work — acknowledging the advice they receive from reviewers and motivating some of their decisions. A good example of such a response is Gove (1962) which gives a response of the editor of the *Webster's Third New International Dictionary of the English Language* (Gove 1961), presumably the most severely criticised dictionary, cf. Sledd and Ebbitt (1962), to the remarks of some of his critics. A more recent example of this type of dictionary criticism is the response by Botha (2005), editor-in-chief of the *Woordeboek van die Afrikaanse Taal*, to the critical review article by De Schryver (2005) and the subsequent response to the response (De Schryver 2005a).

Before publishing reviews of lexicographic publications journal editors will do well to submit these reviews to the authors of the reviewed publications to solicit their response and publish it along with the review. This could ensure a more balanced presentation of dictionary criticism with regard to a specific publication, but it will also stimulate the metalexigraphic discourse.

Peer reviews, category (5), is a type of lexicographic criticism that has not received the necessary recognition as fully-fledged scientific contributions, cf. Bergenholz and Gouws (2015). One of the issues that could be re-assessed in the writing of this type of criticism is the tradition of anonymous peer reviews. Where a reviewer knows who the author of a publication is (either an abstract or a paper) and the author of the publication knows who the reviewer has been the nature and extent of comments and responses could perhaps take cognizance of the relevant context and this could perhaps eventually lead to a significantly enhanced publication. As it is the case with category (5) the importance of the contribution of critical discussions of dictionaries belonging to category (6) has also been neglected in metalexigraphic literature. Many text books on lexicography or metalexigraphic papers contain critical remarks and often

more comprehensive discussions of certain aspects of specific dictionaries or the treatment and presentation of specific data types in certain dictionaries. Albeit that such a discussion might only focus on the way in which a specific dictionary treats one data type in a single dictionary article, it has to be regarded as a significant contribution to the field of dictionary criticism. Metalexigraphers giving this kind of criticism in their publications need to be aware of and should also adhere to the criteria applicable to other forms of dictionary criticism.

This article gives a restricted discussion of reviews, looking only at reviews of dictionaries, with a focus on only dictionary criticism belonging to category (2) and category (3). No further reference will be made to the criticism of metalexigraphic literature; this calls for a separate article.

3. Establishing a dictionary criticism culture

Hausmann (1989: 13) introduced the concept of a dictionary culture, i.e. society adapting to lexicography e.g. by becoming more familiar with dictionary types and dictionary using skills. Gouws (2013) argues in favour of a comprehensive dictionary culture that will not only focus on lexicography dealing with languages for general purposes but also lexicography of languages for special purposes. According to Gouws (2016) such an expanded and refined version of a dictionary culture could be further adapted to include a positive awareness of the need for and importance of a critical approach to dictionaries. A society characterised by a sophisticated and comprehensive dictionary culture will acknowledge the importance of dictionary criticism and will see this genre as a significant method of guiding users to make informed choices when consulting and buying dictionaries. Stronger theoretically-based reviews will lead to an enhanced level of dictionary criticism and therefore to a broadening of the metalexigraphic discourse.

The current article briefly looks at some aspects of dictionary reviews before referring to a comprehensive study of dictionary reviews in the journal *LexicoNordica* (Bergenholtz 2003). Some of the outcomes of this investigation are then briefly compared with those of a much more limited investigation of dictionary reviews in the journal *Lexikos*.

4. Reviews and peer reviews as form of quality assurance

Reviews always deal with new works. In the evaluation of older publications one does not have a review but rather a general discussion or historical presentation.

It is quite natural that a single scientist or even a lexicographic team will rejoice over a positive review of their work — and they will be much less excited about a negative review. But even a few points of well-founded criticism often function as a call for methodological or stylistic changes in future

publications or revisions of the same work. This also applies to dictionaries — both with regard to the correction of concrete errors, e.g. orthographic mistakes, and underlying theoretical problems. Lexicographers should regard it as part of the responsibility of reviewers to give an honest and critical assessment of the quality of the dictionary and they therefore usually appreciate this kind of feedback that can enhance the quality of their future work.

The influence of reviews may never be underestimated and a review may even compel the publisher to withdraw the work from the market and stop any further sales. But even if a publisher does not opt for such a drastic decision after a devastating review, such a review can have a negative influence on further sales of the dictionary. The first three volumes of the six volume Brockhaus/Wahrig (Wahrig et al. 1980–84) received extremely negative reviews, cf. Wiegand and Kučera (1981, 1982) and Ringguth (1982). Albeit very negative, these reviews were published in journals for linguists and due to the very restricted reader group of these journals the reviews did not have much influence. However, a discussion of these reviews in *Der Spiegel*, a popular German weekly magazine with more than a million copies sold weekly at that stage, most probably caused a dramatic drop in sales of the Brockhaus/Wahrig with the end result that the publishing house eventually had to withdraw the six volume dictionary from the market.

In this case it had not been the intention of the reviewers to harm the publishing house but, in order to achieve future quality assurance, to make them aware of deficiencies and mistakes in the dictionary and of problems in the practice of copying from another dictionary. However, reviews may sometimes not primarily focus on quality assurance but may rather endeavour to discredit a given publication. In another, non-lexicographic, field this kind of approach presumably prevailed in the work of four different reviewers of Bjørn Lomborg's *The Sceptical Environmentalist* (Lomborg 2001) that were published in the journal *Scientific American*. A central theme in the book by Lomborg was his scepticism, based on statistical data, regarding the forecast of a sudden occurrence of global warming, caused by humans. Lomborg specifically criticised the huge sums of money allocated to avoid this alleged warming and the fact that a lot of money would still have to be spent in future. Lomborg argues that the money could rather have been used to improve the assistance and self-assistance of the poorest people in the third world. The interpretation of some data was questioned in four extremely negative reviews. Without referring to other positive reviews in other journals these reviews were used as main argument in a complaint submitted to the Danish ethical commission for scientific dishonesty. The complaint explicitly stated in a written motivation its opposition to Lomborg's envisaged appointment as director of a new environmental institute. Lomborg was appointed but a new complaint was launched to the Commission for scientific dishonesty. In spite of protests from several hundred Danish scientists he was convicted of scientific dishonesty. Consequently Lomborg withdrew from the position that he obtained as director of the environmental institute (cf. the article for "Lomborg" in Wikipedia).

From this it is clear that reviews can have a significant influence on the reviewed publication, on future works of the same author or the same team and even on personal career possibilities. Reviews can also be employed as weapons in scientific arguments. A reviewer cannot prevent such misuse but by adhering to the methodological and ethical rules that are discussed in the following sections of this article the reviewer can at least try to produce a review that does not provide any explicit arguments to assist the misuse thereof.

Dictionary reviews can also have a positive influence. The comprehensive multi-volume Afrikaans dictionary the *Woordeboek van die Afrikaanse Taal* (the WAT) had been the target of extremely negative criticism, expressed in a number of reviews of especially volumes 6 and 7, cf. Grobler (1978), Combrink (1979), Odendal (1979) and Gouws (1985). As a result of this criticism the editorial staff of the WAT organised a special seminar to which metalexigraphers and linguists were invited. The purpose of the seminar was to discuss the various points of criticism directed at the WAT. These discussions led to a completely revised editorial policy that resulted in much better lexicographic work in the subsequent volumes of the WAT. In his review article, focusing on volume XIV of the WAT, Swanepoel (2014: 378) argues that the compilers of the WAT have relied on input from, among others, dictionary reviewers. These reviews have enhanced the quality of the WAT (see also Gouws 2016).

In spite of this important role of reviews for individual scientists and science as a whole one unfortunately notices in the internal evaluation of research at universities an increasing lack of consideration for reviews and peer reviews. A single example is given that is typical of universities in Europe, Asia, Africa and North and South America. Table 1 shows how credits were previously obtained at the Aarhus School of Business in Denmark — also for reviews. The University had a system according to which credits were allocated to different types of publications. Each member of the academic staff had to earn a certain minimum number of credits per year and reviews contributed to achieving the prescribed aims.

Peer reviewed paper in a journal or a book	5
Non-peer reviewed paper in a journal or a book	2
Book	10
Presenting a paper at a conference	1
Paper in the proceedings of a conference.	2
Dictionary	10
Textbook	10
Contribution to a dictionary or a textbook	3
Review	1
Peer review for a journal or a book	2

Table 1: Points allocated to research outputs

Today no credits are given for either reviews or peer reviews. Academics are rather requested to refrain from these activities in order to write more papers in peer-reviewed journals. Even when one includes reviews and peer reviews in your research outputs they do not count anything in the evaluation of scientific activities. This tendency at a Danish university is symptomatic of a worldwide tendency that is extremely disadvantageous to science. Fortunately, in spite of this situation there still are researchers who are willing to write reviews and peer reviews and who try to do it as good as they can.

A change in the way reviews and peer reviews are counted is much needed. Reviews are important products of research, also of lexicographical research, for the reviewed scholar anyway, but reviews and peer reviews should count more in strengthening the reputation of lexicographers, also by being recognized as fully-fledged research outputs in the research evaluation process of universities.

It is not only important to recognize the role of reviews in the process of quality assurance; it is also important to acknowledge their role in lexicographic curricula, in the formulation of lexicographic theory and as a topic in metalexigraphic discussions. The curricula of academic programmes and courses in lexicography should give more attention to dictionary reviews. Criteria for dictionary reviews need to be developed and embedded in any lexicography course and potential dictionary reviewers need to be made aware of the relevant criteria.

5. What characterizes a good review?

The following proposals for reviewing dictionaries can be expanded with many more. We only give those that we regard as essential for the review of dictionaries.

Proposal 1: The review of a dictionary has to be fair

In principle a publishing house benefits from every discussion of any of their dictionaries. Obviously publishing houses hope for positive evaluations but even a negative reference to a dictionary increases its exposure and makes it better known. Although lexicographers are pleased with positive reviews, negative comments can have the advantage of helping them to improve the next edition of the dictionary.

A question that does arise, is whether all dictionaries need to be reviewed or only, for example, scientific dictionaries, cf. Wiegand (1998: 40), where the review is published in a scientific journal. Should one completely refrain from reviewing non-scientific or even bad scientific dictionaries? Both these questions can clearly and unambiguously be answered in the negative. Each and every dictionary, good or bad, is an instrument that real users can use in real situations. The user has the right to receive assistance and guidance with regard to the purchasing and use of a dictionary. Metalexigraphic discussion

in general can also benefit from discussions of mistakes made in the lexicographic practice. Lexicography deals with dictionaries; not only good dictionaries. Therefore all dictionaries need to be candidates for reviews. The type of dictionary and the potential target user of the dictionary, the source where the review is to be published and the envisaged reader of the review will necessarily have an influence on the nature and extent of the review. Reviewers should be fair by not applying the same approach in their reviews of dictionaries from different types with different target user groups and different genuine purposes, cf. Wiegand (1998: 298).

No matter where it is published, what dictionary is reviewed or who the target readers will be, any review has to be fair. This is most likely to happen when the reviewer employs criteria for the methodology of reviews, cf. for example Tiisala (2000, 271f). Tiisala discusses dictionary reviews where the reviewers had been impressionistic. He acknowledges that this can happen very easily when reviewers only rely on random sampling of dictionary material for their criticism. A set of criteria that can be applied systematically can lead to criticism that can improve the quality of further editions of a dictionary. Each reviewer can formulate such criteria that can lead to a fair assessment of a dictionary but one can also use existing proposals, for example Steiner (1984), Leisi (1993) or Bergenholtz (2003), whose proposals are given here in a slightly altered version:

- A review should give a brief account of the contents and the extent of the dictionary. But it has to be more than an account. It should also give a real impression of and create interest in the dictionary.
- A review should not be too long but should have room for a presentation of some concrete dictionary articles.
- A review has to include an analysis and an evaluation of the dictionary. This should be motivated by the reviewer by means of a clear indication of the grounds for the evaluation.
- The review should be both an evaluation and a user orientation.
- Reviewers have to identify themselves with the editor without refraining from criticism.
- The criticism should be positive by presenting clear alternatives for the criticized dictionary articles.

Being fair should also imply that a review is factually correct. It has to cite and summarise correctly without omitting important issues and without interpolations for which the reviewed work does not offer any concrete data. A review needs to be both critical and fair. In addition, and this is important, the review also has to be subjective to a certain extent, i.e. it has to present a personal evaluation. A review that does not contain a personal assessment by the reviewer on aspects of the work he/she is reviewing is not a real review but at the most a documenting description. A review should be balanced and should neither be completely objective nor completely subjective. A review is a subjective text,

based on objective criteria, cf. Kassebeer (2015). This is how it can be explained that good reviews may present opposing assessments of the same work, and therefore it is important to look at different reviews when utilizing them for quality assessment. When evaluating reviews it is also important to know who the reviewer is and who the target readers of the review are — this may necessarily influence the nature of the review.

These proposals could be formulated more briefly, as for example in Ripfel (1989: 31) or Bergenholtz and Frandsen (1997: 48): A review should (1) describe, (2) evaluate, (3) motivate the evaluation and (4) recommend or discourage the use of the dictionary.

Proposal 2: The editor of a journal or a book with dictionary reviews should ensure that the review is fair

Editors can follow this proposal if, from the outset, they do not invite potential reviewers where they know that there are close positive or negative relations between the lexicographer and the reviewers. In addition, an editor should not accept reviews for publication when they are in clear breach of the given methodological and ethical demands. Finally, the editor should guarantee the possibility that the lexicographer of the reviewed dictionary has the opportunity to respond to both good and bad reviews with a contribution that could be as comprehensive as the review itself. This should preferably be in the same volume of the journal in which the review is published. If this is not possible, then it should be published in the subsequent volume.

Proposal 3: A reviewer should employ ethical considerations relevant to reviews

A review can only be fair if it adheres to certain ethical demands. Wiegand (1993) uses a biblical metaphor to present a series of these "ethical commandments." We paraphrase them here with regard to dictionary reviews:

- One has to demand ethical claims from dictionary reviews.
- A dictionary is reviewed and not the person that compiled the dictionary. A dictionary is not a person but the lexicographer is a person. Reference to the person of the lexicographer or lexicographers and their scientific qualifications has to be avoided — unless it is a reference to generally known facts that readers can also find elsewhere.
- One should not annihilate the dictionary and thereby its lexicographer. The commandment: 'You shall not kill!' also applies to reviews.
- One should not give too much praise when it is only done to win friends or to support your career.
- Although reviewers should separate the wheat from the chaff they should not only look for the wheat and ignore the chaff. One has to look for the wheat and when the reviewers do not find it they should contemplate whether they are blind in one or in both eyes.

- It is not that much of an achievement to identify a missing comma or a single typing error.
- A review should not be a platform for self-promotion of your own dictionary or scientific contributions. Praise in your own mouth stinks.
- A mocking reviewer is like a professor that ridicules his colleagues in his lectures. He may earn the applause from the audience but also the wrath of the gods because he ridicules a lexicographer that cannot defend himself.

Proposal 4: Each dictionary user can review a dictionary

Many potential reviewers are asked by the editor of a journal to write a review. In exceptional cases people also submit unsolicited reviews. These two situations confront the lexicographer with the question as to who is the best suited to write the review of a given dictionary. Does one have to be an excellent metalexicographer with substantial knowledge of the language(s) of the dictionary or an experienced practical lexicographer or trained linguist? Many people are of this opinion and would therefore criticise a review when realising that the reviewer is not a well-known expert of the language(s) or an expert in the field of metalexicography/lexicography or linguistics. This criticism is often expressed by the lexicographer of the reviewed dictionary. It also happens that reviewers apologetically remark in the introduction to reviews that they have written the review without being a language expert. The criticism mentioned here is unfair if the reviewer has not claimed in the review to be such an expert. The apology is unnecessary but interesting and relevant for the reader of the review but it does not diminish the quality and relevance of the review.

In reality any reviewer, whether metalexicographer, linguist or lay person, should pay special attention to the potential functions and user groups mentioned in the front matter text of the dictionary. To a certain degree a lay person can do this exceptionally well if she/he belongs to the envisaged target user group and writes the review based on her/his own experiences of using the specific dictionary. If this can be done in collaboration with a lexicographer it could be almost ideal for a thorough and relevant review. An example of such an endeavour is Pedersen and Pedersen (1996), a collaboration between a high school student and a lexicographer, or Gundersen (2002), that introduced a child as consultant for the review of a children's dictionary.

Working with the assumption that any dictionary user can be a potential reviewer, it is important that the reader of a review should not only know who the reviewer is but also what the position of the reviewer is with regard to the specific dictionary, e.g. a member of the envisaged user group, a fellow lexicographer or a metalexicographer. Reviews should reflect the position of their authors and the assessment of reviews should be done with regard to the specific status and role of the reviewer in terms of the reviewed dictionary.

6. Reviews in *LexicoNordica*

Bergenholtz (2003) presents a comprehensive analysis of reviews of dictionaries published in the Scandinavian lexicographic journal *LexicoNordica* for the period 1994–2002. Some of the results of his research and some of the tables he used will be given in the following paragraphs to present a basis for comparison when looking at results from the South African lexicographic journal *Lexikos*. Coming from reviews in two lexicographic journals the results from *LexicoNordica* and *Lexikos* are ideal for a comparative study. The criteria applied in Bergenholtz (2003) have therefore also been used in the analysis of *Lexikos*.

The following table from Bergenholtz (2003) gives the number of reviews published in *LexicoNordica* in each year of the period investigated:

LexicoNordica	Number of reviews
1994	17
1995	22
1996	14
1997	9
1998	10
1999	11
2000	11
2001	10
2002	6

Table 2: Number of reviews in *LexicoNordica*

These reviews, a total of 110 over a period of 9 years at an average of 12,2 per annum, were directed at different types of dictionaries, e.g. monolingual general language dictionaries, bi- or polylingual general language dictionaries, bi- or polylingual specialised dictionaries and monolingual specialised dictionaries, lexicons and encyclopaedia.

The core of the analysis by Bergenholtz which will also be compared with *Lexikos* regards the topics referred to in these reviews. The following table illustrates these categories and the number of references to them in the reviews in *LexicoNordica*. The total number of references exceeds the number of reviews because a single review usually contains a discussion of more than one aspect of the specific dictionary and often more than one reference to any given category in the list below. As an example: different references can be found in different sections of a single review to the same aspect, e.g. grammar. Each one of these references has been counted individually. Consequently in e.g. ten reviews the number of references to grammar exceeds the number of reviews under discussion:

	LN 94-02	LN 94-02
	number	%
1. communication functions	101	3,1
2. cognitive functions	78	2,4
3. dictionary user	114	3,5
4. advice to the user	6	0,2
5. price	30	0,9
6. layout/web-design	113	3,5
7. about the lexicographer	40	1,2
8. comparison with other dictionaries	215	6,6
9. history of the dictionary	56	1,7
10. reference to other reviews	9	0,3
11. about the reviewer	19	0,6
12. empirical basis	89	2,7
13. outer texts	172	5,3
14. lemma selection	453	13,9
15. ordering of lemmata	72	2,2
16. access	86	2,6
17. article structure	46	1,4
18. prescriptive/descriptive	48	1,5
19. equivalents	185	5,7
20. grammar	289	8,9
21. orthography	74	2,3
22. pronunciation	110	3,4
23. semantic and encyclopaedic knowledge	247	7,6
24. labeling	117	3,6
25. etymology	48	1,5
26. examples	99	3,0
27. collocations	67	2,1
28. idioms	40	1,2
29. illustrations	58	1,8
30. synonymy/antonymy	41	1,3
31. dictionary-internal references	33	1,0
32. entertainment value of dictionary	10	0,3
33. positive assessment	77	2,4
34. negative assessment	18	0,6
Total	3260	

Table 3: Topics referred to in reviews in *LexicoNordica* 1994–2002

7. *Lexikos* 1991–2015

7.1 Types of reviews

In section 2 of this article reference was made to different types of reviews, including (1) a short description without a critical evaluation, (2) a review directed at a publication as a whole, (3) a review article that offers a more comprehensive and scientifically in depth discussion, (4) response by the author of the metalexigraphic work or the dictionary to a review or a review article. The editors of *Lexikos*, as reflected in the presentation of different categories of contributions in the table of contents, make provision for two types of dictionary reviews. In the table of contents they are classified as *Reviews* and *Review articles*, with the latter infrequently occurring in the various volumes of *Lexikos*. The *Lexikos* category of *Reviews* includes the categories (1) and (2) mentioned above — unfortunately without any distinction that could inform the reader of the nature, scope and extent of the discussion contained in the review.

Just as is the case with dictionaries, dictionary reviews should also have clearly identified envisaged target readers. This target readership is primarily determined by the journal or other publication in which a given review is included. Where reviews in a newspaper are directed at a general reader group, a review in a scientific journal should be directed at readers who are semi-experts and experts in the relevant subject field of that specific journal. A dictionary of physics that is reviewed in a journal of physics should be directed at experts and semi-experts in the field of physics. When the same dictionary is reviewed in a lexicographic journal, the focus is not necessarily on the contents in terms of the needs and demands of physicists because physicists do not typically read a lexicographic journal. Such a review will be directed at the target readers of the specific journal, i.e. people interested in the field of lexicography. The scope of their interest could include a discussion of the contents of the dictionary of physics but could also make provision for remarks regarding e.g. the structures, functions, user-perspective, etc. of the given dictionary; aspects that would most probably not be of interest to the physicists. This target reader group of a review should necessarily influence the approach followed by the specific reviewer and the nature of the review.

In a scientific journal like *Lexikos* one would expect the majority of the reviews to be directed at members of the lexicographic community, i.e. experts and semi-experts in the field. These typical readers of *Lexikos* would typically expect a high level review that goes beyond a brief introduction of the given dictionary. Unfortunately this is not the case in many of the reviews. Too many of these reviews belong to the above-mentioned category (1) and do not contribute significantly to a substantial critical lexicographic discourse. Fortunately there are also ample reviews that are more comprehensive as well as review articles that play an important role in developing dictionary criticism as a component of a general theory of lexicography. By allocating a section in the jour-

nal to reviews and by having a dedicated review editor on the editorial board, the editors of *Lexikos* ensure that this journal does not only give a fine reflection of new dictionaries but it also stimulates the metalexigraphic discussion.

The quality of the contribution of *Lexikos* to the field of dictionary criticism could be enhanced by a more structured approach in the writing of reviews. When soliciting dictionary reviews the review editor could consider the possibility of giving the potential reviewers a brief of the kind of review needed for this journal, cf. Gouws (2016). Too often the reviews do not really live up to the expectations of the target readers of *Lexikos*. Such a brief that deviates from a random approach to say something about a dictionary could also help to develop the expertise of dictionary criticism. This would yet again elevate the level of the metalexigraphic discourse.

7.2 Focal points in the *Lexikos* reviews

Looking at reviews in *Lexikos*, volumes 1–25, and employing the set of criteria used by Bergenholz (2003) it is interesting to note both similarities and differences when comparing these reviews to those in *LexicoNordica*. Compared to *LexicoNordica*'s 110 reviews in 9 years at an average of 12,2 per annum, the 25 volumes of *Lexikos* contain 126 reviews at an average of only 5,04 per annum. There is no fixed number of pages allocated to reviews and no clear indication of a trend in terms of the frequency or extent of reviews or review articles in the 25 volumes. The following table indicates the different categories and the number of relevant references (with different references to the same category in a single review counted individually) in *Lexikos*:

	Lexikos 1–25	Lexikos 1–25
	number	%
1. communication functions	16	2,4
2. cognitive functions	4	0,6
3. dictionary user	42	6,3
4. advice to the user	1	0,1
5. price	75	11,1
6. layout/web-design	11	1,6
7. about the lexicographer	6	0,9
8. comparison with other dictionaries	14	2,1
9. history of the dictionary	11	1,6
10. reference to other reviews	3	,45
11. about the reviewer	3	,45
12. empirical basis	8	1,2
13. outer texts	42	6,2

14. lemma selection	78	11,8
15. ordering of lemmata	13	1,9
16. access	15	2,2
17. article structure	4	0,6
18. prescriptive/descriptive	6	0,9
19. equivalents	17	2,5
20. grammar	36	5,2
21. orthography	16	2,4
22. pronunciation	28	4,0
23. semantic and encyclopaedic knowledge	47	7,1
24. labeling	13	1,9
25. etymology	8	1,0
26. examples	22	3,3
27. collocations	10	1,5
28. idioms	12	1,8
29. illustrations	8	1,2
30. synonymy/antonymy	12	1,8
31. dictionary-internal references	15	2,2
32. entertainment value of dictionary	3	,45
33. positive assessment	51	7,7
34. negative assessment	16	2,4
Total	661	

Table 4: Topics referred to in reviews in *Lexikos* 1991–2015.

Before reflecting on a comparison between the two journals it is interesting to look at some other topics in the reviews in *Lexikos*. Albeit that *Lexikos* is an international journal it is based in South Africa. Many dictionaries from South African publishing houses are submitted for reviews and the editors have to respond to these submissions. Dictionaries are not only compiled for specific target users but also to be used in specific linguistic and cultural environments. It is to be expected that South African dictionaries will reflect aspects of the South African multilingual and multicultural situation and these issues are often addressed in the reviews. This implies that some issues come to the fore in the reviews in *Lexikos* that have no or a far lesser impact in the reviews in *LexicoNordica*.

A number of reviews include references to a typical lexicographic topic like the treatment of homonymy and polysemy. Due to the need for a wide-ranging variety of dictionary types in South Africa, the position of a given dictionary within the typological spectrum is a frequent point of discussion in the reviews. Within some of the African languages lexicographers need to make

definite choices with regard to the lemmatisation procedures to be followed in their dictionaries. In this regard the distinction between a word or a stem lemmatisation approach is relevant, and this is a topic that receives coverage in many reviews. The occurrence of dialectal forms and the reality of regional differences in the South African languages are reflected in reviews. Within a multilingual environment language contact inevitably leads to language influence, and therefore the occurrence of loan words and other borrowings should be reflected in dictionaries. The reviewers are keen to identify loan words that have been included as lemmata in dictionaries. The influence of the political situation on the languages of South Africa and the way in which it is reflected in dictionaries also come to the fore in reviews. Sensitivity for language use that can be offensive is often expressed. This applies to words from the political arena but also language use that could be seen as racist or sexist. Reviewers show their readers how the reviewed dictionaries live up to the expectations of dictionary users to find an objective reflection of the specific language treated in the dictionary.

The following table gives a comparison of the percentages of references to the different categories in the two journals:

	Lexico-Nordica 1994–2002	Lexikos 1–25
	%	%
1. communication functions	3,1	2,4
2. cognitive functions	2,4	0,6
3. dictionary user	3,5	6,3
4. advice to the user	0,2	0,1
5. price	0,9	11,1
6. layout/web-design	3,5	1,6
7. about the lexicographer	1,2	0,9
8. comparison with other dictionaries	6,6	2,1
9. history of the dictionary	1,7	1,6
10. reference to other reviews	0,3	,45
11. about the reviewer	0,6	,45
12. empirical basis	2,7	1,2
13. outer texts	5,3	6,2
14. lemma selection	13,9	11,8
15. ordering of lemmata	2,2	1,9
16. access	2,6	2,2
17. article structure	1,4	0,6
18. prescriptive/descriptive	1,5	0,9

19. equivalents	5,7	2,5
20. grammar	8,9	5,2
21. orthography	2,3	2,4
22. pronunciation	3,4	4,0
23. semantic and encyclopaedic knowledge	7,6	7,1
24. labeling	3,6	1,9
25. etymology	1,5	1,0
26. examples	3,0	3,3
27. collocations	2,1	1,5
28. idioms	1,2	1,8
29. illustrations	1,8	1,2
30. synonymy/antonymy	1,3	1,8
31. dictionary-internal references	1,0	2,2
32. entertainment value of dictionary	0,3	,45
33. positive assessment	2,4	7,7
34. negative assessment	0,6	2,4

Table 5: Comparison of percentages in *LexicoNordica* and *Lexikos*.

Each pairing of categories in this table could be discussed in detail. This will not be done in this article. Only a few remarks will be made with regard to some striking similarities and differences.

Many reviewers are keen to say something about the number of words or new words or the types of words and expressions included in a dictionary. Therefore it can be expected that the lemma selection should be a frequent topic of discussion — the topic most frequently referred to in the reviews in both journals. Aspects about semantic and encyclopaedic knowledge also have a comparable high standing in these tables. It is often, and rightfully, said that few users read the front matter texts, including the users' guidelines text in a dictionary. Reviewers in both journals pay significant attention to the outer texts, i.e. the front and back matter texts. This may be seen as an attempt on the side of the reviewers to help to establish a dictionary culture where users are aware of the fact that a dictionary contains more than just the central list, i.e. the alphabetical section.

This table also shows striking differences, e.g. in categories 5, 8, 33 and 34. These differences can be explained by taking cognizance of editorial traditions in the two journals and by prevailing approaches to dictionary reviews on the side of the reviewing community.

Category 5, the price of the dictionary, shows the biggest deviation between the two journals with *LexicoNordica* scoring 0,9 and *Lexikos* no less than 11,1 — the second most frequent category in *Lexikos*. This is due to a template

introduced in dictionary reviews where the heading of the review includes information on the title, author, publisher, etc. and also gives the price of the dictionary. In category 8 the reviews in *LexicoNordica* refer much more often to other dictionaries, compared to the reviews in *Lexikos*. This could be seen as typical of a more developed culture of dictionary criticism where the reviewed dictionary is not seen in isolation and readers have the opportunity to assess the dictionary by comparing some of its features to those of other dictionaries.

Categories 33 and 34 show that reviews in both journals give an assessment of the dictionaries but such an assessment more frequently occurs in the *Lexikos* reviews. According to proposal 1 (The review of a dictionary has to be fair) given above, a review should give an evaluation of the dictionary. This seems to happen more often in *Lexikos* than in *LexicoNordica*. However, the positive assessment (7,7%) in *Lexikos* does not seem to realistically reflect the quality of dictionaries. Too often too many reviews refrain from a negative overall assessment even when individual categories are criticised negatively. Reviewers should work with the idea that criticism can and should be both positive and negative and that it is the responsibility of the lexicographer to give an honest assessment of the dictionary.

The largely comparable percentages in this table indicate that the categories selected could be regarded as representative of what one typically finds in a dictionary review published in a scientific journal of lexicography. It would be interesting to see the results of a similar comparative study of other major journals of lexicography.

8. Conclusion

As an established international journal *Lexikos* will do well to continue with its policy of having a review editor and publishing both reviews and review articles, reviewing both dictionaries and metalexigraphic publications. The importance of reviews could be emphasised by allocating more pages to this component of the journal's contents. The relevance of reviews as part of the academic discourse could be elevated if lexicographers of the reviewed dictionaries or authors of the reviewed metalexigraphic works are invited to respond to the reviews. The quality of reviews could be enhanced if reviewers are given a brief with an indication of the genuine purpose of the review, cf. Gouws (2016), and the expectations of the intended target readers. These reviews should move away from a mere announcement of the dictionary to a comprehensive discussion of its lexicographic nature and contribution.

When soliciting reviews the review editor could consider to invite reviews from reviewer teams, consisting e.g. of a (meta)lexicographer and a typical target user of the reviewed dictionary. These dictionary users could give a valuable input and their ideas could help the other reviewer but also the lexicographers and the lexicographic community at large to embark in a much more effective way on the planning and production of user-directed dictionaries.

Acknowledgement

This research is supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85434). The Grant-holder 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.

9. References

- Bergenholtz, Henning. 2003. Ordbogskritik i LexicoNordica. *LexicoNordica* 10: 7-26.
- Bergenholtz, Henning and Finn Frandsen. 1997. At anmelde i Hermes — tidsskrift for sprogforskning. *Hermes* 19: 169-183.
- Bergenholtz, Henning and Rufus H. Gouws. 2015. Proposals for the Writing of Peer Reviews in Lexicography. *Hermes* 54: 107-114.
- Bogaards, Paul. 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.
- Botha, Willem. 2005. Concurrent Over- and Under-treatment in Dictionaries: A Response. *International Journal of Lexicography* 18(1): 77-87.
- Chan, Alice Yin Wa and Andrew Taylor. 2001. Evaluating Learner Dictionaries: What the Reviews Say. *International Journal of Lexicography* 14(3): 163-180.
- Combrink, Johan G.H. 1979. Die sesde deel van die W.A.T. *Standpunte* 140, 32(2): 49-64.
- De Schryver, Gilles-Maurice. 2005. Concurrent Over- and Under-treatment in Dictionaries The *Woordeboek van die Afrikaanse Taal* as a Case in Point. *International Journal of Lexicography* 18 (1): 47-76.
- De Schryver, Gilles-Maurice. 2005a. Concurrent Over- and Under-treatment. A Short Reply. *International Journal of Lexicography* 18(1): 89-92.
- Engelberg, Stefan and Lothar Lemnitzer. 2009. *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.
- Gouws, Rufus H. 1985. Die sewende deel van die *Woordeboek van die Afrikaanse Taal*. *Standpunte* 178, 38(4): 13-25.
- Gouws, Rufus H. 2013. Establishing and Developing a Dictionary Culture for Specialized Lexicography. Jesenšek, Vida (Ed.). 2013. *Specialised Lexicography*: 51-62. Berlin/New York: De Gruyter.
- Gouws, Rufus H. 2016. Dictionary Criticism and Dictionary Critics. Bielińska, Monika and Stefan J. Schierholz (Eds.). 2016. *Wörterbuchkritik*. Berlin: Walter de Gruyter. (In press.)
- Gouws, Rufus H. et al. (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin/New York: De Gruyter Mouton.
- Gove, Philip B. (Ed.). 1961. *Webster's Third New International Dictionary of the English Language*. Springfield, Massachusetts: Merriam-Webster.
- Gove, Philip B. 1962. A Letter to the Editor of *Life Magazine*. Sledd, J. and W.R. Ebbitt (Eds.). 1962: 91-92.
- Grobler, Hilda. 1978. 'n Voorlopige toepassing van S.P.E. Boshoff se kriteria vir 'n groot woordeboek op WAT I-VI. *Klasgids* 12(4): 29-46.

- Gundersen, Dag.** 2002. Engelsk stor ordbok. *LexicoNordica* 9: 219-228.
- Hartmann, Reinhard R.K.** 1996. Lexicography as an Applied Linguistic Discipline. Hartmann, R.R.K. (Ed.). 1996. *Solving Language Problems: From General to Applied Linguistics*: 230-244. Exeter Linguistic Studies. Exeter: University of Exeter
- Hausmann, Franz J.** 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, Franz Josef et al. 1989-1991: 1-19.
- Hausmann, Franz Josef et al. (Eds.).** 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*. Berlin: De Gruyter.
- Kassebeer, Søren.** 2015. En anmeldelse er hverken noget rent objektivt eller noget rent subjektivt. *Berlingske Tidende* 14.2.2015.
- Leisi, Ernst.** 1993. Sieben Thesen zum Rezensionswesen. *Deutsch als Fremdsprache* 30: 188-189.
- Lomborg, Bjørn.** 2001. *The Sceptical Environmentalist*. Cambridge: Cambridge University Press.
- Nakamoto, K.** 1994. *Establishing Criteria for Dictionary Criticism: A Checklist for Reviewers of Monolingual English Learner's Dictionaries*. Unpublished M.A. Thesis. Exeter: University of Exeter.
- Nielsen, Sandro.** 2003. Anmeldelse af ordbøger. *LexicoNordica* 10: 27-44.
- Odendal, Francois F.** 1979. Plus positief en plus negatief: Gedagtes na aanleiding van WAT VI (KLA-KLO). *Tydskrif vir Geesteswetenskappe* 19(1): 24-41.
- Osselton, Noel Edward.** 1989. The History of Academic Dictionary Criticism with Reference to Major Dictionaries. Hausmann, F.J. et al. (Eds.). 1989-1991: 225-230.
- Pedersen, Jette and Tine Pedersen.** 1996. Jens Axelsen: Dansk-Engelsk Ordbog. *LexicoNordica* 3: 287-294.
- Ringguth, Rudolf.** 1982. Wörterbücher: Vorsicht angebracht. Zwei Germanisten haben die beiden ersten Bände des "Brockhaus/Wahrig" schonungslos verrissen. *Der Spiegel* 21. 27.12.1982: 217-219. Also online <http://www.spiegel.de/spiegel/print/d-14337292.html>. (Accessed November 2015.)
- Ripfel, Martha.** 1989. Wörterbuchkritik: eine empirische Analyse von Wörterbuchrezensionen. Tübingen: Max Niemeyer.
- Schierholz, Stefan J.** 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25: 323-352.
- Schoonees, P.C. et al. (Eds.).** 1951-. *Woordeboek van die Afrikaanse Taal*. . Pretoria: Government Printer/Stellenbosch: Bureau of the WAT.
- Sledd, J. and W.R. Ebbitt (Eds.).** 1962. *Dictionaries and That Dictionary*. Chicago: Scott, Foresman and Co.
- Steiner, Roger J.** 1984. Guidelines for Reviewers of Bilingual Dictionaries. *Dictionaries. Journal of the Dictionary Society of North America* 6: 166-181.
- Svensén, Bo.** 2009. *A Handbook of Lexicography*. Cambridge: Cambridge University Press.
- Swanepoel, Piet H.** 2008. Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria. *Lexikos* 18: 207-231.
- Swanepoel, Piet H.** 2013. Evaluation of Dictionaries. Gouws, Rufus H. et al. (Eds.). 2013: 587-596.
- Swanepoel, Piet H.** 2014. Evalueringskriteria en die interaksie tussen die leksikografieteorie en -praktyk; die ontwerp van die *Woordeboek van die Afrikaanse Taal* as gevallestudie. *Lexikos* 24: 378-401.
- Tiisala, Seija.** 2000. Multifunktionell allmänordbok. *LexicoNordica* 7: 271-290.
- Wahrig, G. et al. (Eds.).** 1980-1984. *Brockhaus-Wahrig. Deutsches Wörterbuch in sechs Bänden*. Wiesbaden: Brockhaus/Stuttgart: Deutsche Verlagsanstalt.

- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 13-30. *Lexicographica. Series Maior 1*. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst.** 1989. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J. et al. (Eds.). 1989–1991: 246-280.
- Wiegand, Herbert Ernst.** 1993. Wörterbuchkritik. Dictionary Criticism. *Lexikographica 9*: 1-7.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst and Antonín Kučera.** 1981. Brockhaus-Wahrig: Deutsches Wörterbuch auf dem Prüfstand der praktischen Lexikologie. I. Teil: 1. Band (A–BT); 2. Band (BU–FZ). *Kopenhagener Beiträge zur Germanistischen Linguistik 18*: 94-206.
- Wiegand, Herbert Ernst and Antonín Kučera.** 1982. Brockhaus-Wahrig: Deutsches Wörterbuch auf dem Prüfstand der praktischen Lexikologie. II. Teil: 1. Band (A–BT); 2. Band (BU–FZ); 3. Band (G–IZ). Wiegand, Herbert Ernst (Ed.). 1982. *Studien zur neuhochdeutsche Lexikographie II*. *Germanistische Linguistik 3–6/80*: 285-373.
- Wikipedia.** 2015. *Bjørn Lomborg*. https://en.wikipedia.org/wiki/Bj%C3%B8rn_Lomborg. (Accessed November 2015.)

Adapting a Historical Dictionary for the Modern Online User: The Case of the *Dictionary of South African English on Historical Principles*'s Presentation and Navigation Features*

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

Tim van Niekerk, *Dictionary Unit for South African English,
Rhodes University, Grahamstown, South Africa (T.vanNiekerk@ru.ac.za)*

Abstract: A collaboration between the Dictionary Unit for South African English (DSAE) and Stellenbosch University (SU) was initiated in 2016 to address the structure, layout and functionality of the online *Dictionary of South African English on Historical Principles* (DSAEHist). This article focuses on work undertaken during the initial stages of the DSAE-SU collaboration and presents current work in adapting a print historical dictionary for the digital age. As with other print dictionaries that enter the electronic medium, the pilot version of the online DSAEHist still felt and looked like a print dictionary, albeit with a somewhat more user-friendly access structure. Another difficulty that the DSAEHist faces is the need for a dictionary based on historical principles to showcase large diachronic datasets. This requires that the article and microstructure, as well as the access structure, be adapted. Apart from this, the general layout of the dictionary is also being re-evaluated and redesigned to make it more usable on online platforms. To accommodate new users and to bring the DSAEHist into the digital era (with regard to standard web and mobile practices) stylistic, navigational and functional changes are being made. This article presents these changes to show how a valuable cultural and language resource is being brought into the digital era.

Keywords: SOUTH AFRICAN ENGLISH, HISTORICAL DICTIONARY, ONLINE DICTIONARY, PRINT ADAPTATION, MICROSTRUCTURE, MACROSTRUCTURE, ARTICLE STRUCTURE, USER RESEARCH, DATA PRESENTATION, USABILITY, NAVIGATION

* This article was presented as a paper at the Twenty-first Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Xitsonga and Sesotho sa Leboa National Lexicography Units, Tzaneen, South Africa, 4–6 July 2016.

Opsomming: Die herstrukturering van 'n historiese woordeboek vir moderne aanlyn gebruikers: Die geval van die *Dictionary of South African English on Historical Principles* se data-aanbieding- en navigasie-eienskappe.

In 2016 is 'n samewerkingsooreenkoms tussen die Dictionary Unit for South African English (DSAE) en die Universiteit Stellenbosch (US) van stapel gestuur om aandag te skenk aan die onvoldoende struktuur, uitleg en funksionaliteit van die aanlyn *Dictionary of South African English on Historical Principles* (*DSAEHist*). Hierdie artikel bied 'n blik op die aanvanklike resultate van hierdie samewerkingsooreenkoms en wys ook hoe hierdie historiese woordeboek aangepas word vir die digitale era. Soos met talle ander gedrukte woordeboeke wat omgeskakel word na elektroniese aanlyn woordeboeke, het die loodswaerwag van die aanlyn *DSAEHist* nog soos 'n gedrukte woordeboek gelyk en gefunksioneer, al was die toegangstruktuur ietwat meer gebruikersvriendelik. 'n Verdere probleem is dat 'n historiese woordeboek soos die *DSAEHist* 'n groot aantal diachroniese data moet kan vertoon. Daarom moet die artikel- en mikrostruktuur, asook die toegangstruktuur, herontwerp en herbedink word. Daarbenewens moet die algehele uitleg en aanbieding van data ten opsigte van gebruikersvriendelikheid vir die elektroniese sfeer ook verfyn word. Ten einde die *DSAEHist* in staat te stel om nuwe gebruikers te lok en te behou, asook om ooreen te stem met die standaarde en beginsels van moderne internet- en mobiele praktyke, is stilistiese, navigasie- en funksionele veranderinge aangebring. Hierdie artikel het ten doel om te wys hoe 'n kosbare kulturele en taalbron aangepas word vir die digitale era.

Sleutelwoorde: SUID-AFRIKAANSE ENGELS, HISTORIESE WOORDEBOEK, AANLYN WOORDEBOEK, AANPASSING VAN GEDRUKTE WOORDEBOEKE, MIKROSTRUKTUUR, MAKROSTRUKTUUR GEBRUIKERSTUDIES, DATA-AANBIEDING, BRUIKBAARHEID, NAVIGASIE

1. Introduction

At the 20th AFRILEX International Conference in 2015, the pilot online version of *A Dictionary of South African English on Historical Principles* (*DSAEHist*) was presented and critically discussed (cf. Van Niekerk 2015). During this session practical and theoretical obstacles and shortcomings of the online *DSAEHist*, especially with regard to general usability, layout and functionality, were highlighted. Subsequently a collaboration between the Dictionary Unit for South African English (DSAE) and the University of Hildesheim (HU) was formalised to "address the need for a thorough adaptation of the dictionary text to support publication on multiple electronic platforms" (Van Niekerk 2015). In 2016 a further collaboration was established between the DSAE and Stellenbosch University (SU). This article elaborates on the work done on the *DSAEHist* as part of the DSAE-SU collaboration. Since the dictionary adaptation is ongoing and the DSAE-SU collaboration builds on and overlaps with prior DSAE-HU work, it may be helpful firstly to distinguish between areas of adaptation and secondly to give a brief background on the pilot version of the *DSAEHist* and clarify the scope of this article.¹ A final introductory note is that

for the development of user-orientated and functional enhancements as part of the SU collaboration, the ideas set forth by Du Plessis (2015) on electronic and mobile dictionaries were relied upon for theoretical guidance.

2. Scope of projects

The scope of the research and adaptations presented here fit into the wider context of other dictionary adaptation projects that are moving from print to digital, for example the *Woordeboek van die Afrikaanse Taal* (Botha and Fourie-Blair 2016) and Macmillan's dictionaries (see Rundell 2015). The main focus of the DSAE-HU collaboration is restructuring the underlying dictionary dataset to support functionality for enhanced search functionality across the dictionary, to provide more flexible presentation of the article microstructure, and to introduce some new visual components and layout strategies customised for the display of complex historical dictionary data. Some new visual dictionary components, which will be presented in a different article (cf. Van Niekerk et al. 2016a), have also been developed. This restructuring of the dataset has made possible certain layout changes and presentation strategies which were proposed as early prototypes but not yet finally modelled under the DSAE-HU collaboration.

The DSAE-SU collaboration builds on this prior work in that it finalises layout changes which have now become possible, as well as modelling some new presentation features not proposed before. Unlike the DSAE-HU project, which focused on changes to the underlying structure of the dictionary dataset and exploiting the results in article presentation, the DSAE-SU project is specifically focused on the user interface, namely enhancements to overall layout and design at the micro- and macrostructural levels, as well as on substantial improvements to navigation features which had not been addressed before. In short, current DSAE-SU work is concerned primarily (but not exclusively) with questions of design, accessibility and navigation as well as, from the user's point of view, the "look and feel" of the dictionary. It should be noted that the screenshots given in this article are wireframe examples which usually emphasise structural and functional design changes, without final font styling, colour schemes and so forth. Many of these prototypes have not yet been implemented since the project is a non-profit undertaking and operates under resource constraints beyond the control of the project team. The intent of this article is therefore to present current work for three key purposes: (1) to present some examples of the dictionary design possibilities that are becoming available in the still-developing area of electronic lexicography; (2) to show how lexicographical theory can meet practical lexicography in a large-scale dictionary project of complex design; and especially (3) to generate discussion and invite feedback from the lexicographical community.

3. The pilot online edition

The *DSAEHist* is a diachronic comprehensive variety dictionary documenting the vocabulary of South African English from its origins in the 17th century to the date of publication. The aim of the dictionary is "to map and illustrate the complex landscape of [this particular] variety of English" (Silva et al. 1996: vii). To achieve this goal, historical data are always incorporated in the form of "bibliographically documented and dated quotations" (Van Niekerk et al. 2016a). The print version of this dictionary has been out of print since 2004, but a pilot online version was launched in 2014 (<http://dsae.co.za>).

As with many other print dictionaries that enter the electronic medium, the pilot version of the online *DSAEHist* still felt and looked like a print dictionary, albeit with a somewhat more user-friendly access structure. Due to the *DSAEHist*'s need to showcase large diachronic datasets, the article and micro-structure, as well as the access structure, are currently not suited to achieve this goal on an electronic platform. Apart from this, the general layout of the dictionary also needed to be re-evaluated and redesigned to make it more usable on online platforms. Many of the aforementioned issues stem from the "internal representation", which was originally designed purely for print typesetting purposes, being only "partly structured with lexicographic categories" (Van Niekerk et al. 2015). Another area of concern was the problematic and somewhat outdated design of the user interface and limited multi-functionality and interactivity. The XML dataset used for the pilot online version lacked the flexibility required to support layout changes, and the general layout is akin to the structure of the printed version which causes navigational, structural and usability obstacles.

In order to move from print to the electronic sphere, different adaptations need to be made. Although many of these adaptations relate to structural and technological aspects, no change can be made without identifying and catering for the dictionary user. Bergenholtz and Tarp (2003: 172) show that dictionaries should satisfy a range of user needs. Designing or adapting a dictionary therefore relies on the balance between dictionary structures and user needs (Gouws 2005: 56, 60). Consequently a review of the *DSAEHist*'s users is crucial for the adaptation process.

4. Dictionary users

The online *DSAEHist* occupies a rather unusual position among electronic dictionaries. It republishes a scholarly dictionary originally compiled over 25 years of comprehensive historical research with a relatively narrow audience of language specialists, historians, writers and interested laypersons. It was not intended as a popular dictionary — this role is fulfilled by the much smaller print dictionary, *A Dictionary of South African English* by Jean Branford (published in four editions between 1978 and 1991). Yet the online *DSAEHist*, unlike

the *OED* on which it was originally modelled, is a free dictionary which is now consulted by a large audience of lay users who may never have seen the historical dictionary model before, let alone a variety dictionary. The result is a forked user base for the electronic edition, with different sets of expectations from two groups, which can be roughly divided as follows:

- Casual Googlers: laypersons who arrive at a *DSAEHist* entry after consulting a search engine about a South African term; and
- Target-users: the original intended audience of writers, editors, translators, linguists, historians and other more sophisticated dictionary users.

This division, coupled with a significant departure from print design and layout conventions in electronic dictionaries, has long-term implications for the adaptation of the original dictionary, prompting many new requirements which can only be addressed in stages. Du Plessis (2015: 27) posits that it is imperative to take note of the range of user requirements as the user always plays a key role in determining the functionality and usability of a dictionary. Therefore, to help identify and prioritise the aforementioned requirements, the DSAE compiled an online user survey which ran from 17 March to 16 May 2016 (Van Niekerk et al. 2016). Another source of information about users is Google Analytics reports, which give general statistical information about users such as their devices, geographical locations and search engine keywords sorted by date or frequency.

The user survey delivered some interesting results but does not necessarily reflect the true scope of the general *DSAEHist* user group since 49 out of the 75 respondents (65.33%) had not used the *DSAEHist* before (Le Du 2016). Nevertheless, this survey sheds some light on possible user groups, needs and expectations. Many of the changes made are also supported by the data from the survey. A few key points from the survey are summarised below.

- The majority of respondents fall into the 60+ age group (38.67%) followed by the 31–40 age group (22.67%), while the 41–59 age group contained 28% of the respondents. That means only around 10% of the respondents are 30 and younger. The skewed data here is mostly due to the distribution of the survey as shown below.
- The majority of respondents are language professionals (58.67%), with a further 9.33% identifying as working specifically in the field of dictionary compilation, publishing, etc. This result however would be influenced by the distribution list of the survey that included many professional and academic institutions or associations. These figures are still consistent with the previously stated target-user group of the *DSAEHist*.
- The majority of respondents identify as speakers of South African English (81.33%) although most still believe that the concept of this variety should be explained to users of the dictionary (86.67%).

- 84% of respondents indicated that they would not prefer a shorter entry with less information when using a computer, although 58.67% indicated that they would like to control the amount of information displayed as clickable options. This relates to adapting the layout in such a way as not to overload the user with data and to make the dictionary environment more functional.

The user survey gives a solid representation of the target-users and reflects on some of their needs or expectations, but it does not necessarily provide enough detail on possible future users (see Section 7). Gouws (2016) mentions that lexicographers need to be aware of "emerging generations" of dictionary users. Nied Curcio (2015) states that the new generation of modern users, the so called "digital natives", are becoming the primary dictionary users. It is likely that a large proportion of the casual Googler user group falls into this category. Therefore, to accommodate both the target-users and emerging dictionary users, the current and any future enhancements to the *DSAEHist* are additionally focused on the evolving needs, expectations and skills of a possible future audience.

From the Google statistics it became clear that the majority of users access the dictionary from a personal computer or laptop. Due to the increase in mobile resources and society's dependence on mobile technology, an increase in mobile users in the near future is however to be expected (Du Plessis 2015: 24-25). This is also conveyed in the user survey where 45% of respondents indicated that they would use a *DSAEHist* app if available. The ideas and enhancements incorporated during the DSAE-SU collaboration are thus implemented in such a way as to make a long-term shift towards improved mobile support easier. Although there is currently a mobile version available, this version will not be discussed here as it needs first to be critically evaluated during a future stage of the adaptation project. The proposed restructuring and simplification of the desktop design will also be reflected in the mobile version and offer significant improvements.

5. Adaptations to the pilot edition

The following two sections illustrate the different adaptations to the pilot edition of the *DSAEHist*. In some cases prior preparatory work is mentioned as introductory material, but it will be clear from the specified dates or other contextual information how the current collaboration refines and extends existing enhancements while also adapting and improving new areas of the online *DSAEHist*. Many of the initial changes and future changes are supported by the user survey.

5.1 General layout

The first step in modernising the *DSAEHist* was to rethink the general page

layout of the pilot edition including the landing page. As Figure 1 shows, the pilot's general dictionary environment is basic and uninviting and does not conform to modern online user expectations in terms of functionality and interactivity as both are limited and not multidimensional (Rundell 2015: 310). There are some useful features like search autocompletion, the alphabet bar and user-tabs, but they need to be refined and integrated in innovative ways to make them more functional. The pilot edition's stylistic choices, such as the colour scheme, font choices, etc., are outdated and drab. Du Plessis (2015: 115) mentions that although it is imperative not to alienate dictionary users by creating an unfamiliar dictionary environment, online dictionaries need to have interesting and modern style choices in order to attract users and to conform to certain current standard web practices and norms. The design and screenshot structure has therefore been adapted in the new working model so that the search routes, user guidance, access structure and user interactivity become visible and prominent features of the dictionary. Design changes will include new fonts and a less obtrusive colour scheme.

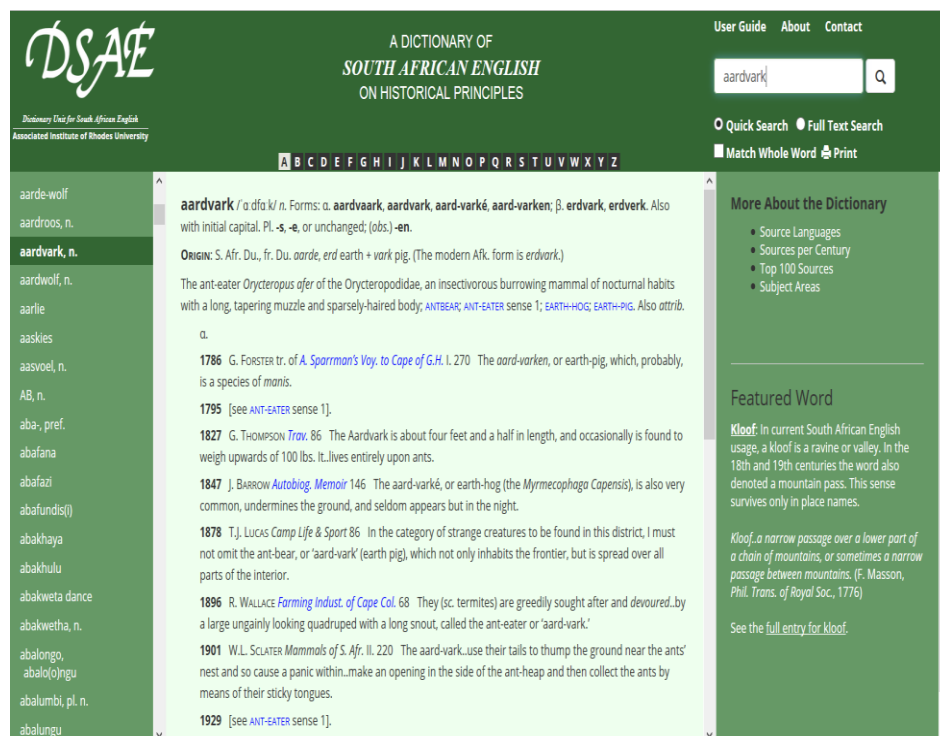


Figure 1: Pilot edition screenshot structure showing entry aardvark

The new working model of the landing page presents the user with a simplified logo and banner and possibly shortened title (pending copyright clarification) on the left of the page (see Figure 2). The newly-modelled logo area removes the abbreviation expansions and the mention of the DSAE's status as an affiliated institute of its host university (this was included originally to acknowledge Rhodes University as the source of initial funding for the pilot edition). Institutional information will henceforth be displayed on the About page, reserving the logo and banner for dictionary branding alone.

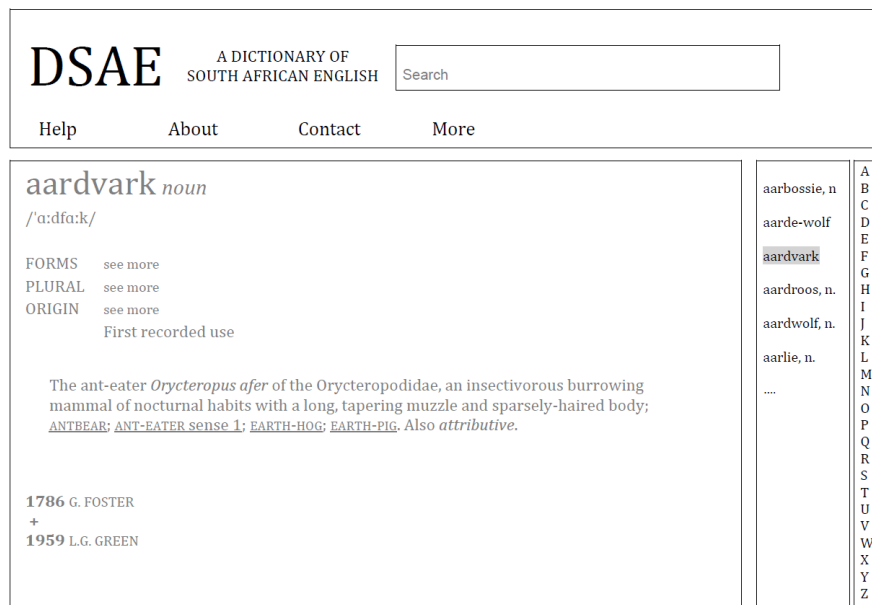


Figure 2: Wireframe prototype of general layout and displayed article

A new lighter colour scheme with readable modern online fonts will be sampled and used. The Print button shown below the search options in Figure 1 will be integrated into the more functional extended microstructure of each displayed dictionary article. Another important change that will improve the functionality and navigability of the landing page, is moving (and enlarging) the search bar to the top-centre of the page. The search bar has become synonymous with web browsing and is the component that online users most frequently access (Terry 2016; Du Plessis 2015: 54). Placing it in a prominent position invites users to start using the dictionary immediately. This change matches current trends in electronic dictionary design and are reflected in the user survey as 50–75% of respondents indicated that clarity, adaptability,

searchability and accessibility are of very high importance, thereby meeting developing user expectations.ⁱⁱ

The new layout also moves menu navigation components (User Guide, About, Contact, and More About the Dictionary content, the latter currently displayed in a side panel) to a set of prominent user-friendly tabs underneath the new search bar and expands on their content (Figure 2). This involves creating easily accessible outer textual tabs to, for example, user guidance, more information about the dictionary and DSAE, and contact information. A label-based Home button will also be introduced. This feature was cut from the pre-publication pilot edition on the basis that the logo functions as a home button, a common convention on the Internet, and that the institutional information under the logo already cluttered the area with text. However, this functionality was insufficient and a Home button is required as an explicit navigation feature to improve usability and navigability, especially for inexperienced online users (Du Plessis 2015: 52, 79).

Online dictionaries should offer multiple search routes to users, especially to target users who are interested in the range of lemmas in the dictionary (Du Plessis 2015: 54-55). With this in mind the alphabet bar and lemma list sidebar were moved to the far-right of the page. The user survey also showed that these components were mostly used by respondents, therefore they cannot be removed. The changes were implemented in order to display the content of greatest interest to the user in the left-centre of the page, in accordance with the F-Layout reading flow pattern followed by most Internet users (see Ambrose and Harris 2010), to provide extra non-intrusive search routes that don't overload the page with data, and to conform to the standard web practice of showing less prominent information, links and options to the right of a webpage (Ambrose and Harris 2010: 168) (see Figure 2).

5.2 User engagement, guidance and feedback features

Given that a historical variety dictionary is unfamiliar to many dictionary users who are better acquainted with general monolingual lookup tools — most South Africans are generally unaware of the extent of South African English — it is also important to offer an engaging initial access point allowing optional exploration of contents which are otherwise opaque on first viewing the dictionary. To facilitate user engagement and improve the general look of the dictionary, low-maintenance but dynamic content (e.g. featured words or interesting facts about South African English) on the landing page is offered as well as under some of the user tabs, with additional dictionary content nested under the More tab. We also propose showing popular searches and user statistics derived from Google Analytics for users to view.

Rather than overloading the user with vast amounts of text, and for the sake of user-friendliness (Tarp 2009: 27), the user guidance will be restructured in the form of frequently asked questions (FAQs), such as "How does the search option work?" or "What do the labels next to the definitions mean?",

with short easy answers that could include step-by-step guidance or example entries. Key FAQs will be incorporated onto the landing page as another means to improve user-friendliness and help new users become accustomed to the page. User interaction can also be gained by including a user feedback form either on the landing page or under the Contact tab since e-mail links are not always effective. Engaging with the user is important to the lexicographic process and adds a mediation function (cf. Fuertes-Olivera 2014) to the dictionary. The possibility of adding images in the form of an image slider were also considered but due to copyright restrictions and added load times to the webpage, as well as the fact that images appear to go against current online dictionary practices, especially for a historical dictionary (Le Du 2016a), this option is under review.

Lastly, as mentioned, dictionary content will be nested under the More tab. Here users will find new dynamic infographics that visually encapsulate and present the user with information on the source languages, sources per century and subject areas, shown as text-heavy tables in the pilot edition. These infographics show interesting extra content that is linked with the dictionary entries. This means that users can find their way to related entries or access the dictionary content in a unique way (see Section 6.2.7 for more information). A multi-layered presentation of data adds to the overall interactivity of the dictionary and is also a different way to navigate through and promote the historical data of the dictionary. These infographics are, however, still a work in progress.

The changes made to the general layout, access structure and screenshot structure inevitably lead to changes in the article and microstructure.

6. Changes to article structure and microstructure

6.1 Initial enhancements

The initial work of restructuring the XML dataset which started during the DSAE-HU collaboration has led to some initial enhancements in the presentation of articles. The following example from *aardvark* in Figures 3 and 4 shows some of these changes (Van Niekerk et al. 2016a):

aardvark /'ɑ:dfɑ:k/ *n.* Forms: α. **aardvaark**, **aardvark**, **aard-varké**, **aard-varken**; β. **erdvark**, **erdverk**. Also with initial capital. Pl. **-s**, **-e**, or unchanged; (*obs.*) **-en**.
ORIGIN: S. Afr. Du., fr. Du. *aarde*, *erd* earth + *vark* pig. (The modern Afk. form is *erdvark*.)
 The ant-eater *Orycteropus afer* of the Orycteropodidae, an insectivorous burrowing mammal of nocturnal habits with a long, tapering muzzle and sparsely-haired body; **ANTBEAR**; **ANT-EATER** sense 1; **EARTH-HOG**; **EARTH-PIG**. Also *attrib.*

Figure 3: Pilot edition before enhancements

aardvark /ɑ:dfɑ:k/ *noun*.

- **FORMS:** α. **aardvaark**, **aardvark**, **aard-varké**, **aard-varken**; β. **erdvark**, **erdverk**. Also with initial capital.

- **INFLECTIONS:** Plural **aardvarke**, **aardvarks**, or unchanged; (*obsolete*) **aardvarken**.

- **ORIGIN:** South African Dutch, from Dutch *aarde*, *erd* earth + *vark* pig. (The modern Afrikaans form is *erdvark*.)

The ant-eater *Orycteropus afer* of the Orycteropodidae, an insectivorous burrowing mammal of nocturnal habits with a long, tapering muzzle and sparsely-haired body; [ANTBEAR](#); [ANT-EATER](#) sense 1; [EARTH-HOG](#); [EARTH-PIG](#). Also *attributive*.

Figure 4: Initial improvements to pilot edition

The 2015 working model shown in Figure 4 reflects changes to content presentation, structure and layout which improve the general usability and readability of data in the *DSAEHist*. These changes include:

- the separation and "labelling of main entry components (forms, plurals and origin) into block-level elements" (Van Niekerk et al. 2016a);
- in the *inflections block*: the manual expansion of abbreviations telescoping such that, for example, "Pl. -s, -e, or unchanged; (*obs.*) -en" is rendered fully as "Plural **aardvarks**, **aardvarke**, or unchanged; (*obsolete*) **aardvarken**";
- in the *etymology block*: the expansion of abbreviations here, where they were numerous, as well as in other areas of the dictionary.

The removal of abbreviations is a substantial undertaking requiring largely manual editorial intervention and proofreading, but this has been newly adopted as a general styling policy for the electronic edition (except in quotations which have their own internal editorial consistency) for two reasons: Firstly, a large set of abbreviations were in use, sometimes for language names or grammatical terms which would be unfamiliar to non-specialists, requiring the user to check against an extensive abbreviation glossary. To remove this burden from the user, and for general readability, abbreviations have generally been removed or are deprecated in styling policy. A second, important reason for removing abbreviations is that they obstruct machine-readability for Internet search engines and the online dictionary's internal search functionality. For example, the search engine keywords "plural aardvark" would be less likely to be matched to the relevant page on the dictionary if it contained only the abbreviated form "Pl."

6.2 Further adaptations: current work

The pilot edition presentation, reflecting the inflexibility of the pilot edition

dataset, displayed the initial entry components as a run-on line as per the print edition:

aardvark /'ɑ:dfɑ:k/ *n.* Forms: α. **aardvaark**, **aardvark**, **aard-varké**, **aard-varken**; β. **erdvark**, **erdverk**. Also with initial capital. Pl. **-s**, **-e**, or unchanged; (*obs.*) **-en**.
ORIGIN: S. Afr. Du., fr. Du. *aarde*, *erd* earth + *vark* pig. (The modern Afk. form is *erdvark*.)

Figure 5: Unadapted pre-definition microstructural components from pilot edition

As shown, some preliminary enhancements to the microstructure and article structure have been investigated and tested. The current article structure is a legacy of the print edition and its ubiquitous space-saving editorial policies, augmented with basic hyperlinks for cross-references between articles and to the bibliography of quotation sources. In order to make the dictionary articles more user-friendly and to showcase the historical data, a repositioning and labelling of data categories (as part of the work done with HU) was undertaken as well as including a more multi-layered or dynamic approach to the presentation of content. Again styling changes have also been employed in the actual dictionary articles:

aardvark *noun*

[/ɑ:dfɑ:k/](#)

FORMS: [aardvaark](#) » [more...](#)

PLURALS: [aardvarke](#) » [more...](#)

ORIGIN: South African Dutch » [more...](#)

[First recorded use 1786](#)

The ant-eater *Orycteropus afer* of the Orycteropodidae, an insectivorous burrowing mammal of nocturnal habits with a long, tapering muzzle and sparsely-haired body; [ANTBEAR](#); [ANT-EATER SENSE 1](#); [EARTH-HOG](#); [EARTH-PIG](#). Also *attributive*.

1786 G. FORSTER tr. of *A. Sparrman's Voy. to Cape of G.H.* I. 270 The *aard-varken*, or earth-pig, which, probably, is a species of *manis*.

[+]

1959 L.G. GREEN tr. of *These Wonders* I. 270 That creature of obscure origin, that champion tunneller of the veld, the erdvark or ant-eater. This pig-shaped freak is not rare, but is seldom captured.

Figure 6: Newly-modelled article layout

The redesigned article layout shown in Figure 6 is made possible only through structural changes to the underlying XML dataset which itself represents a comprehensive map of the dictionary design. Some of this restructuring is complete, and some of it represents current work. In all cases adaptations to the dataset require that structural exceptions be handled, prompting an iterative redesign process. The result of these adaptations is ever-increasing flexibility of

visual and stylistic presentation. The results are as follows, taken in order of the article components.

6.2.1 Headword and part of speech

The headword and part of speech have been restyled so that the headword now occupies a much more prominent position on the page, also functioning as a heading (see Figure 7). This adds a contextual signpost for the user while also playing a role in navigation since the headword and part of speech will, according to the new model, always remain in view when scrolling through the rest of the article contents. An anchored heading-style headword is required here since articles may often be hundreds of words longⁱⁱⁱ and in the pilot edition the headword often disappears from view when scrolling. Research has also prompted the adoption of a mobile-friendly sans serif font (for the headword and part of speech) as it reads more easily and looks similar to fonts used across different popular websites (Beaird 2007: 113). Font choice goes beyond purely visual considerations: the working model illustrated in Figure 7 uses the Google Roboto font for the secondary reason that it can be cached locally on the user's device, thus avoiding a negative performance impact caused by font-loading — one of the ways current design work anticipates the needs of mobile device interfaces.

aardvark *noun*

Figure 7: New lemma and part of speech presentation

The part of speech has been italicised and rendered in a smaller point size than the headword (as shown in Figure 7). This is done as apart from styling, it's also a subtle navigation aid, i.e. it allows the user (once he/she has seen a few entries) to distinguish these components intuitively before even reading the words. In the print edition all parts of speech were abbreviated, except in the case of longer parts of speech. These will now be shown expanded. The heading-style foregrounding of the headword and unabbreviated parts of speech also serve the additional purpose of improving search engine optimisation.

Furthermore, lexicographic and contextual labels will be moved to the start of each sense. A final enhancement in this section relates to unassimilated lexical items, namely borrowings which were not regarded as fully integrated into South African English at the time of publication but were included in the dictionary "in order to document the *S. Afr. Eng.* vocabulary as comprehensively as possible" (Silva et al. 1996: viii). In the print and pilot editions these lemmas are preceded by tramlines (||) as a type of label. This is sometimes confusing or obtrusive, and a colour change or less confusing symbol glossed with

a mouse-over popup will be substituted to improve readability and interpretability.

6.2.2 Pronunciation

The user survey indicated that the pronunciation is a vital component of the dictionary entry and that the IPA symbols are usable to users. The pronunciation has however now been moved to below the headword. This is done to separate components, make the article structure more scannable, and create a more sleek and modern look. The pilot edition's pronunciation indicator is inadequate in that the user must refer to a long table of IPA symbols on another web page in order to make sense of the phonetic notation. In the new model the pronunciation is hyperlinked, so that when a user clicks on it a popup with translations of only the relevant IPA symbols will appear to offer context-specific help (see Figure 8).



Figure 8: Model of new pronunciation popup (prototype)

6.2.3 Variant spellings, inflections and etymology

The data categories variant spellings, inflections and etymology have, after restructuring of the XML markup, been changed from inline to block-level elements. They have also been assigned more easily-understandable labels, namely:

- variant spellings are labelled "Forms"
- inflections are labelled "Plurals"
- etymology is labelled "Origin"

These elements have also been refined so that they don't overwhelm the user with data (even lists of variant spellings are sometimes very long) but are optionally expandable (see Figure 9). 58% of respondents in the user survey indicated that expandable content would be welcomed. Under the "Forms" category, different spelling variation group sections are sometimes indicated using Greek symbols, referring to different quotation paragraphs when these are separated according to spelling patterns (e.g. *aardvaark* has spellings starting with either *a* or *e*, and these are distinguished by having separate spelling variant subsections and quotation paragraphs). These variant subsections will also be restructured into a block presentation with each section being hyperlinked to the relevant quotations. The first recorded use of the headword has also been included and hyperlinked to the first documented quotation under the "Origin" category. These changes have been proposed as a subtle way to emphasise the historical data and cognitive function of the dictionary and to aid interested users in exploring the deeper content of the article further.

FORMS: [aardvaark » more...](#)

PLURALS: [aardvarke » more...](#)

ORIGIN: [South African Dutch » more...](#)

[First recorded use 1786](#)

Figure 9: Expandable newly labelled data categories

6.2.4 Definitions

The paraphrases of meaning have been foregrounded by placing them in a text box or panel. Google Search Analytics statistics indicate that the meaning statement is of key user interest (Van Niekerk et al. 2015). Foregrounding the definitions allows those users not interested in the other historical components of the entry to access the data they need more quickly, rather than being discouraged by surrounding text which is beyond their interests (see Figure 10). A long-term goal here is also to separate the definition data-types, for example into core definitions and cross-references, as another way to facilitate improved presentation of the data.

The ant-eater *Orycteropus afer* of the Orycteropodidae, an insectivorous burrowing mammal of nocturnal habits with a long, tapering muzzle and sparsely-haired body; [ANTBEAR](#); [ANT-EATER sense 1](#); [EARTH-HOG](#); [EARTH-PIG](#). Also *attributive*.

Figure 10: Foregrounded paraphrase of meaning for aardvark

6.2.5 Quotations

Directly below and visually juxtaposed with the paraphrases of meaning are the quotations. As mentioned, the historical data are an integral part of the *DSAEHist* and the goal here is for users to access this section regularly, since they contain additional information which cannot be synthesised in definitions. However, some lemmas have an extensive list of quotations.^{iv} This poses somewhat of a quandary as the historical data needs to be highlighted, but overloading the user with data causes users to look elsewhere for the required information. Therefore to satisfy most users' needs (again linking to the respondents in the user survey's need for expandable content) a content filtering strategy which displays, by default, the first and last quotation per sense, with the option to click a button to expand to the full list, has been developed (also see Van Niekerk et al. 2016a), and the final implementation of this filtering will be refined in the DSAE-SU collaboration. Beyond this, we also propose font and background changes here as a means to i) direct the users' attention to this section, and ii) to make the category of data clear and readable (see Figure 11).

1786 G. FORSTER tr. of [A. Sparrman's Voy. to Cape of G.H.](#) l. 270 The *aard-varken*, or earth-pig, which, probably, is a species of *manis*.



1959 L.G. GREEN tr. of [These Wonders](#) l. 270 That creature of obscure origin, that champion tunneller of the veld, the erdvark or ant-eater. This pig-shaped freak is not rare, but is seldom captured.

Figure 11: Expandable quotations with a dynamic structural indicator

6.2.6 Cross-references and combinations

Related current work includes expanding cross-references between quotations (another print-edition space-saving device which is no longer necessary). For example, under the quotations for **berghaan** in the pilot edition (see Figure 12) the user is sent between these entries for different quotations (1889 and 1867). To give the user immediate access to the data the target quotation will be shown instead (see Figure 13). This change will apply to approximately 2,800

quotations and has involved a combination of semi-automatic updates and case-by-case proofreading, incidentally resulting in improved content where ambiguous or broken cross-references were identified.

berghaan /'berx, ha:n, 'be:g-/ *n.* *obsolescent*.
ORIGIN: S. Afr. Du., *berg* mountain + *haan* cock.

1. The **BLACK EAGLE**, *Aquila verreauxii*.

1844 J. BACKHOUSE *Narr. of Visit* 488 Three species of Eagle occur in South Africa..[including] *Aquila vulturina*, the Berghaan.

1867 [see **DASSIEVANGER**]

Figure 12: Cross-references between quotations in the pilot version

berghaan *noun*
– ORIGIN: South African Dutch, *berg* mountain + *haan* cock.
obsolescent

1. The **BLACK EAGLE**, *Aquila verreauxii*.

1844 J. BACKHOUSE *Narr. of Visit* 488 Three species of Eagle occur in South Africa..[including] *Aquila vulturina*, the Berghaan.

1867 E.L. LAYARD *Birds of S. Afr.* 11 *Aquila Verreauxii*...*Dassie Vanger* and *Berghaan* of Colonists...It is called '*Dassie Vanger*' (coney-eater) and '*Berghaan*' (mountain cock) by the Colonists..from feeding principally on the coney, or rock-rabbit.

Figure 13: Quotation cross-references expanded

Work on presenting combined forms and their respective quotations as separate block-level elements (rather than unbroken paragraphs as in the print edition) is also under way after having been initiated during the DSAE-HU collaboration. Again a semi-automated dataset adaptation is required. The same strategy has been applied to compounds and derivatives, all of which were formerly nested in paragraphs which did not separate these elements — again, to save space in print. For example, the combinations for **biltong** – **biltong-curtain** and **biltong farmer** (Figure 14) will be displayed in a block at the foot of the entry similar to the block for forms, plurals and origin to enhance readability (see Figure 15).

2. *comb.* **biltong curtain** [by analogy with Eng. *Iron Curtain*], a jocular name for the borders of South Africa; **biltong farmer**, **-hunter**, **-jackal**, one who hunts game in order to obtain meat for the making of biltong.

Figure 14: Combinations in pilot edition

2. combinations

- biltong curtain** [by analogy with English *Iron Curtain*], a jocular name for the borders of South Africa;
- biltong farmer, biltong hunter, biltong jackal**, one who hunts game in order to obtain meat for the making of biltong.

Figure 15: Combinations in new layout

6.2.7 Extended microstructure and access structure

As stated in Section 5.2, guiding the user to the historical data contained in the dictionary is an important part of the new working model. An interesting and innovative way to present and link dictionary data is providing users with relevant contextual or subject specific data as part of the extended microstructure. The extended microstructure also functions as an expanded access structure as users will have the option of accessing different entries or parts of the dictionary from an extra search route. Interestingly, 44% of respondents in the user survey indicated that access to related content would be important.

aardvark *noun*

[/ˈɑːdʌrk/](#)

FORMS: **aardvaark** [» more...](#)

PLURALS: **aardvarke** [» more...](#)

ORIGIN: South African Dutch [» more...](#)
[First recorded use 1786](#)

The ant-eater *Orycteropus afer* of the Orycteropodidae, an insectivorous burrowing mammal of nocturnal habits with a long, tapering muzzle and sparsely-haired body; [ANTBEAR](#); [ANT-EATER SENSE 1](#); [EARTH-HOG](#); [EARTH-PIG](#). Also *attributive*.

1786 G. FORSTER tr. of [A. Sparrman's Voy. to Cape of G.H.](#) l. 270 The *aard-varken*, or earth-pig, which, probably, is a species of *manis*.

[+]

1959 L.G. GREEN tr. of [These Wonders](#) l. 270 That creature of obscure origin, that champion tunneller of the veld, the erdvark or ant-eater. This pig-shaped freak is not rare, but is seldom captured.

Other animals:

aardwolf
aasvoel
alikeukel
[» more...](#)

Also first recorded in the 1780s:

eland
Jan Compagnie
lock-shoe
[» more...](#)

Find other entries quoting:

G. Forster
L.G. Green
[» more...](#)

aarde-wolf
aardroos, n.
aardvark, n.
aardwolf, n.
aarlie, n.
aaskies
aasvoel, n.
AB, n.
aba-, pref.
abafana
abafazi
abafundisi
abakhaya
abakhulu
abakweta dance
abakwetha, n.
abalongo, n.
abaloongu
abalumbi, pl. n.
abalungu
abaMbo
abangoma
Abantu,
abaNtu
abanumzane

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Figure 16: Related content shown on the right of the article

The related content found in the extended microstructure will be nested under labelled display blocks in a new sidebar adjacent to the article. These blocks will be expandable and could also be linked to the infographics previously discussed. The labels describing each related content block and the article hyperlinks listed under them (see Figure 16 for an example) can be derived

from categories such as shared origin, shared register or subject specific lexicographic labels as well as other historical and bibliographical data encoded in the dictionary database. Except for comprehensive subject-categories — which are currently being added to supplement existing subject-labels as a prerequisite for advanced search functionality — these semantic categories are already encoded in the dictionary's richly-annotated dataset, and their XML structure makes their values queryable with ease and flexibility. However the implementation of appropriate dynamically generated related content requires careful modelling and testing, and will only be undertaken on completion of current project priorities.

7. Future work

Apart from the ongoing work already mentioned in adapting the XML dataset, general layout and article structure, there are other possible long-term goals that should be pursued.

One area that requires more research is user engagement and involvement. Raising awareness about the *DSAEHist* should become a priority from a marketing perspective. The various ways of promoting the dictionary should also be investigated. One of the objectives of the current adaptation process is specifically to make the dictionary design and user interface appealing to an emerging group of 'digital natives'.

A possibility to increase user involvement in the future is to get users to actively partake in the lexicographic process. This can be done by for example letting users collect data for the dictionary, especially in terms of quotations.

Apart from this possibility, other user-friendly enhancements such as audio options alongside the newly modelled pronunciation indicator should be pursued, as well as the proposal (mentioned in Section 6.2.7) to provide access routes to related content or extra data. In due course, possibilities for offering different search routes, functional and interactive elements and user involvement via an extended microstructure or access structure should also be explored and developed.

The content and functions of outer texts or extra matter, currently occupied by content from the preface of the print version, also need further discussion. This is an area where, for example, external links and important concepts such as the history of South African English can be contained.

Other navigation features such as scroll bars and scrolling methods also need to be tested, as does the integration of other visual components and pop-ups.

Finally, as many of the current adaptations, as well as many of the future possibilities, are in accordance with many of the results of the user survey, especially in terms of adaptability, clarity and access to data, possible follow-up surveys should also be considered for the future.

The work presented here will take time to complete and publish online due to resource constraints and the scale of the undertaking of adapting a his-

torical dictionary originally designed for print to a functional digital online format. Nevertheless this paper presents current work towards bringing a valuable cultural and linguistic resource into the digital era, a process in which the DSAE is being greatly aided by the present collaboration with SU. Such collaborations between research institutions offer the opportunity of showing how guidelines from theoretical lexicography can be applied in practice, and the non-profit model of the project, which allows experimentation and innovation without risk to publishers' commercial imperatives, makes it well-suited for this form of exchange. As such, the intent of this paper is to show the evolution of the dictionary from a print to electronic product to meet evolving user needs, and we renew our invitation to the lexicographical community to provide feedback on the project's current and future development.

8. Endnotes

- i. Please note that this article and any future publications stemming from work done within the DSAE-SU collaboration may be read in conjunction with publications from the DSAE-HU collaboration. See for example Van Niekerk 2016a.
- ii. The changes to search functions are not discussed here as this forms part of the DSAE-HU collaboration (cf. Van Niekerk et al. 2016a).
- iii. Some entries are much longer, e.g. *African* (approximately 2,300 words) and *Hottentot* (approximately 6,000 words).
- iv. On average there are 10 quotations per entry but the count is much higher for many headwords, e.g. *Cape* has 230.
- v. Gary Terry is an IT and web design specialist at Rhodes University.

9. References

- A Dictionary of South African English on Historical Principles*. [Online.] Available: <http://dsae.co.za>.
- Ambrose, G. and P.E. Harris.** 2010. *The Visual Dictionary of Pre-press and Production*. Lausanne: AVA Publishing.
- Beaird, J.** 2007. *The Principles of Beautiful Web Design*. Melbourne: SitePoint.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 17-196.
- Botha, W.F. and H. Fourie-Blair.** 2016. *Demonstration of the Pilot Project for a New Online Woordeboek van die Afrikaanse Taal*. Unpublished paper delivered at the 21st International Conference of the African Association for Lexicography (AFRILEX), 5 July 2016, Tzaneen, South-Africa.
- Nied Curcio, M.** 2015. *The Use of Smartphones in Second-language Teaching*. Colloquium, 3 September 2015, Stellenbosch University, Stellenbosch, South Africa.
- Du Plessis, A.H.** 2015. *'n Analise van die selfoon-WAT: 'n Grondslag vir die verbetering van selfoonwoordeboeke*. Unpublished M.A. thesis. Stellenbosch: Stellenbosch University.

- Fuertes-Olivera, P.A.** 2014. Workshop on Specialised Lexicography. Colloquium, 4 June 2014, Stellenbosch University, Stellenbosch, South Africa.
- Gouws, R.H.** 2005. Oor die verhouding tussen woordeboekstrukture, woordeboekinhoud en leksikografiese funksies. *Lexikos* 15: 52-69.
- Gouws, R.H.** 2016. *Integrating Online Reference Sources into a Comprehensive Dictionary Culture*. Unpublished paper delivered at the 21st International Conference of the African Association for Lexicography (AFRILEX), 5 July 2016, Tzaneen, South Africa.
- Le Du, B.** 2016. Report on User Survey: *Dictionary of South African English on Historical Principles Online*. Dictionary Unit for South African English.
- Le Du, B.** 2016a. Report on Use of Images: *Dictionary of South African English on Historical Principles Online*. Dictionary Unit for South African English.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Silva, P., W. Dore, D. Mantzel, C. Muller and M. Wright (Eds.)**. 1996. *A Dictionary of South African English on Historical Principles*. Cape Town/Oxford/New York: Oxford University Press.
- Tarp, S.** 2009. Beyond Lexicography: New visions and Challenges in the Information Age. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009. *Lexicography at a Crossroads* 90: 17-31. Bern: Peter Lang AG.
- Terry, G.** 2016. Personal interview. 2 March. Grahamstown.†
- Van Niekerk, T.** 2015. *Adaptation of A Dictionary of South African English on Historical Principles for Electronic Platforms*. Unpublished paper delivered at the 20th International Conference of the African Association for Lexicography (AFRILEX), 7 July 2015, University of KwaZulu-Natal, Durban, South Africa.
- Van Niekerk, T., H. Stadler and U. Heid.** 2015. *Data Representation and Data Presentation for Selective Queries to the Dictionary of South African English Online*. Unpublished paper delivered at the Hildesheim Conference on Electronic Dictionaries as Information Tools, 8 October 2015, University of Hildesheim, Hildesheim, Germany.
- Van Niekerk, T., L. Morris, J. Wolvaardt and B. le Du.** 2016. User Survey: *Dictionary of South African English on Historical Principles Online*. Dictionary Unit for South African English.
- Van Niekerk, T., H. Stadler and U. Heid.** 2016a. *Enabling Selective Queries and Adapting Data Display in the Electronic Version of a Historical Dictionary*. Paper delivered at the XVII EURALEX International Congress, 6–10 September 2016, Tbilisi, Georgia. [Online] Available: <http://euralex2016.tsu.ge/publication.html>.

Op pad na 'n omvattende woordeboek-kultuur in die digitale era

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

Abstract: Towards a Comprehensive Dictionary Culture in the Digital Era.

Following a brief introduction this paper offers a discussion of certain aspects of the development of theoretical lexicography. Specific attention is drawn to among others different phases in this development, provision that has to be made for the use of different lexicographic tools and also the user-perspective. The main focus of this contribution is on the development of a comprehensive dictionary culture that unites the traditional division between user-friendliness and dictionary culture in one encompassing comprehensive dictionary culture. A bidirectional relation between society and lexicography is discussed and it is indicated that certain adaptations need to be made in order to improve the comprehensive dictionary culture. This includes both lexicographic and extra-lexicographic adaptations. The distinction between a collective and an individual comprehensive dictionary culture is discussed and it is suggested that dictionary culture should not be a separate main component of dictionary research but should rather be integrated into all four existing main components. Many potential dictionary users within the digital era belong to Generation Z. Some features of this generation are briefly discussed. The need is indicated for an adaptation in lexicography that could motivate this generation to use dictionaries. It is argued that dictionary didactics should play an important role in establishing a comprehensive dictionary culture within Generation Z. This dictionary culture should also be directed at other reference sources.

Keywords: COMPREHENSIVE DICTIONARY CULTURE, DICTIONARY CULTURE, DICTIONARY DIDACTICS, DICTIONARY RESEARCH, DIGITAL ERA, EXTRA-LEXICOGRAPHIC ASPECTS, GENERATION Z, LEXICOGRAPHIC TOOLS, METALEXICOGRAPHY, ONLINE DICTIONARIES, TYPES OF USERS, USER-FRIENDLINESS

Opsomming: Na 'n kort inleiding bied dié artikel 'n bespreking van bepaalde aspekte van die ontwikkeling van die leksikografieteorie. Spesifieke aandag word gegee aan onder meer verskillende fases in dié ontwikkeling, voorsiening wat gemaak moet word vir die gebruik van verskillende leksikografiese werktuie, asook die gebruikersperspektief. Die hooffokus van hierdie artikel is op die ontwikkeling van 'n omvattende woordeboek-kultuur wat die tradisionele verdeling tussen gebruikersvriendelikheid en woordeboek-kultuur onder een oorkoepelende omvattende woordeboek-kultuur verenig. 'n Tweerigtingverhouding tussen samelewing en leksikografie word bespreek en daar word gewys op bepaalde aanpassings wat gemaak moet word ter verbetering van die woordeboek-kultuur. Dit sluit sowel leksikografiese as buiteleksikografiese aanpassings in. Die onderskeid tussen 'n kollektiewe en 'n individuele omvattende woordeboek-kultuur word bespreek

en daar word voorgestel dat woordeboekkultuur nie 'n afsonderlike hoofkomponent van woordeboeknavorsing moet wees nie, maar eerder in al vier die bestaande hoofkomponente geïntegreer moet word. In die digitale era behoort baie potensiële woordeboekgebruikers tot Generasie Z. Kenmerke van hierdie gebruikersgroep word kortliks bespreek. Die behoefte aan 'n aanpassing in die leksikografie om lede van Generasie Z by woordeboekgebruik te betrek, word genoem. Daar word van die standpunt uitgegaan dat woordeboekdidaktiek 'n belangrike rol moet speel in die vestiging van 'n omvattende woordeboekkultuur onder hierdie geslag gebruikers. Dié woordeboekkultuur moet ook op die gebruik van ander naslaanbronne gerig wees.

Sleutelwoorde: AANLYN WOORDEBOEKE, BUIELEKSIKOGRAFIESE ASPEKTE, DIGITALE ERA, GEBRUIKERSTIPES, GEBRUIKERSVRIENDELIKHEID, GENERASIE Z, LEKSIKOGRAFIESE WERKTUIE, METALEKSIKOGRAFIE, OMVATTENDE WOORDEBOEKKULTUUR, WOORDEBOEKDIDAKTIEK, WOORDEBOEKKULTUUR, WOORDEBOEKNAVORSING

1. Inleiding

In die lang geskiedenis van die leksikografiese praktyk is daar blyke van 'n verskeidenheid aanpassings wat gelei het tot die huidige stand van woordeboeke as praktiese gebruiksinstrumente, vergelyk onder meer McArthur (1986) en Gouws, Schweickard en Wiegand (2013). Van die ingrypende aanpassings wat die ontwikkeling van die leksikografie kenmerk, was die oorgang van woordeboeke op kleitablette en papyrusblare na woordeboeke op perkament en later papier, die oorgang vanaf woordeboeke met 'n tematiese ordening na woordeboeke met 'n alfabetiese ordening, die vestiging van 'n preskriptiewe benadering en die geleidelike oorgang vanaf 'n preskriptiewe na 'n deskriptiewe benadering, die oorgang vanaf 'n kaartversameling na 'n elektroniese korpus en, meer onlangs, die oorgang vanaf gedrukte na aanlyn woordeboeke. 'n Groter blootstelling aan en gebruik van woordeboeke het ook 'n oorgang vanaf 'n pre-woordeboekkultuurera na 'n era gekenmerk deur 'n meer gevestigde woordeboekkultuur meebring.

Hierdie aanpassings het tot eiesoortige vernuwing in die leksikografie gelei — vernuwing ten opsigte van medium, ten opsigte van inhoud en werkswyse en in sommige gevalle ook ten opsigte van die geldende teoretiese benadering. In die aktiewe wisselwerking tussen leksikografiëteorie en die leksikografiëpraktyk is dit insiggewend om te let op die rol wat die teoretiese leksikografie in hierdie oorgange gespeel of nie gespeel het nie. Met die praktyk wat veel ouer as die teorie is, was daar talle aanpassings wat uitsluitlik op die praktyk gerig en deur praktiese behoeftes bepaal is, soos onder meer die oorgang na perkament en papier en die benutting van die alfabetiese ordening. In ander gevalle het die leksikografiëteorie, of minstens bepaalde vroeë teoretiese insigte, wel 'n rol gespeel. 'n Voorbeeld hiervan was die vestiging van 'n preskriptiewe benadering in Samuel Johnson se *A Dictionary of the English Language* (1755). Die werkswyse wat hy gevolg het en die vaslegging van sy bena-

dering van "fixing the language" is voorafgegaan deur die dokument wat hy in 1747 gepubliseer het, te wete sy bekende *The Plan of a Dictionary of the English Language* — 'n vroeë voorbeeld van 'n teoretiese uiteensetting wat in die praktyk toegepas kon word.

In onlangse jare, lank na die vestiging van die teoretiese leksikografie, is die oorgang na aanlyn woordeboeke vergesel van pogings om die heersende leksikografieteorie wat veral op gedrukte woordeboeke gerig was, uit te brei en aan te pas om ook as teoretiese basis vir aanlyn woordeboeke te dien. Die oorgang na die aanlyn leksikografie het 'n ingrypende invloed op die ontwikkeling in die teoretiese leksikografie gehad, aangesien die nuwe medium nuwe moontlikhede bied maar ook andersoortige eise stel wat nie sonder meer deur die algemene leksikografieteorie in sy gerigtheid op gedrukte woordeboeke bevredig kan word nie. Daar moes besluit word of daar 'n nuwe teorie vir aanlyn woordeboeke ontwerp moet word en of die bestaande teorie aangepas en uitgebrei moet word om vir sowel gedrukte as aanlyn woordeboeke voorsiening te maak. Die laasgenoemde keuse impliseer dat sommige aspekte van die teorie net op gedrukte, ander op gedrukte en aanlyn en ander slegs op aanlyn woordeboeke van toepassing sal wees. So 'n algemene teorie sal sterker wees as twee afsonderlike teorieë. Dit is die uitgangspunt wat in hierdie artikel gehandhaaf word.

Omdat die praktiese leksikografie veel ouer as die teoretiese leksikografie is, het die teorie baie lank 'n navolgersrol gespeel wat 'n beskrywing moes bied van dit wat reeds in die praktyk plaasgevind het. Algaande het die teorie ingehaal en mettertyd ook 'n leiersrol begin speel deurdat voorstelle gemaak kon word vir vernuwende werk in die leksikografiese praktyk.

Die koms van die digitale era het 'n soortgelyke verhouding tussen praktyk en teorie in die leksikografie tot gevolg gehad as wat vroeër die geval in die ontwikkeling van gedrukte woordeboeke was. Aanlyn woordeboeke wat ontwikkel is, het dikwels nie genoegsaam blyke gegee van 'n metaleksikografiese grondslag nie. Een van die redes hiervoor is daarin geleë dat metaleksikograwe nie voldoende aandag gegee het aan die aanpassings wat in die leksikografieteorie gemaak moes word nie. Ook hier het daar mettertyd 'n verandering ingetree wat weer eens algaande die navolgersrol van die metaleksikografie tot 'n leiersrol omskep het. Die leksikografieteorie maak naamlik toenemend voorsiening vir 'n algemener benadering wat nie mediumspesifiek is nie en wat ook as basis vir die beplanning en samestelling van aanlyn woordeboeke kan dien.

Die oorgang na die aanlyn leksikografie is die gevolg van die koms van die digitale era. Dit is 'n omvattende oorgang in eie reg; nie net in die leksikografie nie. In die leksikografie het dit tot baie aanpassings gelei wat nie hier uitvoerig bespreek word nie. Een groot aanpassing wat nog nie gemaak is nie, maar waarvoor daar in hierdie artikel gepleit word, is die aanpassing in die interpretasie van die begrip *woordeboekkultuur* en die implikasies daarvan vir onder meer woordeboekdidaktiek en die gevolglike ontwikkeling van sowel die teoretiese as die praktiese leksikografie. Dit staan sentraal in hierdie artikel.

2. Ontwikkeling in die leksikografieteorie

2.1 Teorie en praktyk

Leksikografie val in 'n teoretiese en 'n praktiese komponent uiteen. Die leksikografiese praktyk was eerste daar en ouer woordeboeke het in 'n preteoretiese era tot stand gekom. Met die koms van die leksikografieteorie, veral in die laaste drie dekades van die vorige eeu, moes metaleksikograwe eerstens terugskouend werk om 'n teoretiese beskrywing te bied van dit wat reeds in die praktyk beslag gekry het. Mettertyd is die nodige inhaalwerk gedoen en kon teoretici ook vorentoe kyk en vernuwende modelle ontwikkel wat leiding aan praktiserende leksikograwe gebied het vir die beplanning en ontwerp van leksikografiese produkte. Die kontemplatiewe benadering van vroeë metaleksikograwe is ondersteun deur 'n transformatiewe benadering, vergelyk Tarp (2008: 10) vir 'n bespreking van hierdie begrippe.

Die koms van die digitale era het al gou tot die daarstelling van 'n verskeidenheid e-woordeboeke aanleiding gegee. Die eerste woordeboeke wat in die digitale medium saamgestel is, was woordeboeke op CD-ROM. Hierdie woordeboeke was in die meeste gevalle slegs digitale weergawes van bestaande gedrukte woordeboeke. Daar was weliswaar 'n bepaalde toegevoegde waarde, onder meer die vinniger toegangsmoontlikhede, maar die aard en omvang van die woordeboekinhoud en die aanbieding daarvan het nie veel verskil van dié van hierdie woordeboeke se gedrukte teenhangers nie. Wat wel opmerklik was in hierdie fase van die oorgang van gedrukte na e-woordeboeke was die gebrek aan teoretiese bydraes met voorstelle oor hoe hierdie e-woordeboeke ten beste beplan en saamgestel kon word. Die praktyk het weer eens die voortou geneem met uitgewers wat bestaande gedrukte woordeboeke in CD-ROM-formaat beskikbaar gestel het sonder dat metaleksikograwe genoegsaam geraadpleeg is of hoegenaamd leiding gegee het oor nuwe teoretiese modelle wat in die praktyk toegepas kon word.

'n Volgende fase in die ontwikkeling van e-woordeboeke was die toetrede van aanlyn woordeboeke en dit was die aanbreek van die werklike oorgang vanaf gedrukte na digitale woordeboeke. Sommige aanlyn woordeboeke is steeds nie veel meer as digitale weergawes van gedrukte woordeboeke nie. Die vernuwende aanlyn woordeboeke is daardie produkte wat nie 'n gedrukte teenhanger het nie en wat van meet af as aanlyn produk beplan en saamgestel is. Ook hier is die verhouding tussen teorie en praktyk ter sake. Die tegnologie om digitale woordeboeke te maak, het vinniger ontwikkel as die pogings van metaleksikograwe om die leksikografieteorie so aan te pas dat dit genoegsaam vir produkte in die nuwe medium voorsiening gemaak het. Weer eens het die teorie 'n navolger van die praktyk geword en nie 'n leier nie.

2.2 Fases in die ontwikkeling van die leksikografieteorie

In die ontwikkeling van die algemene leksikografieteorie was daar verskillende

fases met verskillende fokusgebiede, te wete 'n fokus op woordeboekinhoud, op woordeboekstrukture en op woordeboekfunksies, vergelyk Gouws (2005). Die formulering van leksikografiese funksies, vgl. Tarp (2008), is op so 'n wyse gedoen dat dit nie mediumsies is nie. Hierdie teoretiese benadering geld gevolglik én vir gedrukte én vir aanlyn woordeboeke. Woordeboekinhoud word grootliks deur die behoeftes van die teikengebruiker bepaal. Aanlyn woordeboeke bied in talle gevalle dieselfde inhoud as gedrukte woordeboeke. Die groot verskil lê in die aanbieding van data wat dit in aanlyn woordeboeke makliker maak vir gebruikers om die verlangde inligting te kan onttrek. Een van dié wesenlike aanpassings wat in die leksikografieteorie gemaak moet word om ook vir aanlyn woordeboeke voorsiening te maak, is op die vlak van woordeboekstrukture.

Sedert die tagtigerjare van die vorige eeu het woordeboekstrukture, veral na aanleiding van die werk van die Duitse metaleksikograaf Herbert Ernst Wiegand, sterk op die voorgrond getree op die terrein van leksikografiese navorsing. Saakmakende publikasies is onder meer Wiegand (1989; 1989a; 1989b; 1996; 1998; 2003; 2008). In die onlangse metaleksikografiese navorsing is daar voortgegaan met werk ten opsigte van woordeboekstrukture vir gedrukte woordeboeke, vergelyk onder meer Wiegand en Beer (2013), Wiegand en Smit (2013), Wiegand, Beer en Gouws (2013), Wiegand, Feinauer en Gouws (2013) en Wiegand en Gouws (2013).

Naas die voortgesette gerigtheid op 'n bespreking van strukture in gedrukte woordeboeke is daar eweneens aandag gegee aan van die moontlike aanpassings van woordeboekstrukture wat vir die benutting daarvan in die aanlyn woordeboeke geskik sal wees. Vergelyk in hierdie verband onder meer Klosa (2009), Müller-Spitzer (2013), Gouws (2014; 2014a; 2015), Caruso (2015), Giacomini (2015), Klosa en Gouws (2015) en Müller-Spitzer en Wolfer (2015). In hierdie navorsing oor woordeboekstrukture vir aanlyn woordeboeke is dit veral aanpassings met betrekking tot die dataverspreidingsstruktuur, raamstruktuur, skermkootstruktuur, makrostruktuur, mikrostruktuur, artikelstruktuur en adreseringstruktuur wat ter sprake gebring is. Gouws en Tarp (2016) dui 'n verskeidenheid vernuwende kenmerke in aanlyn woordeboeke aan, onder meer gesofistikeerde soekmetodes en toegangsroetes, gebruiker-gebaseerde datasifting, die gebruik van versteekte data, die gebruik van opskietvensters en nuwe tipes interne en eksterne skakels. Die voorkoms van hierdie kenmerke en 'n beskrywing wat 'n sistematiese benutting daarvan moontlik maak, vereis ook bepaalde aanpassings in sekere woordeboekstrukture.

Daar is ook reeds heelwat navorsing gedoen om die algemene leksikografie te verruim met 'n bespreking van aspekte wat uitsluitlik op aanlyn woordeboeke gerig is. Vergelyk in hierdie verband die tematiese afdeling van die 2015-uitgawe van die vaktydskrif *Lexicographica* wat gewy is aan strukture in aanlyn woordeboeke (Wiegand 2015).

Belangrik is dit om daarop te let dat hierdie aanpassings in die teorie onder meer dit ten doel het om uiteindelik een algemene en niemediumsiesieke leksikografieteorie te kan formuleer.

2.3 Uitbreiding van die teorie ter wille van verskillende leksikografiese werktuie

Die oorgang na die digitale era het ook tot die benutting van verskillende tipes leksikografiese werktuie gelei. Anders as in die era van gedrukte woordeboeke waar die boek die werktuig was waarin die leksikografiese aanbod weergegee is, sluit die werktuie van die digitale era onder meer rekenaars, tablette en slimfone in. Toegang tot aanlyn woordeboeke kan met behulp van hierdie verskillende werktuie geskied. Waar die aard van die boek as werktuig selde 'n gesprekspunt was in die leksikografieteorie wat op gedrukte woordeboeke gerig was, speel die werktuig en die voordele en beperkinge van verskillende werktuie 'n beduidende rol in die digitale era. Die benutting van 'n bepaalde werktuig vir die daarstelling van 'n bepaalde woordeboek mag nie op 'n arbitrêre manier geskied nie. Die naslaan- en leksikografiese vaardighede van potensiele gebruikers moet ook deur leksikograwe in aanmerking geneem word wanneer daar besluit word oor die werktuig en die aard en omvang van die leksikografiese data wat daarin aangebied word. Verruiming in die bestek van die leksikografieteorie maak ook vir allerlei aspekte van leksikografiese werktuie se beperkings en voordele voorsiening. Dit vereis 'n bykomende uitbreiding in die leksikografieteorie.

'n Aanlyn woordeboek wat saamgestel word met die oog op 'n slimfoon as werktuig se strukture sal verskil van dié van 'n woordeboek wat met behulp van 'n rekenaar geraadpleeg word. Die interne toegangstruktuur moet byvoorbeeld minder divers wees met 'n kleiner verskeidenheid tipografiese en nie-tipografiese struktuurmerkers in 'n verkleinde artikelstruktuur met minder soeksones en 'n beperkter aanduideraanbod. Op die interne soekroete moet daar wel voldoende leiding aan gebruikers gebied word om 'n suksesvolle inligtingsonttrekking per woordeboekraadpleging te verseker. Die keuse van 'n gepaste mikrostruktuurtype is hier belangrik. Indien teksbegrip as kommunikatiewe funksie die enigste funksie van die betrokke woordeboek is, is 'n niegeïntegreerde of selfs 'n rudimentêre mikrostruktuur voldoende. Waar teksproduksie die enigste funksie is, sal 'n geïntegreerde mikrostruktuur help om sukses te verseker. Indien die woordeboek bifunksioneel is met teksbegrip en teksproduksie as kommunikatiewe funksies, kan 'n semi-geïntegreerde mikrostruktuur, vergelyk Wiegand (1996), van waarde wees. Vir teksbegrip sal die geïntegreerde teksblok wat veral vertaalekwivalent- of betekenisparafrase-aanduiders en beperkte koteksaanduiders bevat die nodige leiding kan verskaf. Vir die teksproduksiefunksie van so 'n slimfoonwoordeboek sal koteksaanduiders soos voorbeeldmateriaal- en kollokasie-aanduiders in die niegeïntegreerde blok aangebied word.

In 'n woordeboek wat per rekenaar geraadpleeg word, sal die teksproduksiefunksie deur middel van voorbeeldmateriaalaanduiders in 'n geïntegreerde mikrostruktuur uitgevoer kan word. 'n Slimfoonwoordeboek moet die tipiese gebruikssituasie ter wille wees — daarom kan die artikelstruktuur aangepas word

om 'n absolute verpligte mikrostruktuur, vergelyk Hausmann en Wiegand (1989: 346), oftewel 'n verpligte mikrostruktuur (Gouws 2003: 4) as verstekkomponent te hê. Daarbenewens kan daar 'n uitgebreide verpligte mikrostruktuur wees (vergelyk Gouws 2003: 4) wat op tweede en verdere skermkote aangebied kan word en wat bykomende data bevat.

Ruimte is en was dikwels 'n probleem in die samestelling van gedrukte woordeboeke en leksikograwe moes die inhoud van hulle woordeboeke dien-ooreenkomstig beplan. Waar die ruimtebeperking oorskry is deur die data-aanbod, moes die data-aanbod verminder word of daar moes voorsiening gemaak word vir twee- of meerdelige woordeboeke. In die beplanning van die dataverspreidingsstruktuur, die lemmakeuse en die aard en omvang van die mikrostrukturele aanbod het die fisiese ruimtebeperkinge van 'n boek as leksikografiese werktuig 'n sekere rol gespeel.

Een van die voordele, hetsy 'n gewaande hetsy 'n werklike voordeel, van aanlyn woordeboeke is die haas onbeperkte ruimte. Dit is so dat ruimtebeperking nie deurgaans dieselfde rol speel was wat in die beplanning en samestelling van gedrukte woordeboeke die geval is nie, maar onbeperkte ruimte geld nie altyd en oral in alle aanlyn woordeboeke nie. Lew (Om te verskyn) maak 'n onderskeid tussen die feitlik onbeperkte bergingsruimte in 'n aanlyn databasis en die beperkte aanbiedingsruimte, dit is die ruimte op 'n rekenaar- of ander skerm waar die data vertoon word. Waar 'n woordeboek vir gebruik op 'n slimfoon beplan word, moet die leksikograwe die beperkinge van die werktuig met betrekking tot aanbiedingsruimte in gedagte hou. In die eerste fases van die ontwikkeling van Mobilex, 'n webgebaseerde selfoontoepassing wat gebruikers toegang gee tot 'n leksikografiese bewerking van 'n aantal terme uit enkele vakgebiede, is die data op 'n enkele skermkoot beperk tot 250 karakters, vergelyk Van der Merwe (2015). In die beplanning en samestelling van 'n woordeboek wat met behulp van 'n slimfoon geraadpleeg moet word, is dit belangrik dat daar gewaak word teen data-oorklading, vergelyk Gouws en Tarp (2016). In die leksikografieteorie moet 'n verskynsel soos data-oorklading ook verreken word — nog 'n vernuwing wat in die ontwikkeling van die leksikografie en die oorgang na 'n digitale era ter sake raak.

Die gebruik van 'n slimfoon of selfs 'n tablet of rekenaar waarvan die skerm 'n beperkte aanbiedingsruimte het, impliseer nie noodwendig 'n beperkte leksikografiese bewerking nie. In die aanpassing van die teorie oor woordeboekstrukture het 'n fokus op die artikel- en skermkootstruktuur, vergelyk Gouws (2014), gewys op die waarde van 'n veelvlakkige artikelstruktuur waarvan al die data nie in 'n enkele skermkootstruktuur aangebied word nie. Vergelyk ook Du Plessis (2015). Die aard van die werktuig het dikwels 'n invloed op die gepaste strukture wat gebruik word.

2.4 Die gebruiker

Wiegand (1989c: 251) voer aan dat die leksikografie 'n praktyk is wat daarop

gerig is om woordeboeke tot stand te bring sodat 'n verdere praktyk, te wete die kulturele praktyk van woordeboekgebruik moontlik gemaak kan word. Woordeboeke is praktiese gebruiksinstrumente en die suksesvolle gebruik van hierdie instrumente is 'n kerndoelstelling van die beplanning en samestelling van woordeboeke. Gevolglik is navorsing oor woordeboekgebruik 'n belangrike onderafdeling van die metaleksikografie, vergelyk onder meer Wiegand (1984: 15; 1998: 113). Vanuit hierdie navorsing word woordeboekmodelle geformuleer wat tot beter woordeboekgebruik kan lei.

In die ontwikkeling van die leksikografieteorie was daar algaande 'n sterker fokus op die gebruiker, met 'n toename in navorsing oor woordeboekgebruik en woordeboekgebruikers. Vergelyk onder meer Hartmann (1987), Atkins (1998), Tono (2001), Engelberg en Lemnitzer (2009), Svensén (2009), Welker (2013; 2013a). Gevolglik het Wiegand (1977: 59) se benoeming van die woordeboekgebruiker as die "bekende onbekende" aan geldigheid begin inboet. Die fokus op die gebruiker moet in die leksikografieteorie se aanpassing by e-woordeboeke gehandhaaf en selfs verder uitgebou word.

Soos in die geval van gedrukte woordeboeke staan die gebruiker sentraal in die beplanning en samestelling van aanlyn woordeboeke. Wiegand (1984: 15) het reeds woordeboekgebruik as 'n kernkomponent van die metaleksikografie geïdentifiseer en Hartmann (1989) het die beginsel van 'n gebruikersperspektief aan die orde gestel. Deur die daaropvolgende dekades was hierdie perspektief van kernwaarde in sowel die teoretiese as die praktiese leksikografie. In die metaleksikografie word die belang van die identifisering van 'n teikengebruikersgroep en die vasstelling van die leksikografiese behoeftes en naslaanvaardighede van daardie groep as voorvereistes gestel om die uiteindelijke suksesvolle woordeboekgebruik te help verseker.

In die ontwikkeling van die leksikografieteorie het die erkenning aan die belang van die gebruikersperspektief tot bepaalde aanpassings gelei — weer eens in sowel die teorie as die praktyk. Hausmann (1989: 13) het reeds *gebruikersvriendelikheid* en *woordeboekkultuur* as leksikografiese terme gevestig. Gebruikersvriendelikheid in die leksikografie impliseer dat die leksikografie by die gemeenskap moet aanpas deur aan die gebruikers woordeboeke te lewer wat toeganklik is en waaraan inligting onttrek kan word ter bevrediging van hulle spesifieke leksikografiese behoeftes. Volgens Klotz en Herbst (2016: 33) is twee van die belangrikste faktore wat bydra tot gebruikersvriendelikheid in woordeboeke eerstens 'n duidelike uitleg wat maklik interpreteerbaar is en wat duidelik tussen verskillende inligtingtipes onderskei en, tweedens, 'n duidelik intuïtiewe enkodering van inligting, veral die vermyding van obskure afkortings en kodes wat slegs verstaan kan word na die raadpleging van die woordeboek se voor- of agtertekste. Om 'n woordeboekkultuur gevestig te kry, moet die gemeenskap by die leksikografie aanpas deur vertrouwd te raak met onder meer verskillende woordeboeksoorte, die tipiese data wat in woordeboeke aangebied word, die waarde en beperkinge van verskillende woordeboeksoorte en met naslaanvaardighede om woordeboeke suksesvol te kan gebruik. 'n Woor-

deboekkultuur sluit 'n samelewing se kennis van en vertrouwdheid met die leksikografie in asook die vermoë om woordeboeke op 'n kundige manier te kan gebruik. Vergelyk in hierdie verband ook Gouws en Prinsloo (2005: 12), Hartmann en James (1998) en Nkomo (2015). Waar 'n woordeboekkultuur in 'n bepaalde samelewing of taalgemeenskap geld, sal daar kundige woordeboekgebruikers wees, vergelyk Wiegand (1998: 506).

Die voortgesette ontwikkeling van die leksikografieteorie moet ook voor-siening maak vir en aanpas by die veranderde omgewing waarin woordeboeke gebruik word, spesifiek die digitale era, en die potensiële gebruikers van woordeboeke in hierdie era. Alle fasette van die leksikografieteorie, ook die gebruiksperspektief, moet telkens herbekyk en waar nodig aangepas word.

Al die veranderende invloede in die leksikografiese praktyk en teorie het 'n invloed op die woordeboekkultuur. Dit is daarom belangrik om vas te stel of die tradisionele bestek van die begrip *woordeboekkultuur* nog voldoende is en of daar aanpassings gemaak moet word in die interpretasie van hierdie begrip. Dit word vervolgens bespreek.

3. 'n Omvattende woordeboekkultuur

3.1 'n Oorkoepelende term

In verskeie publikasies, onder meer Gouws (2012; 2013; 2016; 2016a) is daar reeds gepleit vir 'n breër interpretasie van die begrip *woordeboekkultuur*. Gouws (2013; 2016: 108) voer aan dat die term *woordeboekkultuur* veral geïnterpreteer word met betrekking tot woordeboeke vir algemene taalgebruik. Dit is die woordeboeksoort waarmee die meeste woordeboekgebruikers bekend is. Die gewone lid van 'n taalgemeenskap is veel minder bekend met ander woordeboeksoorte, onder meer die groot verskeidenheid vakwoordeboeke en hulle rol as naslaaninstrumente. Hierdie beperkte benadering tot woordeboekkultuur is te wagte as daar in ag geneem word dat woordeboeknavorsing, die metaleksikografie, dikwels 'n wetenskaplike aktiwiteit is wat slegs op algemene taalwoordeboeke gerig is. Vergelyk in hierdie verband Wiegand (1989: 257). Wiegand lewer 'n pleidooi vir die verbreding van kennis oor die leksikografie en die wetenskaplike studie daarvan, vergelyk Wiegand (1989: 258). 'n Vergelykbare verbreding kan ook ten opsigte van die begrip *woordeboekkultuur* en navorsing oor hierdie komponent van die leksikografie verlang word. 'n Omvattender siening van 'n woordeboekkultuur sal die samelewing se vertrouwdheid met woordeboeke verder as 'n kennis van slegs algemene taalwoordeboeke laat strek. Die aanpassing van die samelewing by die leksikografie moet gebruikers ook in aanraking bring met en hulle kennis verbreed van ander woordeboeksoorte, onder meer vakwoordeboeke.

'n Uitbreiding in die bestek van 'n woordeboekkultuur sluit ook ander fasette van die leksikografie in wat nie vroeër as deel van die woordeboekkultuur beskou is nie. Gouws (2016a) bespreek verskillende aspekte van woorde-

boekkritiek en wys onder meer op die verantwoordelikheid van die metaleksikograaf teenoor die gemeenskap om deur middel van resensies die woordeboekgebruikende gemeenskap leiding te gee oor die gehalte en beskikbaarheid van woordeboeke. Woordeboekkritiek is 'n formele komponent van die metaleksikografie, vergelyk onder meer Wiegand (1984: 15; 1989: 262; 1998: 113). 'n Omvattender siening van woordeboekkultuur sluit ook die rol van woordeboekkritiek en 'n deeglike kennis van die rol en waarde van woordeboekkritiek by sowel die gebruiker as die leksikograaf in.

Die bestek van 'n omvattende woordeboekkultuur behoort nog verder te strek as byvoorbeeld 'n fokusverbreding om ook aan ander woordeboeksoorte en aan woordeboekkritiek aandag te gee. Wiegand (1998: 587) pleit ten gunste van die vestiging van die metaleksikografiese term *gebruikersvriendelikheid* maar laat hom nie in dieselfde mate oor die term *woordeboekkultuur* uit nie. Teenoor die term *gebruikersvriendelikheid* wat gereeld ook buite die leksikografie gebruik word, is die term *woordeboekkultuur* meer spesifiek leksikografies van aard. Vanuit hierdie terminologiese tweedeling kan die laasgenoemde term eerder as voorkeurterm in die leksikografie gebruik word, maar dan wel op so 'n manier dat dit as 'n oorkoepelende term dien vir dit wat tans as woordeboekkultuur en gebruikersvriendelikheid bekendstaan. Met die term *omvattende woordeboekkultuur* word daar voortaan in hierdie artikel na die tweerigtingverhouding tussen leksikografie en die gemeenskap verwys. Daar word gepleit vir 'n breë interpretasie van hierdie term wat veel meer insluit as wat tradisioneel as ter sake beskou is vir die suksesvolle verhouding tussen die gemeenskap en die leksikografie.

Die belang van 'n omvattende woordeboekkultuur word so hoog geskat dat die invloed daarvan op 'n algemene leksikografieteorie erken behoort te word. Deur hierdie invloed word daar vanuit die metaleksikografie riglyne aan die leksikografiepraktyk gegee ter daarstelling van woordeboeke wat vir 'n veranderende gebruikersgroep se veranderende behoeftes in 'n veranderende woordeboekgebruiksomgewing en vir 'n veranderde medium en werktuigtipe voorsiening maak.

Naas 'n erkenning van die invloed van woordeboekkultuur op die leksikografieteorie kan 'n studie van woordeboekkultuur 'n formele onderafdeling van 'n algemene leksikografieteorie word. In die uiteensetting van die bestek van die metaleksikografie word daar onder die kategorie *woordeboeknavorsing* tans vir vier hoofkomponente voorsiening gemaak, naamlik woordeboekgebruiknavorsing, kritiese woordeboeknavorsing, historiese woordeboeknavorsing en sistematiese woordeboeknavorsing, vergelyk Wiegand (1998: 118) en Schierholz (2016: 306). Om die leksikografie se verantwoordelikheid teenoor 'n woordeboekkultuur deurgevoer te kry, moet woordeboekkultuur nie as 'n afsonderlike komponent van woordeboeknavorsing gesien word nie. Dit is veel eerder 'n afdeling van die leksikografieteorie wat in elkeen van hierdie hoofkomponente geïntegreer is omdat die navorsing wat ten opsigte van elkeen van hierdie komponente gedoen word dit onder meer ten doel het om die leksiko-

grafie beter by die samelewing te laat aanpas. 'n Omvattende woordeboekkultuur is geïntegreer in 'n algemene leksikografieteorie.

3.2 Individuele en kollektiewe woordeboekkultuur

Die bestaande terme *gebruikersvriendelikheid* en *woordeboekkultuur* word onderafdelings van 'n *omvattende woordeboekkultuur*. Uitbreiding van die bestek van die term *omvattende woordeboekkultuur* geld ten opsigte van sowel dit wat vroeër onder *gebruikersvriendelikheid* as onder *woordeboekkultuur* verstaan is. In die verhouding tussen leksikografie en samelewing moet die leksikografiepraktyk, met die leksikografieteorie as basis, produkte lewer wat die gebruiker meer ter wille is. Die keersy van die wisselwerkende verhouding is dat die vlak van woordeboekkultuur in 'n betrokke taalgemeenskap dit vir leksikograwe moontlik maak om meer gesofistikeerde woordeboeke te kan maak sonder vrees dat die potensiële teikengebruiker dit nie optimaal sal kan benut nie. Daar moet deur albei rolspelers in die aktiewe wisselwerkende verhouding tussen samelewing en leksikografie gepoog word om 'n omvattende woordeboekkultuur so wyd moontlik in die samelewing gevestig te kry.

Gouws (2013: 51) maak voorsiening vir die onderskeid tussen 'n individuele en 'n samelewings- of kollektiewe woordeboekkultuur. Binne 'n samelewing of taalgemeenskap wat gekenmerk word deur die afwesigheid van 'n woordeboekkultuur kan daar individue wees wat wel oor 'n woordeboekkultuur beskik. Dit sal 'n voorbeeld van individuele woordeboekkultuur wees. Waar daar 'n wydverspreide woordeboekkultuur in die samelewing geld, is dit 'n samelewings- of kollektiewe woordeboekkultuur. Die ideaal is dat 'n omvattende woordeboekkultuur ook 'n kollektiewe woordeboekkultuur is waaraan die grootste deel van 'n bepaalde samelewing of taalgemeenskap deel het. Een van die kenmerke van die digitale era is die grootskaalse toegang tot data deur 'n groot aantal gebruikers. 'n Kollektiewe woordeboekkultuur is ook hier nodig om die gebruikers van digitale woordeboeke en ander naslaanbronne vertrouwd te maak met die aard en omvang van die inligtingonttrekking aan aanlyn woordeboeke wat moontlik is. Hier gaan dit veral om die verskeidenheid data-tipes en die gemak van toegang tot hierdie data wat vanuit 'n omvattende kollektiewe woordeboekkultuur bekendgemaak moet word.

3.3 Aanpassings in die samelewing

'n Omvattende woordeboekkultuur impliseer 'n tweerigtingwerking tussen leksikografie en samelewing. Aanpassings in die samelewing kan implikasies vir die leksikografie hê en aanpassings in die leksikografie kan implikasies vir die samelewing hê. Sulke aanpassings moet ook in die konseptualisering en formulering van die woordeboekkultuur weerspieël word. Van die veranderinge in die samelewing wat hier 'n rol kan speel, is onder meer 'n verskuiwing

in die teikengebruikersgroepe van woordeboeke. Daar was byvoorbeeld in die voorafgaande dekades 'n uitbreiding in die woordeboektipologie ten gunste van aanleerderwoordeboeke omdat taalaanleerders 'n belangrike deel van die samelewing geword het en daarom as 'n teikengebruikersgroep van woordeboeke geïdentifiseer is. 'n Woordeboekkultuur moes dus ook voorsiening maak vir aspekte wat deur so 'n verskuiwing van teikengebruikersgroepe meegebring is.

Die werktuie wat gebruik word in die digitale era het hulle eie voordele en beperkinge en dit verg 'n eiesoortige tipe opleiding in woordeboekgebruiksvaardighede — 'n faset wat 'n komponent van 'n omvattende woordeboekkultuur behoort te wees en wat tot sekere aanpassings moet lei.

Leksikograwe moet bewus wees van wie die teikengebruikers van hulle spesifieke woordeboeke is maar ook wat die wyer maatskaplike posisie en behoeftes van hierdie gebruikers is. Daar is ongetwyfeld buiteleksikografiese aspekte binne die samelewing wat 'n wesenlike invloed op die leksikografie kan uitoefen. 'n Voorbeeld van so 'n buiteleksikografiese aspek is die tipiese gebruikssituasie van 'n beplande woordeboek. 'n Woordeboek vir gebruikers wat leerders in 'n omgewing is met onvoldoende fisiese geriewe sal heel moontlik anders lyk as 'n woordeboek van vergelykbare aard en omvang vir gebruikers in 'n gebruiksomgewing met optimale fisiese geriewe. Eersgenoemde woordeboek mag dalk primêr deur gebruikers wat in plakkerskampe woon, gebruik word. Die niebeskikbaarheid van voldoende elektriese lig mag meebring dat ter wille van suksesvolle woordeboekgebruik 'n groter lettertipe in dié woordeboek gebruik word, vergeleke met die vergelykbare woordeboek wat saamgestel word vir leerders wat die woordeboek onder optimale gebruiksomstandighede gebruik. Hier lei gebruikersvriendelikheid nogmaals tot 'n aanpassing van die leksikografie by die gemeenskap.

Die tipiese gebruikssituasie wat voorsien is vir die *Ju/'hoan Tsumkwe Dialect/Prentewoordeboek vir kinders/Children's picture dictionary* (Jones et al. 2014) is nie 'n biblioteek, studeerkamer of goed toegeruste klaskamer nie, maar veel eerder 'n omgewing waar die woordeboek deur sy teikengebruikers, jong kinders, in 'n informeler situasie gebruik word. Die woordeboek se bladsye is van 'n dik en sterk glanspapier gemaak wat verhoed dat dit maklik skeur of vuil raak (Jones, persoonlike mededeling). Die leksikografie pas sy produk ook in terme van buiteleksikografiese aspekte aan by die gemeenskap wat die betrokke woordeboek gebruik. Voorskrifte in die leksikografieteorie moet soepel genoeg wees om vir sulke aanpassings voorsiening te maak.

'n Verandering in die kundighedsvlak van potensiële teikengebruikers van woordeboeke is nog 'n buiteleksikografiese aspek wat 'n invloed op die leksikografiese teorievorming en praktyk kan hê en wat in 'n omvattende woordeboekkultuur verreken behoort te word. Naas die verrekening van verskuiwings in gebruikersgroepe lei die bewusmaking in 'n omvattende woordeboekkultuur van meer woordeboeksoorte as net algemene taalwoordeboeke daartoe dat daar ook aandag gegee moet word aan uiteenlopende gebruikers

van dieselfde tipe woordeboek. In die vakleksikografie maak Bergenholtz en Tarp (1995: 19) voorsiening vir verskillende deelnemers aan vakgesprekke, te wete deskundiges, semideskundiges en leke. Hierdie gebruikers moet ook in ag geneem word by die vasstelling van 'n vakwoordeboek se teikenbruiker-groep en die gevolglike datakeuse vir aanbieding in die woordeboek. In aansluiting by Bergenholtz en Tarp se driedeling stel Gouws (2016: 109) 'n verdere gebruikerskategorie voor, naamlik die ingeligte leek. Dit gaan hier om leke ten opsigte van 'n bepaalde vakgebied wat deur gereelde blootstelling aan terme uit daardie vakgebied 'n sekere vertrouwdheid met die vakgebied verwerf. Waar vakterme in algemene aanlyn woordeboeke bewerk word, stel Gouws (2016: 113) voor dat naas die verstekbewerking vir die leek in die semantiese kommentaar van die woordeboekartikel daar 'n bykomende teksblok in die artikelstruktuur opgeneem kan word waar 'n bewerking aangebied word wat toegevoegde waarde vir die ingeligte leek sal hê. Verdere ondersoek kan gedoen word om vas te stel of kundigheidsvlakke in ander opsigte ook tot 'n aanpassing in die verdeling van teikenbruikers lei. In die pedagogiese leksikografie word daar voorsiening gemaak vir verskillende taalaanleervlakke, naamlik 'n beginnersvlak, intermediêre en gevorderde vlak. Vir die aanlyn leksikografie moet 'n omvattende woordeboekkultuur daarvoor voorsiening maak dat gebruikers vertrouwd raak met doelgemaakte woordeboeke waar daar aan een databasis verskillende woordeboeke onttrek kan word — met die gebruiker wat sy/haar profiel moet opstel sodat 'n woordeboek wat aan hulle behoeftes voldoen aan die databasis onttrek kan word. Vergelyk in hierdie verband Bergenholtz en Bergenholtz (2013). 'n Omvattende woordeboekkultuur is ook op gebruikersvriendelikheid gerig wat impliseer dat die leksikografie aanpassings moet maak om produkte te lewer wat ten beste in al hierdie verskillende gebruikersgroepe se leksikografiese behoeftes kan voorsien. In hierdie verband is dit ook noodsaaklik dat gebruikers vertrouwd moet wees en vertrouwd gemaak moet word met die elektroniese toestel waarop die woordeboeke verskyn. Dit is 'n buiteleksikografiese aspek wat aandag verdien in die uitbreiding van 'n omvattende woordeboekkultuur.¹

Nog 'n buiteleksikografiese aspek wat ingrypende leksikografiese implikasies kan hê, is die omstandighede van potensiële woordeboekgebruikers in die digitale era met verwysing na hulle vertrouwdheid met tegnologiese vordering en hulle houding teenoor gedrukte bronne, veral naslaanbronne. 'n Groot persentasie van die lede van die samelewing wat 'n behoefte aan betroubare naslaanbronne het, behoort tot die jonger geslag. Dit is mense wat in die digitale era gebore is en daarin opgroei. Om te probeer verseker dat daar 'n voortgesette behoefte aan die gebruik van woordeboeke sal wees, moet die woordeboekkultuur ook deur die naslaanbehoefte en naslaanvoorkeure van hierdie geslag potensiële gebruikers beïnvloed word. Die digitale era is die era van Generasie Z en die leksikografie moet op hulle behoeftes reageer. 'n Omvattende woordeboekkultuur moet ook buiteleksikografiese aanpassings in die samelewing in ag neem en die leksikografie moet daarby aanpas.

4. Generasie Z

Generasie Z, ook bekend as Gen Z, sluit diegene in wat na 1990 gebore is, dit is die kinders van Generasie X (gebore 1965–1976) en die kinders of kleinkinders van die Baby Boomers (gebore 1946–64) en soms selfs reeds die kinders van Generasie Y (gebore 1977–1998), vergelyk http://de.wikipedia.org/wiki/Generation_Y. Alhoewel baie Gen Z-lede nog op skool is, behoort hulle binne 'n dekade die grootste deel van die werkkrag uit te maak, aldus Asthana (2008). Dit is 'n geslag wat met die internet opgegroeï het en hulle nie 'n lewe daarsonder kan voorstel nie. Volgens die *urbandictionary*:

This generation is very networked with social networks and many means of communication and means of entertainment at their finger tips such as Cell Phones, iPod, Facebook, YouTube and IM. Digital networking and sharing is common and many have not known a time without the internet ... (<http://www.urbandictionary.com>)

Hierdie netbewoners ("net citizens" of "netizens") stel ander verwagtinge aan naslaanbronne. Volgens *Generation Z: Born in the digital age* (2011) glo hulle dat alles moontlik is met behulp van tegnologie en lei hulle 'n lewe:

that seems a million miles removed from the hopes, dreams and morals of previous generations ... They want everything, everywhere and immediately. They surf on two screens simultaneously. They don't mind paying through the nose for the latest smart phone but turn up their nose at paying for a film or a song when you can get that for free online. ... Even as young as they are, they have already seen so many technologies become obsolete. For this reason, they have become the ultimate "self-educators", learning how to use new stuff via self-help videos on YouTube.

Leksikograwe wat leksikografiese produkte vir die onmiddellike toekoms beplan, moet die tipiese kenmerke van Gen Z in gedagte hou want dit gaan 'n invloed op hulle naslaanbehoefte, woordeboekgebruik en woordeboekgebruiksvaardighede hê. Dit vereis ingrypende aanpassings in die leksikografie-teorie en die leksikografiepraktyk en 'n verdere verbreding van 'n omvattende woordeboekkultuur. Gedrukte bronne is nie Gen Z se voorkeurlesstof nie en hierdie gebrek aan belangstelling in boeke, hulle voorkeur vir gratis aanlyn produkte en hulle vermeende vermoë om hulleself met behulp van die internet op te voed, moet leksikograwe daarvan oortuig dat hulle met 'n deel van die samelewing te doen het onder wie tradisionele woordeboeke min of geen invloed het nie, vergelyk Gouws (2017) en Gouws en Tarp (2016).

Parker (2013) beweer dat indien Gen Z digitale integreerders is

then we can't teach them in analogue, where teachers do the teaching, spoon-feeding them knowledge. Those days are gone.

In 'n omvattende woordeboekkultuur moet daar voorsiening gemaak word vir hierdie potensiële teikengebruikers van nuwe leksikografiese produkte. Woor-

deboekdidaktiek moet ook hierby aangepas word. Die gesprek oor leksikografiese hulpmiddels moet nie tot aanlyn woordeboeke beperk word nie, maar ander inligtingsinstrumente wat deur Gen Z gebruik word, moet ook in gedagte gehou word en leksikografiese produkte moet waar moontlik daarby geïntegreer word. Gen Z is ook die selfoongenerasie omdat selfone 'n onafskeidbare en onvervreembare deel van die meeste lede van hierdie generasie is. O'Farrell (2011) wys op die hoë persentasie van die tienerbevolking wat selfone besit. Drie dinge wat selfone aan hierdie gebruikers gee, is toeganklikheid, gewildheid en verbindbaarheid.

Die gemak en onmiddellike bevrediging van kommunikatiewe behoeftes wat selfone meebring, is heel belangrik vir leksikografiese produkte. Daarom moet leksikograwe probeer om leksikografiese produkte te integreer in slimfone om waarde toe te voeg tot hierdie veelsydige instrumente en daardeur 'n beter kans te bewerk om woordeboekgebruik onder Gen Z-lede te vestig. Dit sluit aan by die gedagtes van Finch (2015):

It's critical that we recognize Gen Z's differences and meet them where they are, rather than where we want them to be.

Dit is ook in die leksikografie ter sake en leksikograwe moet woordeboeke dienooreenkomstig beplan maar ook die woordeboekkultuur dienooreenkomstig aanpas en woordeboekdidaktiek, spesifiek die aanleer van woordeboekgebruiksvaardighede, dienooreenkomstig aanpas. Vergelyk in hierdie verband ook Gouws (2017). Nkomo (2015: 74) sê:

No form of culture can thrive without being inculcated or nurtured by means of education, be it formal or informal. The same applies to dictionary culture ...

Hier moet woordeboekdidaktiek 'n aktiewe rol speel om die sukses van 'n omvattende woordeboekkultuur te help verseker.

5. Woordeboekdidaktiek

Dit is 'n gevestigde siening in die leksikografie dat die opleiding van woordeboekgebruikers 'n noodsaaklikheid is. Binne die tradisionele siening van 'n woordeboekkultuur, dit is die aanpassing van die gemeenskap by die leksikografie, kan woordeboekopleiding as 'n ononderhandelbare komponent beskou word. Whitcut (1986) wys reeds op die belang van die opleiding van woordeboekgebruikers. In 'n omvattende woordeboekkultuur moet woordeboekdidaktiek ook as ononderhandelbare komponent gesien word, maar die opleiding moet op sowel woordeboekgebruikers as leksikograwe gerig wees. Lew (2013: 16) wys op twee belangrike aspekte wat nodig is vir suksesvolle woordeboekgebruik. Dit is hoë gehalte, gebruikersvriendelike woordeboeke en woordeboekgebruikers wat weet wat hulle doen. Hierdie twee aspekte weerspieël die kern van die tradisionele tweedeling van gebruikersvriendelikheid en

woordeboekskultuur en dus ook die kern van die tweerigtingverhouding wat 'n omvattende woordeboekskultuur kenmerk. Om leksikograwe daartoe in staat te stel om hoë gehalte gebruikersvriendelike woordeboeke saam te stel en om gebruikers te hê wat weet wat hulle doen, is opleiding, spesifiek woordeboekdidaktiek, noodsaaklik.

Die behoefte aan woordeboekdidaktiek word gereeld in metaleksikografiese publikasies bespreek. Vergelyk in hierdie verband onder meer Béjoint (1989), Lew en Galas (2008), Engelberg en Lemnitzer (2009), Schierholz (2016) en vir aanlyn woordeboeke Klein (2008). Domínguez Vázquez en Sanmarco Bande (2017) word volledig gewy aan die tema leksikografie en didaktiek. Die behoefte aan formele opleiding vir leksikograwe het gelei tot die instelling van graadkursusse in teoretiese leksikografie, onder meer in Suid-Afrika aan die Universiteit van Stellenbosch en die toonaangewende MA-program EMLex, dit is die interuniversitêre, internasionale European Master in Lexicography, vergelyk <http://www.emlex.phil.uni-erlangen.de/>.

Volgens Klein (2007: 423) is dit vir 'n suksesvolle woordeboekskultuur nodig dat daar voortdurend samewerking tussen leksikograwe en onderwysers moet wees. Die onderwysers is, aldus Klein, 'n belangrike faktor in die ontwikkeling van 'n woordeboekskultuur. Vir die digitale era moet woordeboekdidaktiek gerig wees op vaardighede in die gebruik van aanlyn woordeboeke maar daar moet ook die nodige fokus op die tipiese gebruikers van hierdie aanlyn woordeboeke wees. Die behoeftes en vaardighede van lede van Gen Z stel nuwe uitdagings aan die leksikografieteorie. In 'n samelewing met 'n aktiewe omvattende woordeboekskultuur wat gekenmerk word deur die aanpassing van die gemeenskap by die leksikografie en van die leksikografie by die gemeenskap moet die digitale era 'n invloed hê op die aanpassings wat deur albei groepe rolspelers in die woordeboekskultuur gemaak word — ook ten opsigte van woordeboekdidaktiek. Leksikograwe moet kreatief dink oor vernuwendere maniere waarop woordeboekgebruik en woordeboekgebruiksvaardighede onderrig moet word. Vergelyk in hierdie verband onder meer Dringó-Horváth (2014) en Tiberius en Müller-Spitzer (2015). In hierdie beplanning moet woordeboeke nie in isolasie gesien word nie maar as deel van 'n veel wyer reeks digitale naslaanbronne. Woordeboekvaardighede moet ook vaardighede meebring om ander aanlyn naslaanbronne te raadpleeg. Mettertyd kan 'n omvattende woordeboekskultuur uitgebrei word tot 'n omvattende naslaan-kultuur.

7. Ten slotte

Die sukses van woordeboeke as praktiese gebruiksinstrumente is daarvan afhanklik dat die kundige teikengebruikers weet wanneer om watter woordeboek te kies, weet watter data in watter woordeboek aangebied behoort te word en daartoe in staat is om toegang te kry tot die data waaraan die tersaaklike inligting onttrek moet word. Wanneer 'n woordeboekraadpleging afgesluit

word met die verkryging van inligting wat 'n oplossing bied vir die probleem wat die woordeboekraadpleging gemotiveer het, is daar sprake van sukses. Suksesvolle woordeboekgebruik en die maak van woordeboeke wat aan die geïdentifiseerde teikenbruikersgroep se leksikografiese behoeftes voldoen, word bevoordeel deur die bestaan van 'n omvattende woordeboekkultuur met 'n tweerigtingwerking tussen die leksikografie en die samelewing. Met die oog op die digitale era waar lede van Gen Z oorreed moet word om woordeboeke as naslaanbronne te gebruik en waar aanlyn woordeboeke ook op die behoeftes van Gen Z gerig is, is die vestiging van 'n omvattende woordeboekkultuur maar ook voortgesette metaleksikografiese navorsing om die gehalte van die woordeboekkultuur te verbeter en uit te brei dringend noodsaaklik.

8. Nota

1. My dank aan 'n anonieme keurder wat my aandag op hierdie aspek gevestig het.

9. Erkenning

This research is supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID 85434). The Grant-holder 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.

10. Bibliografie

10.1 Woordeboeke

Hartmann, Reinhard R.K. en Gregory James. 1998. *Dictionary of Lexicography*. Londen/New York: Routledge.

Johnson, S. 1755. *A Dictionary of the English Language*. Londen: J. & P. Knapton/T. & T. Longman et al.

Jones, Kerry L. et al. (Reds.). 2014. *Ju/hoan Tsumkwe Dialect/Prentewoordeboek vir kinders/Children's Picture Dictionary*. Pietermaritzburg: University of KwaZulu-Natal Press.

urbandictionary = <http://www.urbandictionary.com>

10.2 Ander bronne

Asthana, Anushka. 2008. *They Don't Live for Work ... They Work to Live*. <http://www.theguardian.com/money/2008/may/25/workandcareers.worklifebalance>.

Atkins, Beryl T.S. (Red.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. Tübingen: Max Niemeyer.

- Béjoint, Henry.** 1989. The Teaching of Dictionary Use: Present State and Future Tasks. Hausmann, Franz-J. et al. (Reds.). 1989–1991: 208-215.
- Bergenholtz, Henning en Sven Tarp (Reds.).** 1995. *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, Inger en Henning Bergenholtz.** 2013. One Database, Four Monofunctional Dictionaries. *Hermes* 50: 119-125.
- Caruso, Valeria.** 2015. Explorations on Data Condensation and Data Distribution in the Microstructures of e-Dictionaries. *Lexicographica* 31: 38-66.
- Domínguez Vázquez, Maria. J. en M.T. Sanmarco Bande (Reds.).** 2017. *Lexicografía y Didáctica*. Frankfurt: Peter Lang. (Om te verskyn.)
- Dringó-Horváth, Ida.** 2014. *Wörterbuchdidaktik für digitale Wörterbücher*. http://www.daf-online.hu/docs/DringóHorvath_2014_Woerterbuchdidaktik_218-228.pdf.
- Du Plessis, André.** 2015. 'n Funksionele analise van die e-WAT: 'n Grondslag vir die verbetering van internet- en selfoonwoordeboeke. Ongepubliseerde M.A.-tesis. Universiteit van Stellenbosch.
- Engelberg, Stefan en Lothar Lemnitzer.** 2009: *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.
- Finch, Jeremy.** 2015. *What Is Generation Z, And What Does It Want?* <http://www.fastcoexist.com/3045317/what-is-generation-z-and-what-does-it-want>.
- Generation Z: Born in the Digital Age.* 2011. <http://english.astroawani.com/lifestyle/generation-z-born-digital-age-53754>.
- Giacomini, Laura.** 2015. Macrostructural Properties and Access Structures of LSP e-Dictionaries for Translation: The Technical Domain. *Lexicographica* 31: 90-117.
- Gouws, Rufus H.** 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. *Lexikos* 13: 92-110
- Gouws, Rufus H.** 2005. Meilensteine auf dem historischen Weg der Metalexikographie. *Lexicographica* 21: 158-178.
- Gouws, Rufus H.** 2012: Towards a System of Integrated Dictionary Use. Karpova, Olga en Faina Kartashkova (Reds.). 2012. *Multi-disciplinary Lexicography: Traditions and Challenges of the XXI-st Century*: 134-144. Cambridge: Cambridge Scholars Publishing.
- Gouws, Rufus H.** 2013. Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, Vida (Red.). 2013. *Specialised Lexicography*: 51-62. Berlyn/New York: De Gruyter.
- Gouws, Rufus H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Gouws, Rufus H.** 2014a. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3): 481-504.
- Gouws, Rufus H.** 2015. Expanding the Notion of Addressing Relations. *Lexicography*: 1(2): 159-184.
- Gouws, Rufus H.** 2016. Increasing the Scope of the Treatment of Specialised Language Terms in General Dictionaries. Schierholz, Stefan et al. (Reds.). 2016: 101-118.
- Gouws, Rufus H.** 2016a. Dictionary Criticism and Dictionary Critics. Bielińska, Monika and Stefan J. Schierholz (Reds.). 2016. *Wörterbuchkritik*. Berlyn: De Gruyter. (Om te verskyn.)
- Gouws, Rufus H.** 2017. La sociedad digital y los diccionarios. Domínguez Vázquez, Maria. J. en Sanmarco M.T. Bande (Reds.) 2017. *Lexicografía y Didáctica*. Frankfurt: Peter Lang. (Om te verskyn.)
- Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard en Herbert Ernst Wiegand (Reds.).** 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Devel-*

- opments with Focus on Electronic and Computational Lexicography*. Berlyn/New York: Walter de Gruyter.
- Gouws, Rufus H. en Danie J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, Rufus H., Wolfgang Schweickard en Herbert Ernst Wiegand.** 2013. Lexicography through the Ages: From the Early Beginnings to the Electronic Age. Gouws, Rufus H. et al. (Reds.). 2013: 1-24.
- Gouws, Rufus H. en Sven Tarp.** 2016. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 29. (Om te verskyn.)
- Hartmann, Reinhard R.K.** 1987. Four Perspectives on Dictionary Use: A Critical Review of Research Methods. Cowie, Anthony P. (Red.). 1987. *The Dictionary and the Language Learner*: 11-28. Tübingen: Niemeyer.
- Hartmann, Reinhard R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, Franz J. et al (Reds.). 1989: 102-111.
- Hausmann, Franz J.** 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, Franz J. et al. (Reds.). 1989-1991: 1-19.
- Hausmann, Franz J. et al. (Reds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Berlyn: De Gruyter.
- Hausmann, Franz J. en Herbert Ernst Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, Franz J. et al. (Reds.). 1989-1991: 328-360.
- Johnson, S.** 1747. *The Plan of a Dictionary of the English Language*. Faksimilee-uitgawe 1970. Menston: The Scholar Press.
- Klein, Juliane.** 2007. Implementing a Dictionary Culture in South Africa: An Attempt at a General Framework for the African Languages. *Lexikos* 17: 416-423.
- Klein, Juliane.** 2008. Teaching Dictionary-using Skills for Online Dictionaries: An Attempt at a Theoretical Framework for South Africa. Bernal, Elisenda en Janet DeCesaris (Reds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 1265-1271. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Klosa, Annette.** 2009. Außentexte in elektronischen Wörterbüchern. Beijk, Egbert et al. (Reds.). 2009. *Fons Verborum: Feestbundel Fons Moerdijk*: 49-60. Amsterdam: Gopher BV.
- Klosa, Annette en Rufus H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.
- Klotz, Michael en Thomas Herbst.** 2016. *English Dictionaries. A Linguistic Introduction*. Berlyn: Erich Schmidt Verlag.
- Lew, Robert.** 2013. Online Dictionary Skills. Kosem, Iztok et al. (Reds.). 2013. *Electronic Lexicography in the 21st Century: Thinking outside the Paper*: 16-31. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.
- Lew, Robert.** Om te verskyn. Space Restrictions in Paper and Electronic Dictionaries and their Implications for the Design of Production Dictionaries. Bański, Piotr en Beata Wójtowicz (Reds.). Om te verskyn. *Issues in Modern Lexicography*. München: Lincom Europa.
- Lew, Robert en Katarzyna Galas.** 2008. Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-School-Level Polish Learners of English. Bernal, Elisenda en Janet DeCesaris (Reds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona,*

- 15–19 July 2008: 1273-1285. Barcelona: Universitat Pompeu Fabra.
- McArthur, Tom.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablets to the Computer.* Cambridge: Cambridge University Press.
- Müller-Spitzer, Carolin.** 2013. Textual Structures in Electronic Dictionaries. Gouws, Rufus H. et al. (Reds.). 2013: 367-381.
- Müller-Spitzer, Carolin en Sascha Wolfer.** 2015. Vernetzungsstrukturen digitaler Wörterbücher. Neue Ansätze zur Analyse. *Lexicographica* 31: 173-199.
- Nkomo, Dion.** 2015. Developing a Dictionary Culture through Integrated Dictionary Pedagogy in the Outer Texts of South African School Dictionaries: The Case of *Oxford Bilingual School Dictionary: IsiXhosa and English.* *Lexicography* 2(1): 71-99.
- O'Farrell, Nicholas.** 2011. <http://www.ning.com/blog/2011/07/generation-mobile1.html>.
- Parker, Phil.** 2013. <http://www.sec-ed.co.uk/blog/how-generation-z-is-different>.
- Schierholz, Stefan.** 2016. Einige kurze Überlegungen zur Metalexikographie in der Lehre. Schierholz, Stefan, et al. (Reds.). 2016: 295-308.
- Schierholz, Stefan, et al. (Reds.).** 2016. *Wörterbuchforschung und Lexikographie.* Berlin: De Gruyter.
- Svensén, Bo.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making.* Cambridge: Cambridge University Press.
- 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.
- Tiberius, Carole en Carolin Müller-Spitzer.** 2015. Research into Dictionary Use. Wörterbuchbenutzungsforschung. 5. *Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“.* Mannheim: Institut für Deutsche Sprache.
- Tono, Yukio.** 2001. *Research on Dictionary Use in the Context of Foreign Language Learning. Focus on Reading Comprehension.* Tübingen: Max Niemeyer.
- Van der Merwe, Michele.** 2015. *Perceptions of the MAAL Generation in Higher Education on the Use of LSP Dictionaries on Mobile Phones.* Paper presented at the StelleLex-Colloquium, Stellenbosch, March 2015.
- Welker, Andreas.** 2013. Empirical Research into Dictionary Use since 1990. Gouws, Rufus H. et al. (Reds.). 2013: 531-540.
- Welker, Andreas.** 2013a. Methods in the Research of Dictionary Use. Gouws, Rufus H. et al. (Reds.). 2013: 540-547.
- Whitcut, Janet.** 1986. The Training of Dictionary Users. Ilson, Robert (Red.). 1986. *Lexicography: An Emerging International Profession:* 111-122. Manchester: Manchester University Press.
- Wiegand, Herbert Ernst.** 1977. Nachdenken über Wörterbücher: Aktuelle Probleme. Drosdowski, Günther, Helmut Henne en Herbert E. Wiegand (Reds.). 1977. *Nachdenken über Wörterbücher:* 51-102. Mannheim/Wenen/Zürich: Bibliographisches Institut.
- Wiegand, Herbert Ernst.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, Reinhard R.K. (Red.). 1984. *LEX'eter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983:* 13-30. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst.** 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, Franz J. et al. (Reds.). 1989–1991: 371-409.
- Wiegand, Herbert Ernst.** 1989a. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, Franz J. et al. (Reds.). 1989–1991: 409-462.

- Wiegand, Herbert Ernst.** 1989b. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, Franz J. et al. (Reds.). 1989–1991: 462-501.
- Wiegand, Herbert Ernst.** 1989c. Der gegenwärtige Status der Lexikographie. Hausmann, Franz J. et al. (Reds.). 1989–1991: 246-280.
- Wiegand, Herbert Ernst.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, Herbert Ernst (Red.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger lexikographischen Kolloquium*: 1-82. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Mit 159 Illustrationen im Text.* Berlyn/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wiegand, Herbert Ernst.** 2008. Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. *Lexicographica* 24: 209-315.
- Wiegand, Herbert Ernst (Red.).** 2015. *Lexicographica* 31. Berlyn: De Gruyter.
- Wiegand, Herbert Ernst en Sandra Beer.** 2013. Access Structures in Printed Dictionaries. Gouws, Rufus H. et al. (Reds.). 2013:110-149.
- Wiegand, Herbert Ernst, Sandra Beer en Rufus H. Gouws.** 2013. Textual Structures in Printed Dictionaries: An Overview. Gouws, Rufus H. et al. (Reds.). 2013: 31-73.
- Wiegand, Herbert Ernst, Ilse Feinauer en Rufus H. Gouws.** 2013. Types of Dictionary Articles in Printed Dictionaries. Gouws, Rufus H. et al. (Reds.). 2013:314-366.
- Wiegand, Herbert Ernst en Rufus H. Gouws.** 2013. Macrostructures in Printed Dictionaries. Gouws, Rufus H. et al. (Reds.). 2013: 73-110.
- Wiegand, Herbert Ernst en Maria Smit.** 2013. Microstructures in Printed Dictionaries. Gouws, Rufus H. et al. (Reds.). 2013: 149-214.

10.3 Webblaaie

http://de.wikipedia.org/wiki/Generation_Y.

<http://www.emlex.phil.uni-erlangen.de/>

The Application of the Prototype Theory in Lexicographic Practice: A Proposal of a Model for Lexicographic Treatment of Polysemy*

Ana Halas, *Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia (anahalas@gmail.com)*

Abstract: The aim of this paper is to point to fundamental implications of the prototype theory for lexicographic practice using them as a solid foundation for developing and proposing a model for systematic, consistent, precise and, above all, reliable treatment of polysemy in dictionaries. The main requirement of such a model is to make transparent the main features of a polysemous structure including its hierarchical organization, mutual relatedness of senses based on family resemblance and motivated derivation of senses. The proposed model includes three most challenging but key aspects of lexicographic treatment of polysemy: sense discrimination, sense defining and structure and organization of a dictionary entry. The prototype-based view of a polysemous structure is illustrated with a diagram representing a proposed upgrade of the radial set model introduced by Brugman and Lakoff (1988) and providing a pattern for tailoring a dictionary entry for a polysemous word. After practical aspects of the proposed model are explained in detail, an illustration follows setting forth a sample entry for a highly polysemous verb formed according to the proposed model. A solid theoretical foundation of the model is expected to ensure its practical efficiency primarily concerning quick and easy detection and understanding of senses.

Keywords: LEXICOLOGY, LEXICOGRAPHY, COGNITIVE LINGUISTICS, PROTOTYPE THEORY, POLYSEMY, DICTIONARY ENTRY, MODEL, SENSE DISCRIMINATION, SENSE DEFINING, ENTRY STRUCTURE

Opsomming: Die toepassing van die prototipeteorie in die leksikografiese praktyk: 'n Voorstel vir 'n model vir die leksikografiese hantering van polisemie. Die doel van hierdie artikel is die aantoon van basiese implikasies van die prototipeteorie vir die leksikografiese praktyk en die gebruik daarvan as stewige grondslag vir die ontwikkeling en voorstelling van 'n model vir sistematiese, konsekwente, noukeurige en veral betroubare hantering van polisemie in woordeboeke. Die belangrikste vereiste vir so 'n model is die verheldering van die hoofkenmerke van 'n polisemiese struktuur met inbegrip van hiërargiese samestelling, onderlinge verwantskap van betekenis geïmpliceer deur gemeenskaplike ooreenkomste en gemoti-

* This paper is based on the author's doctoral dissertation written under the supervision of Professor Tvrtko Prčić and defended at the University of Novi Sad.

veerde afleiding van betekenis. Die voorgestelde model sluit drie uiters uitdagende maar sleutel-aspekte van die leksikografiese hantering van polisemie in: betenisonderskeiding, betekenisdefiniëring en die struktuur en samestelling van 'n woordeboekinskrywing. Die prototipe-gebaseerde beskouing van 'n polisemiese struktuur word geïllustreer deur 'n diagram. Hierdie diagram is 'n voorgestelde verbetering van die radiale reeks-model wat deur Brugman en Lakoff (1988) bekendgestel is en wat 'n ontwerp verskaf vir die skep van 'n woordeboekinskrywing vir 'n polisemiese woord. Nadat die praktiese aspekte van die voorgestelde model in besonderhede uitgelê is, volg daar 'n illustrasie van 'n voorbeeldinskrywing vir 'n hoogs polisemiese werkwoord wat geskep is volgens die voorgestelde model. Daar word verwag dat 'n stewige teoretiese grondslag vir die model die praktiese doeltreffendheid daarvan, veral betreffende die vinnige en eenvoudige vasstelling en begrip van die betekenis, sal verseker.

Sleutelwoorde: LEKSIKOLOGIE, LEKSIKOGRAFIE, KOGNITIEWE LINGUISTIEK, PROTOTIPETEORIE, POLISEMIE, WOORDEBOEKINSKRYWING, MODEL, BETEKENISONDER-
SKEIDING, BETEKENISDEFINIËRING, INSKRYWINGSTRUKTUUR

1. Introduction

The advent of cognitive linguistics in the 1980s marked a turning point in the development of semantic theories and their view of meaning bringing the notion of polysemy into focus again. One of the main advances in the perception and understanding of polysemy within the framework of cognitive linguistics was the incorporation of the prototype theory into its account of word meaning. The prototype theory represented a strong and direct opposition to the classical view of words and their meaning relying on psychology and its empirical findings, viewing meaning as a form of categorization and relating it to mental representations. Such an account of meaning and polysemy is considered here as the one that can be successfully applied to lexicographic treatment of polysemy. However, the analysis of polysemy treatment in British and American lexicography (represented by six monolingual general-purpose dictionaries and five learner's dictionaries of English) that I have conducted as a part of my doctoral dissertation (Halas 2014a) has shown that most of these dictionaries still adhere to the principles of traditional lexicography and do not base their practice on contemporary lexicological theory. *The New Oxford Dictionary of English* (NODE, Pearsall 1998) is the first dictionary to be compiled taking into account achievements of cognitive semantics so that its main characteristic is the differentiation between a lexeme's main uses labelled as core senses and their subsenses, i.e. more specific uses (Halas 2014b: 160) as well as the organization of senses within entries that shows that they "hang together" (Van der Meer 2000: 421) and that they are not mutually independent. However, after analysis of this dictionary's practice, Van der Meer (2000: 428) concludes that it is not completely coherent and consistent in applying the principles of the given theoretical basis.

The main premise in this paper is that complete fulfilment of users' needs

results from a successful interplay of the contemporary lexicological theory and lexicographic practice. Furthermore, the prototype theory has been recognized in this paper as highly beneficial to lexicographic treatment of polysemy including all its most challenging aspects - sense discrimination, sense defining and structure of the entry. Therefore, this paper presents an attempt to clearly point to the ways in which the application of the prototype theory can increase efficiency and user-friendliness of lexicographic practice concerning the treatment of polysemy as well as reliability of data presented in a dictionary entry. Thus, there will be formulated a proposal of a model for lexicographic treatment of polysemy completely based on the principles of the prototype theory.

2. The prototype-based view of polysemy

The notion of polysemy is typically associated with a multiplicity of senses of a single lexeme. However, the key feature of polysemy that differentiates it from other cases of lexical ambiguity, such as homonymy, is defined by Cruse (2004: 108) as a motivated relationship between senses. This relationship is motivated since each sense is derived from another one in the same polysemous structure through a mechanism of sense extension, such as: specialization, generalization, metaphor, metonymy, etc. This mutual relatedness of senses within a polysemous structure became especially prominent when the cognitive linguistic view of polysemy arose. Thus, when defining polysemy, Evans (2005: 33) also emphasizes that polysemy is a phenomenon "whereby a single linguistic form is paired with a number of distinct but related meanings or senses".

Cognitive linguistics has studied polysemy as a form of categorization (Lewandowska-Tomaszczyk 2007: 140). Wittgenstein (1980) defined categorization as speakers' ability to perceive entities as members of particular groups or categories, while each category has its own typical features. Wittgenstein also concludes that categories have blurred boundaries while their members are related on the basis of family resemblance since their common features "overlap and criss-cross" just like traits in one family (Ravin and Leacock 2006: 13).

On the basis of the results of her psychological experiments, Rosch (1977) reaches the conclusion that people categorize entities on the basis of their resemblance to the best exemplar of the given category that exhibits the most salient features of the category. The central member of a category is the prototype, the best representative of the given category. The other members of the category exhibit different degrees of their similarity with the prototype. For example, if the category of birds is taken into consideration, a sparrow could be regarded as the prototypical bird, i.e. the prototype of the category of birds. A pigeon, swallow or canary exhibit a greater degree of similarity with the prototype due to which, in a visual representation of this category, they should be positioned closer to the center, while a penguin or a platypus share a smaller number of features with the prototype so that they are regarded as peripheral members (Prčić 2016: 59). Therefore, some members are better exemplars of the

category, while others are less representative, which leads to the conclusion that the category membership is graded. Hence, the internal structure of a category develops from its core, i.e. the prototype and extends gradually over the central members as more typical ones to peripheral, less typical members. Actually, the main criterion for determining the position of an individual member in a category can be regarded as goodness-of-exemplar (Croft and Cruse 2004: 77).

The key aspects of the prototype-based view of categories can be summarized in four prototypicality effects or features formulated by Geeraerts (1989):

- (1) Prototypical categories exhibit degrees of typicality.
- (2) Prototypical categories are blurred at the edges.
- (3) Prototypical categories cannot be defined by a single set of (necessary and sufficient) features.
- (4) Prototypical categories exhibit a family resemblance structure.

Since polysemy is viewed as a form of categorization within the cognitive linguistic theoretical framework, the four prototypicality features apply to a polysemous structure as well. Interpreted in the specific context of polysemy, the first feature refers to the difference among senses in their structural weight due to the fact that there is a prototypical sense lying at the core of the structure while the other senses can be more or less central or peripheral depending on the number of features, or in this particular case – sense components, they share with the prototype. The second prototype effect indicates the flexibility of prototypical categories, which means that, as Taylor (1989: 53) claims, they can include new members without any restructuring of the whole category. In the case of a polysemous structure, these new members refer to newly derived senses since a polysemous structure can be enriched and expanded through meaning extension. A polysemous structure is also defined in terms of a cluster of partial descriptions since none of the members of a particular category exhibits all of its typical features, i.e. none of the senses in a polysemous structure possesses all the prototypical components. Hence, there is not a single set of features typical of a particular category common as a whole to all of its members. Thus, Lewandowska-Tomaszczyk (2007: 146) explains that there is not a set of necessary and sufficient features by which a category can be defined but rather categories are defined in terms of a cluster of partial descriptions. Still, all the senses within a single polysemous structure are organized around and derived directly or indirectly from the same prototype so that they are all mutually related sharing some common features and overlapping, which indicates that mutual relatedness of senses within a polysemous structure is based on family resemblance.

Lakoff (1987) sets forth his view that polysemous words should be understood as radial categories. Brugman and Lakoff (1988) present the radial set model for the description of a polysemous structure. According to this model, the central member, i.e. the prototype is a cognitive model that motivates all

the other senses in the given structure and they are all positioned relative to the center. The visual representation of Lakoff's radial set model is given below:

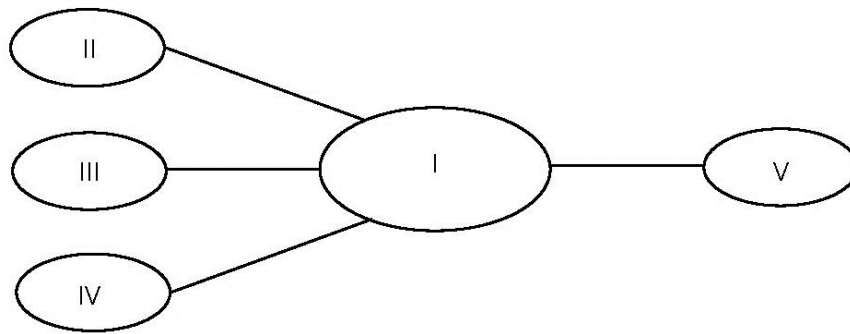


Figure 1: A radial set (Lewandowska-Tomaszczyk 2007: 156)

Evans (2005: 34) follows Lakoff's model and, thus, arrives at a similar description of a polysemous structure claiming that all senses in one such structure form a motivated semantic network organized around a central sense that this author refers to as the sanctioning one. Senses in a polysemous structure are linked to the prototype through various mechanisms (e.g. metaphor, metonymy, etc.) by means of which they are derived from it. As Evans and Green (2006: 332) explain, the operation of these mechanisms results in the formation of meaning chains creating a polysemous structure of a lexeme. As the main advantage of the radial set model, Evans (2005: 40) emphasizes the fact that it clearly shows degrees of relatedness between senses since some senses are more closely related to the prototype while others can be more closely related to some derived sense in the structure. The afore-mentioned author sees this as clustering of senses according to patterns that reveal possible paths of derivation. Actually, senses in a structure are mutually connected by relational links representing pathways along which one sense has been derived from another. As Geeraerts (2001: 11) notices, the radial set model emphasizes the multidimensional nature of a polysemous structure arising from mutual derivational links among senses.

2.1 The upgrade of the radial set model by Brugman and Lakoff

The prototype-based model offers a deeper insight into the internal organization of a polysemous structure focusing upon its center as the core from which the whole structure has developed and, thus, emphasizing mutual relatedness of all the senses in a structure based on family resemblance, which is successfully visually represented by the radial set model introduced by Brugman and Lakoff (1988). However, the fact that should not be neglected is that the proto-

type-based model of polysemy also emphasizes that there are differences among senses in their structural weight and centrality. Basically, the organization of a polysemous structure follows the cline from center to periphery (Geeraerts 2001: 11). Moreover, noticed differences in structural weight and centrality among senses strongly suggest that a polysemous structure is hierarchically organized, which provides the basis for proposing an upgrade of the radial set model that will present the internal structure and organization of a polysemous structure in even greater detail.

The visual representation of this model consists of a set of concentric circles where each circle represents a level of semantic derivation. The basic or prototypical sense of the given polysemous structure is positioned in the center of the diagram, which is in accordance with its primary and central status in the structure, as can be seen in the illustration of the model shown below.

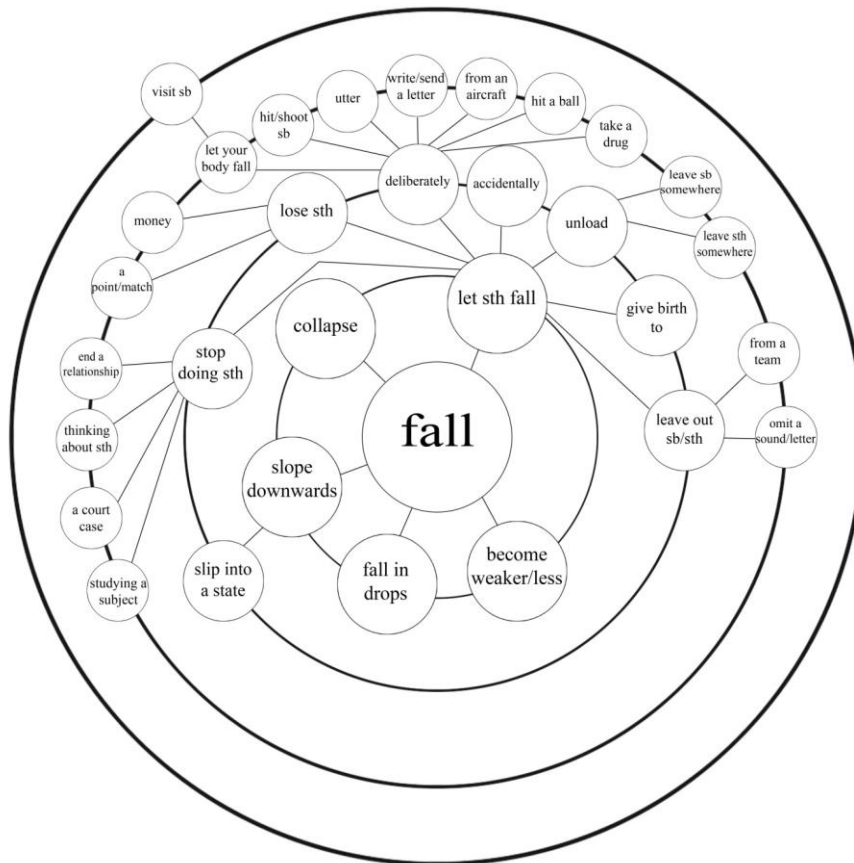


Figure 2: The polysemous structure of the verb *drop* according to the proposed upgrade of the radial set model

This is a diagram reflecting the highly polysemous structure of the verb *drop*, including a considerable number of senses so that as such it is an illustrative example of the way in which this model actually functions. The prototypical sense of this verb is 'fall'. The question that arises here is how the prototypical sense is established. Evans (2005: 44) proposes four criteria for determining the prototypical sense of a polysemous structure: 1) historically earliest attested meaning, 2) predominance in the semantic network in the sense of type-frequency, 3) predictability regarding other senses, 4) a sense which relates to lived human experience, i.e. experience at the phenomenological level. Since the model proposed in this paper is based on the view that a polysemous structure is characterized by family resemblance while representing a motivated network of senses, the criterion applied for the identification of the prototypical sense presents the interplay of the second and third listed criteria. The second criterion suggests that the prototype should be determined on the basis of the meaning component that is the most frequent in the given structure. The third criterion is in accordance with Evans' claim that sense derivation is principled and motivated so that the prototype is the sense from which the other senses "most naturally" have been derived (Evans 2005: 44). Hence, the prototype is established as the sense component that is directly or indirectly incorporated into the other senses in the given structure representing the core from which, most plausibly, they have been derived directly or indirectly.

The center of the diagram is surrounded by a certain number of evenly spaced concentric circles. In the case of the polysemous structure of the verb *drop*, there are four concentric circles since the given structure encompasses four levels of semantic derivation. Thus the number of concentric circles surrounding the center of the diagram is actually dictated by the complexity of the polysemous structure in question, i.e. by the number of its derivational levels. The first concentric cycle or level of semantic derivation (the one closest to the center) contains senses directly derived from the prototypical one. They will be referred to as primary senses. In this particular case, there are five primary senses derived from the prototypical sense through various mechanisms: metaphor ('become weaker/less'), metonymy ('let something fall'), and specification ('fall in drops', 'slope downwards', 'collapse'). These primary senses can have their own potential for further semantic dispersion and act as the basis for the derivation of other, new, more peripheral senses less closely related to the prototype. These senses indirectly derived from the prototype will be referred to as secondary senses and they are positioned on the second derivational level (i.e. the second concentric circle). Two of the five identified primary senses of the verb *drop* are semantically productive in the sense that they have a semantic potential for derivation of new senses. One of them is the sense 'let something fall' that is the base from which several secondary senses have been derived through metaphor ('stop doing something', 'lose something', 'unload', 'give birth to', 'leave somebody out') and specialization ('let something fall deliberately', 'let something fall accidentally'). The mentioned primary sense acts, thus,

as the superordinate to its subsenses, i.e. the mentioned more peripheral senses directly derived from it and only indirectly from the prototype. Hence, the relation of subordination is established between most closely related senses, i.e. the superordinate sense and its subsenses. This fact proves the existence of a hierarchy within a polysemous structure. Moreover, derivational paths for all senses descending from the first to the second derivational level, i.e. from superordinate senses to their subsenses become transparent in this model. However, the sense derivation process is a recurrent one, which means that every secondary sense from the second derivational level can also be the base for further sense derivation and become the superordinate sense to the senses derived from it directly, i.e. to its subsenses positioned on the next, third level being even more peripheral so that the same relation of subordination and hierarchy is established between the senses on the second level and the corresponding derived senses on the third level. An illustrative example is the sense 'let something fall deliberately' from which a number of more peripheral senses have been derived as it can be seen on the diagram. The senses on the third level can also become derivational bases motivating further semantic dispersion. Hence, this recurrent nature of the process of semantic derivation ensures further extension of a polysemous structure.

3. The implications of the upgraded prototype-based model for the lexicographic treatment of polysemy

As Atkins and Rundell (2008: 130) claim, a grounding in relevant theoretical ideas or those that have direct application to lexicographers' work is an invaluable tool when tackling the task of analyzing data and tailoring precise and neat dictionary entries. It is considered here that only lexicographic practice with a solid foundation in the contemporary lexicological theory can provide a valid and credible description of a semantic structure within a dictionary entry. As it has been concluded that the prototype theory provides an insight into the internal organization of a polysemous structure, it can be used as a firm and, above all, useful theoretical basis for the lexicographic treatment of polysemy. The proposed upgraded model of representing a polysemous structure makes this insight even deeper and more detailed on the basis of which it is possible to analyze the semantic structure of the given entry. When the given semantic structure is analyzed and represented according to this model, lexicographers have a clear and detailed depiction of this structure that serves as a pattern for tailoring a dictionary entry for the given lexeme taking into account all the key aspects of polysemy treatment: sense discrimination, sense definition and structure and organization of senses within an entry. In this way, this challenging and particularly important aspect of lexicographic practice ceases to be arbitrary, random and subjective and becomes systematic, theoretically well-founded but also practically efficient, which will be proved in the following sections of the paper.

The fundamental implications of the proposed upgraded prototype-based model of representing a polysemous structure for the lexicographic treatment of polysemy are set forth below:

- (1) A polysemous structure is organized around the prototype as the motivational core from which all the other senses in a structure are derived directly or indirectly. Therefore, when establishing the semantic structure of a polysemous lexeme, lexicographers' initial step is to identify the prototype as the basic sense in the given structure.
- (2) Derivation of senses within a polysemous structure is motivated. As it has already been explained, a new sense is derived from another one in the structure through the activation of a sense derivation mechanism. This clearly suggests, as Csábi (2002: 250) points out, that lexicographers should do the motivational analysis of senses, which includes identifying a derivational path for each sense as well as recognizing the activated mechanism. In this way, there are established groups of closely related senses comprised of a superordinate sense and its subsenses.
- (3) Mutual relatedness of senses based on family resemblance naturally stems from previously explained motivated sense derivation. Motivated derivational links existing between a superordinate sense and its subsenses as well as the fact that all senses in a structure originate directly or indirectly from the same prototypical sense strongly imply that one sense cannot be defined as an independent, individual use of the given lexeme but as relative to its closely related senses. Accordingly, in order to provide a complete and reliable description of a sense, definitions should reflect these family resemblances as well as the hierarchy of a polysemous structure.
- (4) Due to the organization according to the principle 'from center to periphery' and the relation of subordination between a superordinate sense and its subsenses, it follows that only a hierarchically structured dictionary entry does full justice to the multilayer internal organization of a polysemous structure.

These implications are underpinning theoretical principles of the proposed model for the lexicographic treatment of polysemy. Practical aspects of this model, an illustration of its application and its advantages will be presented in the following sections of the paper.

4. Practical aspects of the proposed model for lexicographic treatment of polysemy

Practical aspects of the proposed lexicographic model will be explained through reference to three key aspects of polysemy treatment: sense discrimination, sense defining and structure and organization of an entry.

4.1 Sense discrimination

Sense discrimination is in this paper understood as a lexicographic procedure including sense identification and sense differentiation. Sense identification subsumes two tasks: identification of the basic/prototypical sense and identification of other senses in a structure, both primary and secondary ones. After a thorough search of a corpus and gathering all possible uses that form the semantic network of the given lexeme, a lexicographer should identify its semantic core, i.e. the semantic base from which all the other members of the given structure have been derived directly or indirectly (Halas 2014b: 161-162). The analysis of the semantic network of the verb *drop*¹ clearly shows that the most plausible and natural semantic core of the polysemous network of the verb *drop* is the sense 'fall vertically'.

Identification of other senses in a structure, both primary and secondary ones, consistently follows detected sense derivation paths descending from a superordinate sense to its subsenses. For each specific use of a particular lexeme, a lexicographer should identify the sense which it has been derived from, i.e. its superordinate as well as a mechanism of its derivation. Subsequently, the semantic potential of the given use is to be analysed with the purpose of identifying all its subsenses. It follows from this that it is necessary to establish each derivational path in its entirety descending from the first to the last derivational level in a structure. Thus, the hierarchical relation among senses in a structure is another principle that governs the sense identification process according to this model. An example of a derivational path within the polysemous structure of the verb *drop* is the following: the primary sense 'let something fall' is derived directly from the prototypical sense 'fall vertically' through the mechanism of metonymy (according to the pattern MOVE ON ITS OWN (OF AN OBJECT)-MAKE IT MOVE) but it also acts as a superordinate sense from which several secondary senses have been derived, such as the sense 'stop doing something' (through metaphor according to the pattern DISCONTINUING SOMETHING IS LETTING IT FALL) positioned accordingly on the second derivational level. However, this derivational path reaches up to the next, third derivational level since there are several specific uses of the verb *drop* that are identified as subsenses of the mentioned secondary sense, such as: 'end a relationship', 'stop studying a subject' (derived by means of specialization realized by specifying a referent, e.g. a relationship, a school subject), etc. The analysis of their derivational origin leads to the sense 'stop doing something' as their derivational base. Generally, the initial step in the sense identification process is to recognize the specific nature of a particular use in the corpus, which is followed by the reconstruction of its derivational path through gaining an insight into its relation to the other senses in the structure.

When a specific use of the given lexeme is identified and established, the questions that arise are firstly, whether it is possible and secondly, whether it is needed to decompose the given use into several more specific, more narrowly

defined senses. This is the task of sense differentiation that is to be in accordance with the hierarchical organization of a polysemous structure so that its procedure is based on a proper combination of lumping and splitting of senses. This means that superordinate senses should be identified as more general uses while their subsenses should be finely differentiated one from another so that subtle distinctions among them are visible and prominent enough especially if their specific features are semantically productive and, thus, significant for further sense derivation. The described procedure is exemplified by the use 'let something fall' that is established as a quite broad superordinate sense whose wide range can encompass various more specific uses as its subsenses. On the second derivational level, there are two recognized variations of this general sense involving different ways of letting something fall denoted by its subsenses 'let something fall deliberately' and 'let something fall accidentally'. It is completely justified to split the mentioned more general superordinate sense into these two subsenses and establish them as individual senses since the distinction between them reflected in the opposition of sense components 'deliberately-accidentally' is of essential importance for further sense derivation. Namely, several new senses are identified as derived from the sense 'let something fall deliberately' which is obviously highly semantically productive as it can be seen in the figure 2. Therefore, differentiae among the subsenses on the third derivational level should be elicited in order to make these specific senses prominent enough.

However, as Stock (1984: 131) underlines, it is of crucial importance to establish the adequate extent of splitting more general senses into very subtle nuances of meaning which become individuated in this way although there are rather slight differences among them. Therefore, there is a risk of making such differentiation redundant and unnecessarily detailed. This can lead to a too great number of subsenses, which requires too much effort from a lexicographer to organize them into a neat and clear list of senses as well as for a user to easily and quickly spot the desired sense. For example, the subsense 'to let a part of your body fall suddenly' could be further split into at least three subsenses by eliciting differences among them that refer to the exact part of the body which is dropped (e.g. mouth, jaw, eyes). However, this difference is not of relevance for further semantic derivation and individuation of each of these variations would only unnecessarily increase the number of individual senses and make the organization of the entry too complex and detailed, thus, harming its transparency. It is not advisable either to entirely neglect these variations of the given more general sense or miss to make them visible. An optimal solution in such cases could be to specify these various referents in sentences or phrases serving as illustrative examples following and supporting the definition of the mentioned sense. The same practice can be applied to the primary sense 'become weaker, less, lower' that can be split into several individual subsenses on the basis of various referents it is used with, such as voice, wind and speed. However, a more practically efficient solution would be to lump these

possible variations into a single individual sense and specify them in illustrative examples so that this certainly would not harm the preciseness of representing the sense inventory of the verb *drop*. The recommended practice will be illustrated in the proposed dictionary entry for the verb *drop* that will be presented in the following chapter of the paper.

In conclusion, it can be stated that, according to the proposed theoretical model, sense differentiation is to follow the systematic combination and balancing of sense lumping and splitting following the rule which states that superordinate senses tend to be generalized, while their subsenses should be established through splitting and subtle sense differentiation especially if the established differentia is significant for further semantic dispersion and if it contributes to the necessary preciseness of representing the sense inventory without harming its clarity and neatness.

4.2 Sense defining

According to the proposed model, the crucial requirement that sense definitions are expected to fulfil is to reflect motivated nature of sense derivation. This basically indicates that mutual relatedness of senses within a polysemous structure based on family resemblance is to be made transparent with a proper formulation of sense definitions. In order to make this relatedness among senses easily noticeable, it is necessary to provide a formal indication of this relation in the very sense definitions of every superordinate sense and its subsenses. That formal indication is the identical formulation of the common semantic base or the common dominant feature shared between a superordinate sense and a sense directly derived from it, i.e. the feature that motivated the given sense derivation. This also implies that the formulation of each sense definition should be such that every derived sense can be easily related to its superordinate sense and that its derivational path including the employed derivational mechanism is transparent. Thus, the formulation of the common semantic core shared by the prototypical sense of the verb *drop* and all the primary senses directly derived from it is 'to fall vertically' and the identical formulation of this core is contained in the definitions of all the primary senses. In this way, it becomes obvious that they are related to the prototypical sense. Consequently, it also becomes obvious that these senses are all mutually related due to their common derivational origin and family resemblance. The same principle applies to defining of secondary senses.

Thus, according to the proposed model, the formulation of sense definitions adopts the following procedure: the basis of a sense definition is established by excerpting the formulation of the common dominant component or semantic core from the definition of its superordinate sense (i.e. from the definition of the sense it has been derived from) so that the identical formulation of the shared basic component is present in the definitions of the two most closely related senses. Then, the explanation of the component(s) specific for the sense

being defined is added to the formulation of the shared basic component. The explained procedure is illustrated by definitions of senses belonging to the same derivational path descending from the prototype to the second derivational level. The definition of the primary sense 'to make something fall vertically' contains the component on the basis of which it has been derived from the prototype which is 'to fall vertically'. The text of the given definition further incorporates the component specific for this primary sense that reflects the metonymic pattern motivating this derivation, (MOVE ON ITS OWN (OF AN OBJECT)-MAKE IT MOVE), so that the final formulation of this definition as a whole is: 'to make something fall vertically'. This primary sense has a considerable semantic potential since a number of new senses have been derived from it as it has already been noticed. One of its subsenses or senses from the second derivational level is 'to deliberately make something fall vertically'. Its definition has obviously been formulated in the following way: the dominant component it shares with the superordinate sense has been incorporated into the definition using the same formulation. Then, its specific semantic component, 'deliberately', has been added. In this way, the relation between this subsense and its superordinate sense has been made explicit and transparent while its narrower scope and more specific nature have been pointed out as well. Moreover, the hierarchical relation between the superordinate sense and its subsense has been reflected by their definitions since the superordinate sense is defined more broadly so that its scope encompasses all its more specific subsenses.

One can put the question about the application of this model to the case of defining metaphorical senses whose link with their superordinate is not obvious to the same extent as in the previous case since it is reflected in the metaphoric pattern according to which the given metaphorical sense has been derived. The model of defining metaphorical senses will be exemplified by the sense 'to stop doing something'. It has been derived from the sense 'to make something fall vertically' through the metaphorical pattern STOPPING DOING SOMETHING IS MAKING AN OBJECT FALL TO THE GROUND. Atkins and Rundell (2008: 287) claim that figurative extension of meaning (or the derivation of metaphorical senses) involves an implied resemblance between the literal and newly derived non-literal sense. In this particular example, the derivation of the new sense is based on the similarity with the component of the literal sense referring to the act of making something reach an endpoint, which, in this case, is the point at which an object reaches the ground. In the given metaphorical sense, the endpoint is the one at which one stops performing some activity. This metaphorical pattern is to be made explicit in the sense definition so that the derivational link between the metaphorical sense and its superordinate sense becomes transparent. This can be achieved by incorporating the metaphorical pattern into the formulation of the definition in the following way: 'to stop doing something (which resembles making an object fall to the ground)'.

Svensén (2009: 224-225) states that it is not always possible to determine a set of necessary and sufficient distinctive features to be included into a sense

definition since borderlines of categories are fuzzy so that a lexicographer might choose to define such a phenomenon by describing the typical case often introducing it by words such as *typically, usually, especially, often* or by listing typical exponents of a category preceded by expressions like *for example, e.g., such as* etc. Jackson (2002: 95) refers to such definitions as 'typifying' ones. Similar definitional practice is applied in the model proposed in this paper and it can be illustrated by the following definitions: 'to stop having a relationship with someone, especially a romantic one' or '(of an animal) to give birth to young, especially a foal (which resembles making an object fall to the ground)'. In these prototype-based definitions, the typical case is clearly pointed out by the expression *especially*. Jehle (2004: 58) underlines that pointing to typical subjects and circumstances of a verbal activity in sense definitions helps users to understand but also create idiomatic sentences in the given language by thinking of referents that are matched against typical ones on the basis of family resemblance.

Therefore, the system of sense definitions within one entry established in the previously explained manner contributes largely to the reliable representation of a polysemous structure as a unified whole based on family resemblance.

4.3 Structure and organization of entries

According to the proposed model, the entry is structured hierarchically since it is claimed here that this type of structure can most reliably depict the multi-level nature of a polysemous structure. An entry can have a two, three or even four-layer hierarchical structure. Each entry is divided into as many numbered paragraphs as there are identified primary senses in the given polysemous structure. However, there is one paragraph at the very head of the list of senses clearly separated from the others and underlined. This is the paragraph for the prototypical sense of the given polysemous structure marked by number one so that its central position and status in a polysemous structure is clearly indicated. Every next paragraph in the list of senses encompasses a group of most closely related senses, i.e. the superordinate sense and all its subsenses or one entire derivational path descending from the first to the last derivational level in the structure. The most prominent element in one such paragraph is the primary sense acting as the superordinate one and it is always numbered. Its direct subsenses are listed immediately after it and are marked by letters following the model a,b,c, etc. Each of these subsenses is presented within a separate subparagraph slightly indented in relation to the paragraph of its superordinate sense forming, thus, the second hierarchical layer in the entry. If some of the subsenses from the second derivational level act as the bases for further semantic derivation, it is necessary to form the third layer in the structure. Actually, a subsense marked by a letter that has such a semantic potential becomes secondarily superordinate to its direct subsenses each of which is positioned within its own sub-subparagraph slightly indented in relation to the

paragraph of its superordinate sense and it is marked according to the following model: a.1,a.2,a.3; b.1,b.2,b.3 etc. New or deeper layers in the hierarchical structure of the entry are formed using the same model and are marked in the following way: a1.1, a1.2, a1.3, etc.

Such a detailed and precise hierarchical structure of an entry effectively contributes to the transparency of mutual relatedness among senses and derivational paths of individual senses.

The question that has remained unanswered so far refers to sense ordering within a dictionary entry. The senses belonging to the same hierarchical layer or derivational level are, as the figure 2 shows, equally distant from the prototype so that their order cannot be established on the basis of the declining scale of their closeness to the prototype or 'from center to periphery' (Halas 2013: 276). Hence, the only logical criterion for ordering of senses on the same hierarchical layer is the frequency of their occurrence in the contemporary language use. Moreover, an insight into the frequency of various senses of a word is of great value to dictionary users. The frequency of a particular sense can be determined by analyzing a corpus of the contemporary language use. Senses on the same hierarchical layer are ordered according to the descending scale of their frequency in the contemporary language use. This principle is applied to every hierarchical layer in a structure. However, it should be noticed that only the prototypical sense is always listed first in an entry regardless of its frequency in the corpus in order to give it its due prominence.

5. Illustration of the proposed model

The application of the proposed model is illustrated by tailoring a dictionary entry for the verb *drop*:

drop verb /BrE drɒp, AmE dra:p/ **drops, dropping, dropped**

1. FALL² [I] to fall vertically:
[drop+adverb/preposition: from, onto, into...] *The cup suddenly dropped from her hand.³ I noticed the moment when her earring dropped onto the green floor.*

2. MAKE STH FALL [T] to make something fall vertically:
She dropped her bag as soon as she entered the room.
 - a. DELIBERATELY [T] to deliberately make something fall vertically:
[drop something+adverb/preposition: from, onto, into...] *I dropped the noodles into the water and gave them a stir. He and his son were dropping stones from the bridge into the river.*

- a.1.** BODY [T/I] to deliberately make a part of your body fall suddenly:
*She immediately dropped her head against the pillow. When she saw me, her **mouth dropped/jaw dropped**⁴ (= she was very surprised). FORMAL When she asked him to tell her the truth, he just **dropped his eyes** (= looked downwards).*
- a.2.** FROM AN AIRCRAFT [T] to deliberately make supplies, equipment or troops fall from an aircraft by parachute:
The ally dropped food supplies in the refugee camp.
- a.3.** WRITE A LETTER [T] to write or send a letter (which resembles deliberately making letters fall to the paper):
*Just **drop me a line/note** from time to time.*
- a.4.** UTTER [T] to say something informally or indirectly (which resembles deliberately making words fall from the mouth):
*Mary **dropped a hint** that she would move to England the following year.*
- a.5.** HIT A BALL [T] SPORT to make a ball fall into a hole, basket or goal:
You can drop a putt into the hole easily if the speed is good.
- a.6.** TAKE A DRUG [T] INFORMAL to take a drug by making it fall down your throat:
*They knew that she **dropped acid** a couple of times.*
- b.** ACCIDENTALLY [T] to accidentally make something fall vertically:
She dropped her favourite plate while doing the washing-up. I dropped the needle on the carpet and now I can't find it.
- c.** STOP DOING STH [T] to stop doing something (which resembles making an object fall to the ground):
Can you drop everything for a moment and give me a hand?
- c.1.** TALKING [T] to stop talking about something or discussing it:
*Can you just **drop it**? I don't want to listen about it any more. After an hour of discussion, they finally **dropped the subject**.*
- c.2.** RELATIONSHIP [T] INFORMAL to stop having a relationship with someone, especially a romantic one:
[drop somebody] After ten years, she dropped Tom.

- c.3.** SCHOOL SUBJECT [T] to stop studying a school subject:
In her third year, Mary dropped German and chose another language.
- c.4.** COURT CASE [T] to stop dealing with a court case:
*The prosecution decided to **drop the case**.*
- d.** LEAVE SB/STH OUT [T] to leave somebody or something out accidentally or deliberately (which resembles making an object fall to the ground):
[drop somebody/something from something] *The editor asked them to drop that word from the title. The topic considered controversial was dropped from the list of the offered ones.*
- d.1.** FROM A TEAM [T] SPORT to leave somebody out from a team or side:
[drop somebody from something] *Thomson had a bad season so the coach dropped him from the team.*
- d.2.** SOUND/LETTER [T] to leave out a sound or a letter while pronouncing or writing a word:
He always drops his aitches when speaking in front of the audience. You dropped one 'r' every time you wrote 'tomorrow'.
- e.** LOSE STH [T] to fail to win something (which resembles making an object fall to the ground):
- e.1.** POINT/MATCH [T] SPORT to lose a point, match or competition:
The Blues drop points as Manchester City go top of the table. He won the tournament without dropping a single match.
- e.2.** MONEY [T] INFORMAL to lose or spend a particular amount of money, especially in gambling:
I dropped a thousand dollars in the casino last night. She dropped 50 euros for a portion of lobster.
- f.** UNLOAD [T] to leave passengers or goods at a particular place (which resembles making them fall to the place):
- f.1.** LEAVE SB SOMEWHERE [T] to take somebody by car and leave them at a particular place, especially on the way to another place:
Just drop me near the post office and I'll go by bus from there. I'm driving to work so I can drop you in the town centre.
- f.2.** LEAVE STH SOMEWHERE [T] to take something and leave it at a particular place without staying there very long:

[drop something round/off] *I found your wallet so I'll drop it round to your place when I finish my work. Just drop it off on your way home.*

- g. GIVE BIRTH TO [T] (of an animal) to give birth to young, especially a foal (which resembles making an object fall to the ground):

We are expecting a mare to drop a foal in a few days.

3. BECOME WEAKER/LESS/LOWER [I,T] to become or make something reduced to a lower level, rate, number, intensity etc. (which resembles falling or making something fall vertically):

*He suddenly **dropped** his **voice** when he realized that the baby was sleeping. **[drop off]** The production **dropped off** during last year. The temperature **dropped sharply/dramatically/suddenly** over last few days. The **wind** finally **dropped** on Sunday so we could continue our journey. **[drop to]** The price of shares **dropped** to 1,000 \$. Flats **dropped in price** by 15%. You need to **drop** your **speed** when it is raining.*

4. FALL IN DROPS [I/T] to fall or make something fall vertically in drops of liquid:

*The rain was **dropping** on the roof. Can you **drop** it into my eye?*

5. COLLAPSE [I] to fall vertically to the ground unexpectedly or suddenly, as from a shot, wound, exhaustion, death, etc:

[drop down, onto, into...] INFORMAL *I was working so hard that at the end of the day I was **ready to drop**. We heard a shot and saw him **dropping** to his knees in the mud. When she came back, she just **dropped** into her favourite armchair feeling completely exhausted.*

6. SLOPE DOWNWARDS [I] (of a ground) to slope steeply (as if falling vertically):

[drop away, down] *It was a wonderful view – the land **dropped away** to the river. The path **dropped down** sharply.*

- a. SLIP INTO A STATE [I] to enter or pass into a particular state or condition without conscious effort (resembling a ground sloping steeply):

[drop into] *He can ordinarily **drop into** sleep at times when he should be occupied.*

6. Conclusion

The concluding remarks of this paper will concisely point out the benefits of

the proposed model for both lexicographers and dictionary users. Namely,

- (a) The proposed model of lexicographic treatment of polysemy has a solid theoretical foundation built on the principles of a contemporary lexicological theory (the prototype theory).
- (b) The proposed model reliably reflects the nature of a polysemous structure including its hierarchical organization, mutual relatedness of senses based on family resemblance and motivated derivation of senses, making all these prominent features of a polysemous structure fully transparent.
- (c) Consequently, this model of polysemy treatment can be justifiably expected to achieve practical efficiency that refers to proper and complete fulfilment of users' needs including, primarily, quick detection of a particular sense and its memorization and comprehension with less effort. Also, this model can satisfactorily meet research and scholarly needs of dictionary users.
- (d) In addition, the visual representation of an entry is consistent, systematic, clear and precise, which significantly contributes to the previously mentioned practical efficiency of the model.
- (e) In conclusion, the proposed model is characterized by solid scientific foundation, consistency, systematicity and uniformity, which is expected to lead to considerable reduction of arbitrariness in the treatment of polysemy in dictionaries.

The idea underpinning the proposed model is that the prototype theory offers a firm basis for defining a set of strategies that will successfully tackle some of the most challenging lexicographic issues, such as the treatment of polysemy. Actually, the proposed model represents an illustration of the premise that such a firm theoretical basis ensures enhanced efficiency of lexicographic practice.

Endnotes

1. For the purpose of illustrating the proposed model of polysemy treatment, the sense network of the verb *drop* has been created on the basis of a sense inventory compiled after the excerption of all various senses offered within entries for this verb in various monolingual dictionaries of English, including general-purpose and learner's dictionaries.
2. The proposed model involves the use of sense indicators in front of definitions, which is considered as good practice already adopted by certain learner's dictionaries of English (*Oxford Advanced Learner's Dictionary*, *Longman Dictionary of Contemporary English*, *Cambridge Advanced Learner's Dictionary*) and for example, *Encarta World English Dictionary* as a general-purpose dictionary of English. A sense indicator is a word or a short phrase expressing the most prominent or specific component of the given sense. Such practice enables quick and easy search of an entry.

3. For this particular purpose of tailoring a single example entry to illustrate the proposed model, example sentences have been formulated by the author of the paper herself as her attempt to provide illustrations of the typical context in which each sense is used.
4. Lexical collocations provided in illustrative examples are marked by red colour.

References

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Brugman, C. and G. Lakoff. 1988. Cognitive Topology and Lexical Networks. Small, S., G. Cottrell and M. Tannenhaus (Eds.). 1988. *Lexical Ambiguity Resolution: 477-507*. San Mateo: Morgan Kaufman.
- Croft, W. and D.A. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Cruse, A. 2004. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Second Edition. Oxford: Oxford University Press.
- Csábi, S. 2002. Polysemous Words, Idioms and Conceptual Metaphors. Cognitive Linguistics and Lexicography. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, 13-17 August 2002: 249-254*. Copenhagen: Center for Sprogteknologi, University of Copenhagen.
- Evans, V. 2005. The Meaning of Time: Polysemy, the Lexicon and Conceptual Structure. *Journal of Linguistics* 41(01): 33-75.
- Evans, V. and M. Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Geeraerts, D. 1989. Prospects and Problems of Prototype Theory. *Linguistics* 27: 587-612.
- Geeraerts, D. 2001. The Definitional Practice of Dictionaries and the Cognitive Semantic Conception of Polysemy. *Lexicographica* 17: 6-21.
- Halas, A. 2013. The Representation of a Multidimensional Semantic Structure in English Monolingual Learner's Dictionaries. Gudurić, S. and M. Stefanović (Eds.). 2013. *Languages and Cultures in Time and Space II: 269-280*. Novi Sad: Faculty of Philosophy.
- Halas, A. 2014a. *Polysemy in Dictionaries of English and Serbian: Theoretical, Methodological and Practical Aspects of its Treatment*. An unpublished doctoral dissertation. Novi Sad: University of Novi Sad.
- Halas, A. 2014b. The Treatment of Polysemy in the *Oxford Dictionary of English*. Prčić, T. et al. (Eds.). 2014. *Festschrift in Honour of Draginja Pervaz English Language and Anglophone Literatures in Theory and Practice: 155-172*. Novi Sad: Faculty of Philosophy.
- Jackson, H. 2002. *Lexicography. An Introduction*. London/New York: Routledge.
- Jehle, G. 2004. Prototype Semantics and Learners' Dictionaries of English. Herbst, T. et al. (Eds.). 2004. *Lexikografie, ihre Basis- und Nachbarwissenschaften: 53-69*. Tübingen: Max Niemeyer.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lewandowska-Tomaszczyk, B. 2007. Polysemy, Prototypes, and Radial Categories. Geeraerts, D. and H. Cuyckens (Eds.). 2007. *The Oxford Handbook of Cognitive Linguistics: 139-169*. Oxford: Oxford University Press.

- Pearsall, J. (Ed.).** 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Prčić, T.** 2016. *Semantika i pragmatika reči*. Third electronic edition. Novi Sad: Faculty of Philosophy.
- Ravin, Y. and C. Leacock (Eds.).** 2006. *Polysemy. Theoretical and Computational Approaches*. Oxford: Oxford University Press.
- Rosch, E.** 1977. Human Categorization. Warren, N. (Ed.). 1977. *Advances in Cross-Cultural Psychology*. Volume 1: 1-72. London: Academic Press.
- Stock, P.F.** 1984. Polysemy. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography, Exeter, 9-12 September 1983*. Lexicographica. Series Maior 1: 131-141. Tübingen: Max Niemeyer.
- Svensén, B.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Taylor, J.R.** 1989. *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Clarendon Press.
- Van der Meer, G.** 2000. Core, Subsense and the *New Oxford Dictionary of English* (NODE). On How Meanings Hang Together, and not Separately. Heid, U. et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, 8-12 August 2000*: 419-431. Stuttgart: Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.
- Wittgenstein, L.** 1980. *Philosophische Untersuchungen*. Frankfurt am Mein: Suhrkamp.

Lemmatisation of Fixed Expressions: The Case of Proverbs in Northern Sotho*

I.M. Kosch, *Department of African Languages,
University of South Africa, Pretoria, South Africa*
(koschim@unisa.ac.za)

Abstract: The purpose of this article is to make a quantitative and qualitative assessment of the lexicographic treatment and listing of proverbs in the *Wörterbuch der Sotho-Sprache* (Endemann 1911) in comparison to selected Northern Sotho dictionaries. In order to accommodate proverbs, which are fixed multiword expressions, they are customarily entered as sub-lemmas under a particular simple headword, usually one of the key components of a proverb. The selection of a key component relies on the subjective judgement of the lexicographer. This selective approach may result in proverbs falling between the cracks if none of the components strike the compiler as prominent enough to justify the inclusion of a proverb under a particular headword. This seems to have been the case in the dictionary under investigation, given the dearth of proverbs taken up in this work. On the other hand their omission could simply be ascribed to a practical consideration such as limited space in a printed dictionary. A dictionary user might find it challenging to look up a desired proverb, especially if the individual words have a very low general frequency or are even obsolete in modern life. In that case, an electronic format of a dictionary would be most enabling, allowing for an electronic search. Special purpose dictionaries dedicated to culturally-birthed sayings such as proverbs, will go a far way in safeguarding their knowledge for posterity.

Keywords: MULTIWORD EXPRESSIONS, PROVERBS, LEXICOGRAPHIC TREATMENT, KEY COMPONENT, HEADWORD, GENERAL DICTIONARY, SPECIAL PURPOSE DICTIONARY

Opsomming: Lemmatisering van vaste uitdrukkings: Die geval van spreekwoorde in Noord-Sotho. Die doel van hierdie artikel is om 'n kwantitatiewe en kwalitatiewe bepaling te maak van die leksikografiese hantering en lemmatisering van spreekwoorde in die *Wörterbuch der Sotho-Sprache* (Endemann 1911) vergeleke met geselekteerde Noord-Sotho woordeboeke. Ten einde spreekwoorde, wat vaste meervoudige woorduitdrukkings is, te huis-

* Paper presented at ALASA Conference (African Language Association of Southern Africa) at Walter Sisulu University in Umtata, South Africa, 30 June 2014.

ves, word hulle gewoonlik as sub-lemmas onder 'n bepaalde enkelvoudige hoofwoord geakkommodeer, gewoonlik een van die sleutelkomponente van die spreekwoord. Die keuse van 'n sleutelkomponent hang af van die subjektiewe oordeel van die leksikograaf. Hierdie selektiewe benadering kan tot gevolg hê dat spreekwoorde uit die oog verdwyn as geen van hul komponente prominent genoeg deur die samesteller geag word om die insluiting van die spreekwoord onder 'n bepaalde hoofwoord te regverdig nie. Dit blyk die geval te wees in die woordeboek wat hier onder bespreking is, gegewe die skaarsste van spreekwoorde wat daarin opgeneem is. Aan die ander kant kan die weglating van spreekwoorde ook eenvoudig aan 'n praktiese oorweging toegeskryf word, soos beperkte spasie in 'n gedrukte woordeboek. 'n Woordeboekgebruiker mag dit uitdagend vind om 'n verlangde spreekwoord op te soek, veral as die individuele woorde 'n baie lae frekwensie het of selfs in onbruik geraak het in die moderne lewe. In dié geval, sal 'n woordeboek in elektroniese formaat die gebruiker bemagtig deur die gebruik van 'n elektroniese opsoekmetode. Woordeboeke vir spesiale doeleindes wat hul toespits op kultuur-gebaseerde segswyses soos spreekwoorde, speel 'n groot rol om dié kennis vir die nageslag te bewaar.

Sleutelwoorde: MEERVOUDIGE WOORDUITDRUKKINGS, SPREEKWOORDE, LEKSIKOGRAFIESE BESKRYWING, SLEUTELKOMPONENT, HOOFWOORD, ALGEMENE WOORDEBOEK, SPESIALE DOEL WOORDEBOEK

1. Introduction

A dictionary, if it is to serve as an "authoritative container of knowledge" (Gouws and Prinsloo 2005: 1), needs to be a reflection of the culture of a particular linguistic community. Hence it "compels lexicographers to contextualise the language in terms of the more general world of the relevant speech community" (Gouws and Prinsloo 2005: 2 referring to Zgusta 1971). Proverbs are bearers of culture and hence they also deserve to be lemmatised. Lemmatisation refers to how and where a lexicographer chooses to list a word or phrase in a dictionary. Lexicographers are faced with a challenge when it comes to the lemmatisation of proverbs — a fact which is evident from the haphazard inclusion or exclusion of proverbs in different dictionaries for a language like Northern Sotho, for example. Proverbs are multiword expressions, but the design of the traditional dictionary is not conducive to their lemmatisation. On the contrary, single-word items are the elements which by default constitute the macrostructure of a dictionary. This is corroborated by Gouws and Prinsloo (2005: 86) who state that "dictionaries have often been characterised and dominated by a word-bias". Thus any other items smaller or larger than a word (sub-lexical or multilexical respectively), would not constitute independent look-up items, but would only be found as sub-entries under main entries as illustrated in Figure 1.

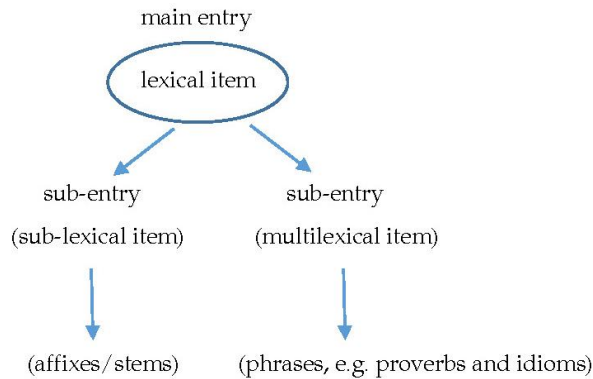


Figure 1: Lemmatisation of sub-lexical and multilexical items

Gouws and Prinsloo (2005: 86) argue that in a lexical-based approach to the macrostructure of a dictionary *all* types of lexical items should have their own lemmatic status, be they sub-lexical items such as affixes and stems or multilexical items such as proverbs and idioms. They convey meaning like any other lexical unit and therefore deserve comprehensive treatment in a dictionary, just like traditional lexical items. Mulhall (2010: 1355) adds that words and phrases share an equal status as units of meaning in the lexicon. Despite this, lexicographers usually make a distinction in how they record words as opposed to phrases: Words are usually listed as main-entries while phrases are listed as sub-entries of main-entries and do not enjoy their own lemmatic status.

In this article the lemmatisation of proverbs in a number of Northern Sotho dictionaries is compared and particular reference is made to the *Wörterbuch der Sotho-Sprache* (Dictionary of the Sotho languages) by Endemann, published in 1911, to highlight some of the complexities of lemmatising proverbs.

2. Identification of proverbs in Endemann's dictionary

Endemann's dictionary is a bilingual, unidirectional Sotho-German dictionary. The term 'Sotho' in the title encompasses all three of the Sotho languages, namely Tswana, Southern Sotho and Northern Sotho. The proverbs were easily identified by the label *Spruchwort* (German for 'proverb'), which appears in the internal article structure of an entry, invariably given either in full or as an abbreviation (*Sprichw.*), with or without parenthesis, before or after a proverb (see labelling in bold in examples in (1) to (3)). For the purpose of this discussion the current orthography is used instead of Endemann's. In the examples below, Endemann's entries are followed first by the researcher's own literal English translation and then by an interpretation extracted from another source which lists the relevant proverb.

(1) Endemann (1911: 242)

kgori, Pl. di-, "wilder Pfau". Kgori e bona mae, moutlo ga e o bone. "der wilde Pfau sieht die Eier, die Falle sieht er nicht" (**Spruchwort**).

Lit: 'The kori bustard sees the eggs, it doesn't see the trap.'

Interpretation: 'Have your mind on what you want to do without thinking of the consequences; if one has set one's heart on something one is inclined to see its bright side only' (Ziervogel and Mokgokong 1975: 681).

(2) Endemann (1911: 692)

boganolemare "Schmeichelei". Vergl. unser "Speichelleckerei". **Spruchwort**: Gano le mare ga di lemogwe "Schmeichelei wird nicht gemerkt".

Lit: 'The mouth cavity and saliva are not recognised (they don't understand each other).'

Interpretation: *Man kann es einem Lügner und Betrüger nicht ansehen, ob er die Wahrheit spricht oder lügt* (Kuhn 1929/1930: 48). 'One cannot perceive whether a liar and deceiver is speaking the truth or is lying' (own translation).

(3) Endemann (1911: 527)

tadi, Pl. di-, "gestreifte Feldmaus". Ngwana-tadi o tsebja ka mereto "das Kind vom *tadi* wird an den Streifen erkannt" (**Spruchw**).

Lit: 'The child of the field mouse is recognised by its stripes.'

Interpretation: 'like father like son; a chip of the old block; a tree is known by its fruit' (Ziervogel and Mokgokong 1975: 890).

In total 24 labelled proverbs were found in the 690 pages of the dictionary. The search was done by manually paging through the dictionary and checking each entry. This figure might be slightly higher, as at least one proverb was encountered which had not been marked with the label 'proverb'. Other unmarked proverbs might have escaped detection.

3. Internal rigidity of proverbs

Proverbs, like idioms, are multiword expressions, but they are more rigid in structure than idioms. They have their own fixed microtext, but small changes may nevertheless occur, compare (3) above with the same proverb in Ziervogel and Mokgokong (1975: 890) in (4) below:

(4) Ngwana wa tadi o tsebja ka mebala/merêtô

In (4) the full possessive form (*ngwana wa tadi*) is used instead of the compound

word *ngwana-tadi*. An alternative lexical item *mebala* (colours) is given in addition to *merêtô* (stripes).

Manyawu (2012: 214) describes both proverbs and idioms as "invariable metaphoric utterances that are almost as stable as standard lexical items". As part of the lexicon of the language, even these slightly unstable forms should be incorporated in a dictionary. According to Čermák (2014) proverbs can come in two forms, namely unmodified (reproduction) or based on a partial change (modification). The partial changes may be syntagmatic, paradigmatic or mixed. In his study of English proverbs, Čermák (2014: 25) found that the longer a stable and familiar proverb is, the greater the possibility of variability. It is their very familiarity which gives speakers or writers the creative freedom to quote only the first half of a proverb, if they so wish, because they know their listeners or readers will easily be able to infer the second half from the context, as in *lerumo le tee ...* 'one assegai...' which the addressee will be able to complete with *...ga le bolaye kgomo* '...will not kill a head of cattle', meaning that a person on his/her own cannot succeed in a task. An investigation into the types of variations of proverbs falls outside the scope of this paper, but could form the topic of another study, especially if the proverbs can be culled from a large corpus of written texts and oral passages, based on real usage.

As is well known, the meanings of proverbs can seldom be guessed from the meanings of the individual words which make up the proverbs. As stated by Akande and Mosobalaje (2014: 35) "Proverbs are, often, short value-laden expressions with multi-layered meanings that can be decoded only by those who possess a good mastery of the oral art and culture that produce them". They are figurative and didactic expressions and embrace the inherited wisdom and experience of a people. They are reflective of a nation's values, norms and morality and are used to bring across messages to guide, edify and admonish.

4. The problem of proverb selection and headword selection

When a dictionary design includes the lemmatisation of fixed combinations of words such as proverbs, the lexicographer will need to ask him-/herself the following questions:

- Which type of source is most likely to contain proverbs for inclusion in the dictionary?
- How many and which proverbs should be included?, and thirdly
- How and where is the best place to enter the proverbs in the dictionary?

In a study conducted on the British National Corpus, Čermák (2014: 28) advanced the following regarding the texts most likely to contain proverbs:

It is now generally recognized that proverbs are not limited to a single text domain only, although there might be a tendency for them to occur more often in some. What one may only presume so far, is a different and *perhaps higher use and distribution in the spoken language*. [own emphasis].

The suggestion that proverbs are more prevalent in spoken than in written language makes sense, as proverbs have a communicative function and entail the impartation of wisdom, caution or admonition, normally by one person to another or to an audience in a verbal exchange. Should the lexicographer therefore base his data on an entirely written corpus, he is not likely to encounter many proverbs. His corpus should include examples from written as well as spoken language to increase his chances of coming across proverbs, more particularly well-established proverbs. The written texts should cover a wide spectrum of types, such as periodicals, newspapers, novels (fiction), books (non-fiction) and other miscellaneous works (Čermák 2014: 28). Čermák (2014: 48) concludes that the "nonexistence of proverbs, for example, in chemistry, mathematics or physics, may suggest that proverbs appear in the traditional fields linked to practical human life and its recurrent aspects and repeated patterns shared by the whole community over a period of time. Thus, it seems that proverbs in their coverage and use belong to social sciences (including linguistics), rather than to exact ones ...". In Northern Sotho written texts, proverbs are most likely to occur in works of a didactic nature. They also seem to have a heightened occurrence as titles of books, films or theatres, as opening and concluding lines. Sometimes they may also be used as the heading of a chapter or section, for example *bana ba tau ga re jane, re molokomong* 'children of the lion we do not eat each other, we are family' (*Tšhupa-Mabaka a Kereke* 1959: 36, a Church Bulletin) to aptly convey the gist of a message.

Regarding the second question on the *number and type* of proverbs, the lexicographer has to keep the needs of the target user in mind and make a selection of the most suitable proverbs from the overwhelming number of proverbs that exist in Northern Sotho. It goes without saying that a comprehensive listing of all existing proverbs is not possible within the constraints of a paper-based dictionary and that any selection will of necessity be incomplete and subjective, because criteria such as popularity and frequency of use are relative. No frequency studies have been carried out on proverbs for Northern Sotho to determine the most common ones expressing the most common meanings. One gets the impression that the inclusion of proverbs is not necessarily a deliberate exercise in a general purpose dictionary, but that proverbs are included randomly as example sentences where they can serve as a suitable context in which to illustrate the use of a lemma. The *Oxford Bilingual School Dictionary. Northern Sotho and English* (De Schryver 2007) is an example of a dictionary in which the proverbs which occur as example sentences were not purposefully chosen by the lexicographer, but emanated from the corpus on which the dictionary was based. Users may not be aware that they are being presented with proverbs as example sentences in an article treatment in de Schryver (2007), since the fixed expressions are not marked by a specific label. Had these proverbs been marked by a special label, this would have served a useful additional educational purpose. For example, an unmarked proverb *Se bone thola boreledi, teng ga yona go a baba* is used as an example sentence in de Schryver (2007: 28) under the head word *boreledi* (smoothness) with only its literal

English translation 'don't look at the smoothness of the bitter apple, the inside is bitter'. The figurative meaning (i.e. not everything is as it seems) is not provided. A comparable example in English would be: 'not all that glitters is gold'.

The third question is probably the most challenging for a lexicographer, namely where to *enter* the proverbs in a dictionary. It requires the lexicographer to identify the most suitable allocation lemma from the elements which constitute the multiword expression.

In the conventional word-based approach, according to Mulhall (2007) "the decision to list an idiom, or any phrase, as a sub-entry necessitates the lexicographer to choose an element of the phrase, which they [sic] believe to be the most suitable point of entry as well as being the most identifiable to the dictionary user listing". In this process one lexicographer might choose one lexical item as key component, while another may find another element more suitable as key component of the same proverb. This is responsible for discrepancies in the lemmatisation of proverbs across dictionaries. In Endemann (1911: 527), for example, the proverb in example (3) is listed under *tadi* (field mouse), while it features under the lemma *ngwana* (child) in Ziervogel and Mokgokong (1975: 874) and under both *ngwana* and *mereto* (stripes) in Kriel (1965: 149 and 114 respectively).

In some cases the search for a proverb under any of the assumed key words is unsuccessful. In Endemann, for example, one would have expected with a greater than chance probability that the proverb *ngwana yo a sa llego, o hwela tharing* (the baby that doesn't cry, dies in the carry sling, i.e. if you don't make your needs known, you will not receive assistance) would have featured under the most easily-guessed-at lemma *thari*. Instead, only an explanation of *thari* is given (sling made of skin to carry a baby on its mother's back), while the proverb does not feature in the dictionary at all. The absence of this proverb under the headword *thari* is also observed in other Northern Sotho dictionaries that were consulted (Kriel 1965, 1976a, 1976b, 1977; 1983; Kriel, Van Wyk and Makopo 1989; Prinsloo and Sathekge 1996; Kriel, Prinsloo and Sathekge 1997; Mojela, Mphahlele, Mogodi and Selokela 2006 and De Schryver 2007). One would have thought that a lexical item such as *thari* would have triggered the entry of the proverb in these dictionaries because of the collocational sense relation between this word and the proverb, but this was evidently not the case. In all fairness, the omission of the proverb might have been dictated by the nature of the dictionaries, some of which are concise dictionaries, designed for learners at an elementary level where it is assumed that the priority is finding the meaning of single word items, rather than that of fixed expressions. Only Ziervogel and Mokgokong (1975) list this proverb, and they do so under the lemmas *ngwana* as well as *thari*. As discussed in the next section, it is thus possible for the same proverb to be listed more than once in the same dictionary, that is, under different lemmas, to make it more accessible to users who will be guessing at its location in the dictionary. This reveals that the choice regarding the inclusion of proverbs is a very subjective issue, but partly also dictated by the dictionary plan and purpose.

4.1 Entry points

Given their multilexical composition, there are a number of possible entry points for proverbs, which complicates the look-up process, because users may not be able to predict these entry points.

If a proverb has multiple listings, in other words, if it is listed under more than one of the lexical items contained in the proverb, it inevitably takes up extra space in a traditional paper dictionary. A further disadvantage is that, as sub-entries, proverbs only enjoy a reduced visibility due to the volume of surrounding information. In this case a special symbol or label to mark the proverb would be useful to make the multiword expression stand out among the other information. On the positive side, multiple listings will create more chances for a user to find a proverb while he/she conducts his/her search by trial and error. Cross-references could be used to guide the user to the sought-after proverb. Multiple listing is a compromise on the part of the lexicographer who would otherwise have to rely on his intuition and subjective judgement as to which of the elements deserves to be the most suitable key word under which a proverb should be listed.

The following example illustrates that the same proverb can have one listing at one entry point or multiple listings under various lemmas. In Ziervogel and Mokgokong (1975) the proverb *moeng o naka di maripa* is listed under the lemmas -ENG, -NAKA and RIPA. In Kriel (1977 and 1983) the same proverb is listed at two entry points, namely MOENG and MARIPA. In Endemann, however, only one listing was found, namely under -RIPA:

(5) **Proverb:** *moeng o naka di maripa*

Ziervogel and Mokgokong (1975)	Kriel (1977, 1983)	Endemann (1911)
-ENG, mo-/ba- a visitor should not interfere in matters concerning his host (p. 187)	MOENG 'n gas mag nie sy gasheer kritiseer nie, hy moet stilbly (1977: 147, 1983: 191)	-
-NAKA, le-/ma- a newcomer/a visitor does not meddle in other people's affairs (p. 837)	-	-
RIPA a visitor should always be well-behaved (p. 1118)	MARIPA 'n gas mag nie uitgesproke wees nie (1977: 132, 1983: 172)	-RIPA der Gast ist einer, dem die Hörner verstutzt sind, d.h. ein Gast darf nicht stossen, nicht grob sein (p. 436)

With the problems that have been outlined, it becomes clear that due to their structural nature, proverbs cannot satisfactorily be integrated in a dictionary for general purposes. Their inclusion and placement are not predictable, as theoretically, one lexicographer may decide to enter a proverb under one lemma, while another may decide to include the same proverb under another lemma, which seems more appropriate or salient to him or her. Mpofu (2007: 364) calls such prominent lexical components, 'semantically heavy' words. In the African languages there are many proverbs in which it would be difficult to identify a 'semantically heavy' word, for example, the subject may be unexpressed or the constituents may be of equal lexical and semantic importance. Here it would be difficult to decide under which lemma the proverb should be listed, e.g.

(6) *A di fule, di hlakane* (Erasmus n.d.: 7)

'Let them (domestic animals) graze and mix' (said of herd boys who should reconcile and become friends again after a quarrel — their livestock should graze together again).

The basic form of any of the verb stems in the above example (*-fula* 'graze' and *-hlakana* 'come together, mix') could technically serve as a point of entry. In this regard one could ask whether some proverbs fall between the cracks for the very reason that none of their items are considered prominent enough to trigger their inclusion under a particular lemma.

There is a tendency by users to look up proverbs under the first major component or first key word in an expression. Korhonen (2011) says that for German, the first noun, verb or adjective of a proverb usually determines under which lemma a proverb should be listed. This is perhaps more easily said of European languages than of African languages. In African languages such salient components may only appear towards the end of a proverb and one component may not necessarily be more salient than the other. It is thus difficult to decide which component of a proverb should be chosen as its allocation lemma. Example (6) above illustrates the challenge of deciding on an appropriate key component, as the example does not even contain an overt subject noun, while both verbs carry equal semantic weight. Looking up a proverb under the first key word (e.g. *moeng* in example (5)), would have been rewarded immediately in Ziervogel and Mokgokong (1975) and Kriel (1977, 1983), but not so in the case of Endemann's (1911) entry. Between the three dictionaries in example (5), it would have taken the longest to look it up in Endemann's dictionary, as it only uses the last key word (*-ripa*) as its entry point.

In special purpose dictionaries (see next section), on the other hand, in which proverbs are arranged in alphabetic order according to the first part of the phrase, be it a sub-lexical or full lexical item, no thought needs to be spared for salient lexical items.

4.2 Lemmatisation principle in special purpose dictionaries: First-element or index-based

4.2.1 Lemmatisation according to the first element

In dictionaries for special purposes such as for multiword expressions, the lemmatisation of proverbs can be done alphabetically according to the first element that they begin with. In the latter case, no decision needs to be taken on which element(s) constitute(s) salient lexical items for lemmatisation purposes, as the proverbs are merely listed according to the first part of the phrase, be it a unit smaller than a word (a morpheme) or a full lexical item. The listing of proverbs in such a publication is deliberate or purposive unlike in dictionaries for general purposes, where proverbs are often included at the lexicographer's discretion.

Rakoma's (1949) impressive collection of idioms and proverbs (*Marema-ka-dika tša Sesotho sa Transvaal*), was inspired by the concern that the older generation would take the knowledge of many fixed expressions to their grave, if not documented. Apart from idioms, the collection contains 1041 proverbs with explanations in Northern Sotho. The entries are arranged alphabetically according to the first element of the proverbs (introducers), but follow no specific alphabetical order in the microstructure under each alphabetical entry. The introducers are of various kinds, from lexical to non-lexical elements which one would not necessarily encounter as lemmatised items in general purpose dictionaries, such as negative morphemes, subject concords, hortative particles, etc.

In his collection of proverbs entitled *Uitgesoekte Noord-Sotho spreekwoorde* (Selected Northern Sotho proverbs) Erasmus (n.d.) also followed an alphabetic listing according to the introducers of the proverbs. Lemmatisation according to initial components is conducive to the user finding the required proverb quickly, even if the same element appears numerous times as the first component under a particular article stretch, e.g. G, as illustrated in the examples in (7) beginning with "Ga go" ('there is not') (Afrikaans explanations by Erasmus rendered in English by the researcher).

(7a) **Ga go** moedi mo-tlhoka-semenya (Erasmus n.d.: 25)

'there is not a valley which lacks a hollow' — nobody is perfect, every household carries a burden

(7b) **Ga go** 'šaka la poo-pedi (Erasmus n.d.: 26)

'there is not a kraal of two bulls' — you cannot serve two masters

(7c) **Ga go** sekiswe khutswane, rakhudu a le gona (Erasmus n.d.: 26)

'the little tortoise is not charged as long as father tortoise is there' — parents are responsible for their children's misdeeds, because they educate them

Finding these proverbs in a general purpose dictionary will take more effort, because the first element (which is often a sublexical item such as a prefix) may not occur in the macrostructure, but rather in the microstructure of another lemma in a word-based dictionary. Even if **ga** should occur as a headword, the users would have to navigate through a long list of sub-entries for **ga** before finding the sought-after proverb. When there are a number of proverbs which start in the same way, the microstructure of the entry becomes congested, particularly if, in addition, the proverbs are quite long. This would complicate or delay successful information retrieval. In a dictionary dedicated specially to proverbs, items can be arranged according to the letters of the alphabet as Rakoma and Erasmus have done and the length of proverbs would not constitute a problem.

A special label to mark an expression as a proverb would facilitate the identification of a proverb under a particular lemma. This practice has not been observed in any of the Northern Sotho dictionaries consulted for this research, except in the dictionary by Endemann. On the other hand, if proverbs are listed alphabetically in a special purpose dictionary, users will find it to be user-friendly, because they would approach such a dictionary with a different expectation, compared to general purpose dictionaries where the proverbs would be concealed under various entries.

4.2.2 Lemmatisation which is index-based

Another collection of proverbs in which a different approach was used compared to Erasmus' (n.d.) approach referred to above, appeared in an article published by Kuhn in the *Zeitschrift für Eingeborenen Sprachen* in 1929/1930. Kuhn listed 702 proverbs, which he remembered from his childhood days, growing up amongst the Bapedi as a missionary child. The proverbs are listed numerically in random alphabetic order, but a list with German key terms/concepts in alphabetic order appears in the appendix, with numerical references to the proverbs in which the key terms occur. For 'mouth/mouth cavity' (*Mund/Mundhöhle*), for example, there are references to 27 proverbs. Of these 27 proverbs, 11 are indexed by a second key term, 6 by a third key term and 7 by a fourth key term. Thus, this is a very useful index, as the same proverb can be located using more than one key term, increasing the chances of the user to find the sought-after proverb.

In 1938 a list of 124 proverbs in Tlokwa, a dialect of Northern Sotho, was published by Krüger, but not in any particular alphabetic or thematic order. Nevertheless, the list constitutes an important part of cultural heritage.

More recently, also driven by the passion to preserve proverbs for future generations, Motana (2004) produced an illustrated booklet of 102 Sepedi (Northern Sotho) proverbs, listed alphabetically under the initial letter of key-words that appear in his English translations; compare the following examples

under the letter B (cf. italicised *baboon*) and V (cf. italicised *vulture*) respectively in (8a) and (8b):

(8a) *Go diega ga tšhwene ke go gadimela morago*

To look backward delays a *baboon*.

People who keep on delaying, or do not focus on what they are supposed to do, often fail to accomplish a task. (Motana 2004: 11)

(8b) *Nong ye kgolo ga e rutwe go fofa.*

An old *vulture* is not taught how to fly.

You cannot teach an expert the basics. (Motana 2004: 49)

5. Dearth of proverbs in general purpose dictionaries

The *Groot Noord-Sotho Woordeboek* by Ziervogel and Mokgokong (1975) by far provides the largest number of proverbs compared to the other dictionaries that were consulted for this study (Endemann 1911, Kriel 1965, 1976a, 1976b, 1977, 1983; Kriel, Van Wyk and Makopo 1989; Prinsloo and Sathekge 1996; Kriel, Prinsloo and Sathekge 1997; Mojela, Mphahlele, Mogodi and Selokela, 2006 and De Schryver 2007). To date Ziervogel and Mokgokong's (1975) publication is still the most comprehensive dictionary that exists for Northern Sotho, but there are nevertheless some proverbs encountered in Endemann which couldn't be found in Ziervogel and Mokgokong, underlining the need for dedicated dictionaries on proverbs. One could surmise that the limited coverage of proverbs in most of these dictionaries can be ascribed to the fact that they are general purpose dictionaries whose priority is to explain the meaning of single-word tokens at word-level, and not to offer information at phrase-level. In dedicated volumes of proverb collections, on the other hand, proverbs do not have to compete for space under lemmas which are judged to be key words in the proverb and the listing is much more comprehensive.

The sporadic occurrences of proverbs in Endemann (1911), speak of an incidental inclusion of proverbs. The proverbs serve as illustrative sentences to demonstrate the application of a particular headword in context. They seem to have been included in a 'by-the-way' fashion, probably as and where a lexical item had a collocational sense relation with a proverb (in other words where there was a mutual expectancy that the word and the fixed expression would co-occur).

6. Microstructure of proverbs in Endemann (1911)

Looking at the microstructure of proverbs in Endemann's dictionary, one notices that not all proverbs have been treated equally. In some cases only a literal translation is provided, with no explanation, in others both a literal

translation and an explanation are provided, while in others a literal translation and an equivalent proverb in the target language are provided. Since there are gaps in the provision of meanings or interpretations of the proverbs, it would appear that the lexicographer took a decision on behalf of the user as to which proverbs are transparent and self-explanatory and which ones needed further elucidation.

The different combinations of the treatment of proverbs in Endemann (1911) as opposed to the treatment of the same proverbs in Ziervogel and Mokgokong (1975) are exemplified in examples (9) to (11). The captions of the examples in (9) to (11) reflect the information as encountered in Endemann. (Note that for the purpose of this discussion complete entries are not given, but only the headword and information pertaining to the proverb under that particular headword).

(9) Proverb, literal translation, no equivalent, no explanation

Endemann (1911: 72)	Ziervogel and Mokgokong (1975: 215)
Lemma: mo-fatamollo	Lemma: FATA
Proverb: <i>Mofatamollo wa marega o a iphatela</i>	Proverb: <i>sefatamollo se a iphatela</i>
Lit. "der im Winter das Feuer scharrt, der scharrt für sich selbst"	Lit. <i>None</i>
Equivalent: <i>None</i>	Equivalent: reap what one sows
Explanation: <i>None</i>	Explanation: <i>None</i>

(10) Proverb, literal translation, no equivalent proverb, an explanation

Endemann (1911: 279)	Ziervogel and Mokgokong (1975: 768)
Lemma: le-loka	Lemma: -LOKA
Proverb: <i>Tlou e hlabja ka lloka, tšhukudu mošimane</i>	Proverb: <i>Tlou e hlabja ka diloka, tšhukudu e be mošimane</i>
Lit. "der Elefant wird vom lloka gestochen, das Nashorn ist ein Junge"	Lit. <i>None</i>
Equivalent: <i>None</i>	Equivalent: a labourer is worthy of his hire
Explanation: Auch ein Grosser kann durch ein kleines Ding überwunden werden.	Explanation: if you work hard you attain your goal.

It is interesting to note that the imperative form of the verb is employed in the above proverb in Kuhn (1929/1930: 124) and Rakoma (1949: 135) (*Hlabang tlou ka dilokwa, tšhukudu mošimane*). The respective explanations given by these authors for this proverb amount to the following: 'through small things one can

achieve something great' and 'through patient perseverance one will reach one's goal'.

(11) **Proverb, literal translation, equivalent proverb, no explanation**

Endemann (1911: 259)	Ziervogel and Mokgokong (1975: 752)
Lemma: le-leka	Lemma: -LEKA
Proverb: <i>Maleka ga se makgone, makgone ke maboeletša</i>	Proverb: <i>Maleka ga se makgone, makgone ke maboeletša</i>
Lit. "Neulinge sind keine Künstler, Künstler sind die (etwas) immer wiederholen"	Lit. <i>None</i>
Equivalent: Übung macht den Meister	Equivalent: Rome was not built in a day
Explanation: <i>None</i>	Explanation: perseverance will be rewarded

7. **Conclusion**

In a general purpose dictionary certain proverbs may fall into oblivion, if their components are not considered prominent enough to trigger their inclusion under a particular headword. Special purpose dictionaries dedicated to the documentation of proverbs, on the other hand, will be able to document such proverbs as custodians of the culture of a people, be they of a bygone era or of modern origin. Proverbs are an integral part of a society, although life has changed to such a degree that many proverbs are no longer used. They can give insight into the way of life of earlier periods in a nation's history or cultural development and play a "social role as repositories of the ancient wisdom of a given speech community" (Manyawu 2012: 214). In the words of Čermák (2014: 47) "there seems to be a consensus that there is no substitute for them in social communication, even in our modern times and society".

Traditionally proverbs occur as sub-lemmas in the microstructure of a lemma instead of the macrostructure of a dictionary. They convey meaning like any other lexical unit and thus deserve comprehensive treatment just like other single lexical items. Otlogetswe (2012: 232) bewails the situation whereby only single words are considered as candidates for dictionary entry because it "impoverishes a dictionary and betrays a rudimentary understanding of what constitutes a word in a language". Mphahlele (2003: 163) argues that "[m]ulti-lingual items are lexical items that consist of more than one word. This combination of words is always a unit and should be treated likewise in a dictionary. Although multiword lexical items consist of more than one word, they should, according to Gouws (1991: 78), be regarded as single lexical items. These items should therefore be included as multilexical lemmata in the macrostructure of dictionaries."

Mpofu (2007: 364) states that "As a result of the changing trends in lexi-

cography, dictionaries are now lemmatising units larger than the word in a bid to meet the needs of different users". A multi-word lexical unit is easier accessed from the macrostructure "than to search for it within another entry in the microstructure" (Mpofu 2007: 361). Even easier would be the location of proverbs in an electronic dictionary (a discussion which falls outside the aim of this article), which allows access at various points and does not have the restrictions of a paper dictionary. It would also counteract the "old, unacceptable trend trying to squeeze the complex meaning of the proverb under a single word label" (Čermák 2014: 145).

A paper-based dictionary for general purposes cannot adequately cater for all the proverbs one could possibly encounter in a language. The choice as to which proverbs should be included and under which main lexical items they should be lemmatised, remains the lexicographer's prerogative and is also dictated by the target user's needs. The inclusion of fixed expressions in Endemann (1911) is clearly not a user-driven one, as entries of proverbs are ad hoc and incidental to the lemmatisation process. Only a handful of proverbs were encountered in his otherwise quite comprehensive dictionary which also lemmatises many sub-lexical items such as subject concords, possessive concords, class prefixes, etc.

In a word-based dictionary, multiple lemmatisation of proverbs under different headwords, increases the user's chances of finding the proverb and it avoids the lexicographer having to subjectively judge which element is more important compared to another.

Mphahlele (2003:166) proposed that lexicographers have a choice of treating fixed expressions such as idioms and proverbs as multilexical lemmas in the macrostructures of general or special dictionaries. General dictionaries receive more attention as "tools for achieving language standardisation, documentation and preservation", but not as repositories for "specialised knowledge domains" (Chabata 2013: 55). Specialised dictionaries, on the other hand, can facilitate access to proverbs, arranged according to topics or themes (e.g. health, God, ancestors, social relations, etc.), thus narrowing down the search field of the user and making information retrieval much more effective.

From a linguistic perspective culturally-birthered sayings and their phraseology are of particular interest to language historians. Special purpose dictionaries, dedicated to the documentation of proverbs, would go a far way in safeguarding their knowledge for posterity. E-dictionaries would particularly ensure quick and easy access to a variety of proverbs and their (minimal) individual variations. An advantage is that the dictionary searching skills of users are not taken for granted. Space restrictions typical of a paper dictionary are not a limiting factor and new proverbs can be added any time. To avoid the situation whereby proverbs are permanently lost to posterity, dedicated dictionaries for proverbs should continually be developed. If a dictionary of the calibre of Rakoma (1949) could be produced as a bilingual resource, for example, it would reach a wider readership, more so, if made available electronically.

References

- Akande, A.T. and A. Mosobalaje.** 2014. The Use of Proverbs in Hip-hop Music: The Example of Yoruba Proverbs in 9ice's Lyrics. *Proverbium* 31: 35-58.
- Čermák, F.** 2014. *Proverbs: Their Lexical and Semantic Features*. Burlington, Vermont: University of Vermont.
- Chabata, E.** 2013. The Language Factor in the Development of Africa: A Case for the Compilation of Specialised Dictionaries in Indigenous African Languages. *South African Journal of African Languages* 33(1): 51-58.
- De Schryver, G-M. (Editor-in-chief).** 2007. *Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English*. Cape Town: Oxford University Press South Africa.
- Endemann, K.** 1911. *Wörterbuch der Sotho-Sprache*. Volume VII of Abhandlungen des Hamburgischen Kolonialinstituts. Hamburg: L. Friedrichsen & Co.
- Erasmus, J.G.** n.d. *Uitgesoekte Noord-Sotho spreekwoorde*. Johannesburg, Cape Town: APB.
- Gouws, R.H.** 1991. Toward a Lexicon-based Lexicography. *Dictionaries* 13: 75-90.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PRESS.
- Korhonen, J.** 2011. *Phraseologie und Lexikografie: Phraseologismen in ein- und zweisprachigen Wörterbüchern mit Deutsch*. Burlington, Vermont: University of Vermont.
- Kriel, T.J.** 1965. *Pukantšu Noordsottho-Afrikaans, Afrikaans-Noordsottho*. Pretoria: Dibukeng.
- Kriel, T.J.** 1976a. *Popular Northern Sotho Dictionary (N.Sottho-English/English-N.Sottho)*. Second edition. Pretoria: J.L. van Schaik.
- Kriel, T.J.** 1976b. *The New English-Northern Sotho Dictionary (English-Northern Sotho/Northern Sotho-English)*. Fourth revised edition. King William's Town: Educum.
- Kriel, T.J.** 1977. *Pukuntšu Woordeboek. Noord Sotho-Afrikaans, Afrikaans-Noord Sotho*. Second revised and extended edition. Pretoria: J.L van Schaik.
- Kriel, T.J.** 1983. *Pukuntšu Woordeboek. Noord Sotho-Afrikaans, Afrikaans-Noord Sotho*. Third edition. Pretoria: J.L van Schaik.
- Kriel, T.J., D.J. Prinsloo and B.P. Sathekege.** 1997. *Popular Northern Sotho Dictionary, Northern Sotho-English, English-Northern Sotho*. Fourth edition. Cape Town: Pharos.
- Kriel, T.J., E.B. van Wyk and S.A. Makopo.** 1989. *Pukuntšu Woordeboek (Noord Sotho-Afrikaans, Afrikaans-Noord Sotho)*. Fourth revised and extended edition. Pretoria: J.L van Schaik.
- Krüger, F.** 1938. Tlokoa Überlieferungen. *Afrikanische Studien* 41(3): 16-35.
- Kuhn, G.** 1929/1930. Sotho-Sprichwörter. *Zeitschrift für Eingeborenen-Sprachen* 20: 34-73, 120-148, 208-221.
- Manyawu, A.T.** 2012. An Intertextual and Interdiscursive Analysis of Contemporary Thinking about Womanhood in Mokitimi's (1997) *The Voice of the People: Proverbs of the Basotho*. *South African Journal of African Languages* 32(2): 213-222.
- Mojela, M.V., M.C. Mphahlele, M.P. Mogodi and M.R. Selokela.** 2006. *Sesotho sa Leboa-English Pukuntšu ya dipolelo tše pedi le dihlošo/English-Sesotho Bilingual and Explanatory Dictionary*. Cape Town: Phumelela Education.
- Motana, N.** 2004. *Sepedi (Northern Sotho) Proverbs*. Cape Town: Kwela Books.
- Mphahlele, M.C.** 2003. The Lexicographic Treatment of Sublexical and Multilexical Items in a Northern Sotho Monolingual Dictionary: A Challenge for Lexicographers. *Lexikos* 13: 154-167.

- Mpofu, N.** 2007. The User Perspective in Lexicography: The Lemmatisation of Fixed Expressions in *Duramazwi Guru reChiShona*. *Lexikos* 17: 360-371.
- Mulhall, C.** 2007. *Semantics and the Lemmatisation of Idiomatic Expressions: An Analysis of Bilingual Italian–English Dictionaries*. Available at: https://www.essex.ac.uk/langling/documents/langue/langue_2007_proceedings.pdf. [Accessed 29 January 2016.]
- Mulhall, C.** 2010. A Semantic and Lexical-based Approach to the Lemmatisation of Idioms in Bilingual Italian–English Dictionaries. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*: 1355-1371. Ljouwert: Afûk.
- Otlogetswe, T.** 2012. Enriching a Dictionary Database with Multi Word Expressions. Ndinga-Koumba-Binza, H.S. and S.E. Bosch (Eds.). 2012. *Language Science and Language Technology in Africa: A Festschrift for Justus C. Roux*: 215-236. Stellenbosch: SUN PRESS.
- Prinsloo, D.J. and B.P. Sathekge.** 1996. *New Sepedi Dictionary (English–Sepedi (Northern Sotho)/Sepedi (Northern Sotho)–English)*. Pietermaritzburg: Shuter & Shooter.
- Rakoma, J.R.D.** 1949. *Marema-ka-dika tša Sesotho sa Transvaal*. Pretoria: Van Schaik.
- Tšhupa-Mabaka a Kereke.* 1959. Eerste Fabrieke, Pretoria: Northern Sotho Book Depot.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Pukuntšū ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*. Pretoria: J.L. van Schaik.

What French for Gabonese French Lexicography?*

Blanche Nyangone Assam, *Department of Foreign Languages,
University of the Western Cape, Bellville, South Africa (bassam@uwc.ac.za)*

and

Hugues Steve Ndinga-Koumba-Binza, *Centre for Text Technology,
North-West University, Potchefstroom, South Africa and
Département des Sciences du Langage, Université Omar Bongo,
Libreville, Gabon (hugues.ndingakoumbabinza@nwu.ac.za)*

and

Virginie Ompoussa, *Département des Sciences du Langage,
Université Omar Bongo, Libreville, Gabon (vompoussa@yahoo.com)*

Abstract: This paper is a response to Mavoungou (2013a) who has pleaded for the production of a dictionary of Gabonese French as variant B of the French language. The paper intends to comprehend the concept of "Gabonese French". It gives an outline of the situation of French within the language diversity of Gabon as a contribution to the theoretical perspective toward the inception of Gabonese French lexicography. Answers are given to the following questions: What French is described in existing Gabonese French studies? What French is and should be presented in Gabonese French dictionary production?

Keywords: GABONESE LEXICOGRAPHY — FRENCH — GABONESE LANGUAGES — LANGUAGE DIVERSITY — DICTIONARIES

Résumé: *Quel français pour la lexicographie du français gabonais?* Cet article est une réponse à Mavoungou (2013a) qui plaide pour la production d'un dictionnaire du français gabonais en tant que variante B de la langue française. Il vise à comprendre le concept de « français gabonais ». Il donne un aperçu de la situation du français au sein de la diversité linguistique du Gabon comme une contribution à la perspective théorique vers la création de la lexicographie du français gabonais. Des réponses sont données aux questions suivantes: Quel français est-il décrit dans les études existantes du français gabonais? Quel français est-il et devrait-il être présenté dans la production des dictionnaires du français gabonais?

Mots-clés: LEXICOGRAPHIE GABONAISE — FRANÇAIS — LANGUES GABONAISES — DIVERSITE LINGUISTIQUE — DICTIONNAIRES

* An earlier version of this article was presented as a conference paper at the annual joint conference of the Linguistics Society of Southern Africa (LSSA), the Southern African Applied Linguistics Association (SAALA) and the South African Association for Language Teaching (SAALT) held at the Potchefstroom Campus of the North-West University, South Africa, 24–26 June 2015.

1. Introduction

According to Lafage (1993: 216), in Africa, *there does not appear to be any challenge in the foreseeable future to the role of French in the functions for which it is used: administration, international relations, teaching, the media, trade, transport, tourism, science and technology, literature and so on, despite sociolinguistic and socioeconomic situations and political options being very different*. This description perfectly suits the current position of the French language in Gabon, an African country where language studies are experiencing a fast-growing and emerging discipline of lexicography.

Although it has its origins as a foreign language, being the sole official language, French has been granted the rank of national language through socio-political, economic and contextual means in Gabon (cf. Ndinga-Koumba-Binza 2004, 2005a and 2011, and Pambou 1998). This leads to ask the question whether Gabon should stop importing French dictionaries in order to make its own. In his attempt to circumscribe *a lexicographic plan for Gabon within the Gabonese language landscape*, Ndinga-Koumba-Binza (2005b: 132) explicitly asked the same question in these words: *Should Gabon keep importing its French dictionaries from France, or should the Gabonese compile their own French dictionaries, including French words and expressions exclusively used in Gabon?* The answer to such a question inevitably leads to another important one, which is to define and comprehend the place of French in the emerging Gabonese lexicography.

The paper covers the following topics:

- (i) a brief outline of the issues that present the current Gabonese French dictionaries,
- (ii) an outline of the status and the characterization of the French language within the Gabonese language landscape and Gabon's language planning perspectives, and
- (iii) a perspective toward the development of Gabonese French lexicography.

Although this article is set to respond to Mavoungou (2013a), it however aims at contributing to the theoretical foundations for the inception of Gabonese French lexicography. It comes within the framework of the strategic planning of Gabonese lexicography as developed by Emejulu (2000, 2001, 2002 and 2003), Mavoungou (2001 and 2010) and Ndinga-Koumba-Binza (2005b).

2. Problems in Current Gabonese French Dictionaries

In his survey of Gabonese French dictionaries and prospects for Gabonese French lexicography, Mavoungou (2013a: 255) indicated that the *modern era of Gabonese lexicography has witnessed the production of various dictionaries focusing on this African variety of French*. In fact, five monolingual Gabonese French dic-

tionaries (including wordlists and lexicons) had been published by the time of Mavoungou's survey, i.e. Boucher and Lafage (2000), Dodo-Bounguendza (2008 and 2010), Ditougou (2009) and Moussounda Ibouanga (2011). These dictionaries which have been recently followed by two others, i.e. Dodo-Bounguendza (DB2013) and Mavoungou et al. (2014), contrast with the majority of dictionaries produced in Gabon: they are monolingual and solely focus on the Gabonese variety of the French language, whereas Gabonese dictionary production has always been bilingual or multilingual including French and native African languages (cf. Nyangone Assam and Mavoungou 2000, Mihindou 2001).

The current Gabonese French dictionaries present a number of common problematic grounds. In this paper, attention will only be given to a few of these metalexicographical and methodological issues in the production of these dictionaries.

2.1 Theoretical and Structural Shortcomings

It is unfortunate that, except for Mavoungou et al. (2014), none of these Gabonese French dictionaries was based on a theoretical planning. Although the current Gabonese French dictionaries have never been theoretically assessed or reviewed — except for the work by Ditougou (2009) which was reviewed by Ondo-Mebiamé and Ekwa Ebanega (2011) —, it can be admitted that the lack of prior theoretical planning might explain the number of metalexicographical shortcomings that can be observed in these dictionaries.

Among others of these metalexicographical insufficiencies, one can for instance notice the numbering of the lemma entries of the dictionary as shown below in **Figure 1**. The figure displays two dictionary articles from the dictionary (DIT2009) by Ditougou (2009).

436- **Haoussa** : N. immigrants ouest-africain ou musulman. ☞
Aofien.
437- **Haut** : N. m. chemisier. *Change de haut, celui-là ne convient pas avec ce pantalon.*

Figure 1: Numbering of headword on the lemmata (DIT2009: 75)

A limited microstructure denoted by the shortness of dictionary articles can also be found in most of these dictionaries as seen below in **Figure 2**. It is also a display of two dictionary articles extracted from DIT2009.

443- **Hein non ?** *Expres. N'est-ce pas ?* ☞ *Ha bon hein ?*
444- **Houe** : N. f. Honte.

Figure 2: Very short dictionary articles and limited microstructure (DIT2009: 75)

Ondo-Mebiame and Ekwa Ebanega (2011: 356) assert that the DIT2009 presents a number of shortcomings that may require the revision of the entire dictionary for better quality.

Figure 3 below presents a dictionary article from Dodo-Bounguendza (2010)'s dictionary (DB2010) where the shortness of the article is also a reflection of a limited microstructure.

Midi - ↓ *Je veux déjeuner avec toi ce midi* .↑ *Je veux déjeuner avec toi à midi*
.Fréquent, oral, citoyen ordinaire.

Figure 3: Very limited microstructure (DB2010: 63)

An unaware reader of such an article from this dictionary would not know whether the word "Midi" is a noun, an adjective or an adverb if he/she does not understand the sentences given as examples in the dictionary article. The reader would also struggle to understand the meaning of the downward and upward arrows, and the difference between the bold-marked *midi* and the underlined *midi* without a comprehensive user guide.

Other deficiencies that can be found in the Gabonese French dictionaries are the lack of a user's guide or the inconsistency of the cross-references. Figure 4 below shows that the signs = and Syn. seem to denote the same meaning in the dictionary (DB2013) by Dodo-Bounguendza (2013) according to these three dictionary articles.

Reums n.f.inv. *Mère. Même ma Reums est alamaba. 'Ma mère est aussi de là-bas'. Syn. Match.*
Ribi (Pluriel *des ribis*) n.f. *Paire de lunettes. Ta ribi est de couleur noire. 'Ta paire de lunettes est de couleur noire'. Syn. Ribande, riby.*
Rick-ross n. m.inv. *Grosse personne. J'ai vu le gars de ta sœur, c'est un rick-ross'. 'J'ai rencontré l'ami de ta sœur il est gros (=énorme)'.*

Figure 4: Inconsistency of cross-referencing (DB2013: 113)

Moreover, although it also contains a few theoretical weaknesses, the dictionary by Mavoungou et al. (2014) was an attempt grounded on a metalexicographical planning by Mavoungou (2002a and 2011), Nsa Ndo (2010) and Nsafou (2010) as clearly stated in the preface of the dictionary (Mavoungou et al. 2014: 10). It appears to be a fact that when the compiler or one of the compilers is a trained and professional lexicographer, the dictionary is often based on some theoretical grounding and shows little metalexicographical discrepancies.

2.2 Influence of the Compilers' Initial Training and Occupation

Speaking about the dictionary by Mickala Manfoumbi (2004), Ndinga-Koumba-Binza (2006a: 299) believes that the author's occupation and initial training influenced the quality and the structures of the dictionary. It is likely the case with Gabonese French dictionaries in particular, and most of Gabonese dictionaries in general. None of the authors of available dictionaries for Gabonese languages has a formal training in theoretical and practical lexicography, except for Professor Paul Achille Mavoungou who jointly compiled and published the Yilumbu dictionary with Bernard Plumel (cf. Mavoungou and Plumel 2010), and the latest Gabonese French dictionary to date (cf. Mavoungou et al. 2014) with two of his colleagues.

In addition, most of the compilers of the Gabonese French dictionaries are by no means professional lexicographers. Professor Moussounda Ibouanga, a co-compiler in Mavoungou et al. (2014) and who has a dictionary of his own previously published (i.e. Moussounda Ibouanga 2011), is a trained sociolinguist and teaches sociolinguistics and dialectology at Omar Bongo University in the Department of Language Sciences. Apart from the two dictionaries (i.e. Moussounda Ibouanga 2011 and Mavoungou et al. 2014), most of his research work is related to language variation in Gabonese French and native Gabonese languages.

Doctor Jean-Aimé Pambou, also a co-compiler in Mavoungou et al. (2014), has formal university training in sociolinguistics and didactics. He teaches sociolinguistics and didactics in the Department of French Studies at *Ecole Normale Supérieure* (Teachers' Education College) in Libreville. His research works focus on French didactics and use in the Gabonese school context. Doctor Eric Dodo-Boungouenza is also a trained linguist with no known formal education in lexicography. He has so far published three Gabonese French dictionaries (DB2008, DB2010 and DB2013), but most of his available research works focus on language description and language planning. He teaches research methodology and historical linguistics at Omar Bongo University in the Department of Language Sciences. As for Doctor Lucien Ditougou, he has no known training in lexicography or linguistics. His available research outputs are mainly concerned with French-speaking African literature. He has been the Head of the Department of African Literatures at Omar Bongo University for the past four years.

A few common grounds can be found among the authors above who are the compilers of the current Gabonese French dictionaries. First, except for Lucien Ditougou, they are all trained linguists and teach linguistics-related subjects at higher education level. Second, except for Paul Achille Mavoungou and Eric Dodo-Boungouenza, they deal with subjects related to French for their research works and/or teaching subjects. Thus, it can be assumed that their research works and teaching subjects exposed these authors to the lexical particularities of Gabonese French to the point of compelling to work on such dictionary projects.

2.3 The French Language of the Current Gabonese French Dictionaries

The third major problem of available Gabonese French dictionaries and that leads to the present article in response to Mavoungou (2013a) is what it is understood as "Gabonese French". Gabonese French has actually been the topic of many linguistic studies. These studies were conducted in various linguistic sub-fields such as

- (i) *dialectology* (Mouloungui Nguimbyt 2002, Italia 2011),
- (ii) *didactics* (Ompoussa 1998, Itembo 1999, Makanga Mboumba 2005, Pambou 2011a and 2011b),
- (iii) *language situation* (Couvert 1982, Ogden 1984, Blanchon 1994, Mitchell 2004, Van der Veen 2006, Idiata 2008),
- (iv) *lexicology* (Bagouendi-Bagère 1999, Boucher 1997, Ompoussa 2008, Ondo-Mebiame 2008),
- (v) *morphology* and *syntax* (Ntsaga-Oyouni 1998, Mbondzi 1998, Italia 2000 and 2006, Mitchell 2009, Italia and Queffélec 2011),
- (vi) *phonology* (Tomba Moussavou 2001) as well as
- (vii) *pragmatics* (Boutin-Dousset 1989, Boucher 1998, Ompoussa 2011), and
- (viii) *sociolinguistics* (Moussirou Mouyama 1984, Boucher 1999, Bangouendi-Bagère Bonnot 2007, Moussounda Ibouanga 2008a).

From most linguistic studies on Gabonese French and from the lemmata contained in currently available Gabonese French dictionaries, it appears typically clear that the concept of Gabonese French is understood as the set of particular lexical items, collocations and locutions displayed in the French language spoken in Gabon, which are known as Gabonese particularisms or "*Gabonismes*", including *Toli-bangando* (a variety of Gabonese French to be presented further below). Mavoungou (2013a: 260) seems to support such a definition when he states that Gabonese French *is a repertoire of a variety of lexical items and expressions*. He further indicates *the idiosyncratic features of Gabonese French in vocabulary and syntax* (Mavoungou 2013a: 260) as being archaic, obsolete or old usage, abbreviation, modification of fixed expressions, usual graphic modification, change of connotation, change of denotation, loanwords from Gabonese native languages, Anglicisms and neologisms.

The present study refutes such a reductive definition of what Gabonese French is comprised of. It is herein believed that the French language that is the official language of the Republic of Gabon (according to its Constitution), the second language of the current political, intellectual and administrative elites of Gabon, and the mother-tongue of the majority of the Gabonese youth cannot be a form of language made only of lexical, phrasal and pronunciation particularisms. As it tries to give an accurate understanding of the concept of Gabonese

French, this paper suggests reconsidering the place of the French language within the Gabonese lexicography with a conception of what Gabonese French is in reality.

3. The French Language of Gabon: Status, Variants and Characterization

As stated earlier, the French language, as it is spoken in Gabon, has been the topic of various research studies, especially in the sociolinguistic domain. The most comprehensive are notably studies by Couvert (1982), Moussirou Mouyama (1984), Itembo (1999) and Mindze M'Eyeghe (2001). It is known that the language was introduced in Gabon by French colonialists and catholic missionaries in the first half of the 19th century (cf. Moussirou Mouyama 1986). This section will try to describe the Gabonese French in the three following aspects. First, it is a national and local language of Gabon. Second, it is at the centre of Gabon's language policy and language planning perspectives. And third, it is absolutely a variety of a major language, but it is also a local language with its own structural varieties.

3.1 A National and Local Language

The status of French as a Gabonese language or as a language spoken in Gabon has been scrutinized on various grounds (cf. Ndinga-Koumba-Binza 2011 and 2004, Makanga Mboumba 2007, Pambou 1998, Blanchon 1994, Moussirou Mouyama 1984 and Ogden 1984). The language has always been a language with multiple statuses in Gabon. Pambou (1998) tries to summarize these in indicating that French is simultaneously in Gabon a second language, a foreign language, a mother-tongue, first language or initial language, a real initial language, and the official language and language of education. However, Ndinga-Koumba-Binza (2011) suggested a revision of these statuses as being the following:

- (i) initially a foreign language,
- (ii) the second language of a portion of Gabon's population, comprising the colonial generation, the post-colonial generation and the younger vil-lager generation,
- (iii) the official language for all public affairs, business and administration,
- (iv) the mother-tongue and initial language of another portion of Gabon's population (mainly the youth, cf. Makanga Mboumba 2007, Mabika Mbokou 2008 and 2012, Idiata 2009), and
- (v) the national and local language, for its role as the language of intercultural and interethnic communication (between Gabonese various ethnic groups as well as between nationals and foreigners in the absence of a lingua franca, cf. Ndinga-Koumba-Binza 2006b and 2007).

The present paper adheres to the view by Ndinga-Koumba-Binza (2011: 146) who pointed out that French *has become a 'Gabonese language' and it is no longer a foreign language per se* in Gabon. French has gained a national status in Gabon. This status is strengthened by the fact that French is actually the sole dominant language at national level (cf. Ndinga-Koumba-Binza 2007: 107-108) in a country that has been inventoried with 62 native speech forms, including languages and dialects (cf. Kwenzi Mikala 1998). Ndinga-Koumba-Binza (2005b and 2007) pointed out the existence of what he coined as the Gabonese language landscape (GLL), i.e. a countryside picture of *the languages, both foreign and native, used in Gabon* (Ndinga-Koumba-Binza 2005b: 133). According to Ndinga-Koumba-Binza (2005b: 134 and 2007: 113), the GLL can be schematized as follows in **Figure 5** with French as one the local languages.

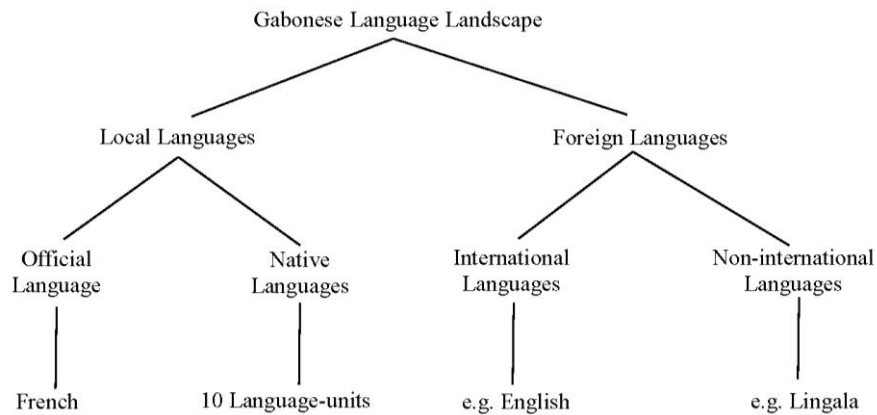


Figure 5: Schema of the Gabonese Language Landscape

The schema reflects the view that French is a Gabonese language with the same rights (or even more rights in this specific case) as a native language.

3.2 Language Policy: All Rise and Fall on French

Gabon is one these Francophone countries where the sole language policy statement is to adopt French as the official language of the nation. In fact, the Constitution (Act 2, paragraph 8) clearly stipulates: *The Republic of Gabon adopts French as the official language. Furthermore, it endeavors to protect and promote the national languages.* Ndinga-Koumba-Binza (2007: 108) believes that this statement is the *entire language policy* of Gabon.

This is true on the account that no other official, constitutional or administrative document exist in connection with languages. For instance, there is no

legal or administrative document which would define the process, the procedures or the actions which Gabon intends to implement in order to *protect and promote the national languages*. The government willpower to promote any other language than French limits itself to public speeches and support to linguistics workshops and seminars (cf. Dodo-Bounguendza 1999 and 2008). Thus, it can be argued that if Gabon's language policy does exist, French is certainly at the centre of it.

Ndinga-Koumba-Binza (2005a: 67) states that the current language policy of Gabon is indeed a continuation of the language policy implemented by the colonial government of France in the entire French colonial empire, especially in Africa. In fact, it is only with the constitutional amendment of 1994, i.e. thirty-four years after attaining independence, that the provision to *protect and promote the national languages* was added to the Act 2, paragraph 8, of the Constitution (cf. Idiata 2002: 74).

Thus, before 1994 the language policy was constitutionally more like in the colonial era when language use in French territories was organized according to two government decrees of the French government of the time (cf. Kwenzi Mikala 1990: 123). The first was the Villets-Cotterêts decree of 1539 which forbid the use of any other language than French in all official and administrative domains in the French territories (cf. Trudeau 1983, Fumaroli 1992, Boulard 1999 and Cohen 2003).

The second decree, the decree of the 14th February 1922 (cf. Spencer 1971, Garcia 1971 and Wakely 2000), was concerned with language in education making the difference between government public education, i.e. formal general education, and private education which comprised only religious education. In all African territories of the former French empire, the formal general education ought to be exclusively in French and the religious education regarding Bible teaching and any religious training ought to be in native languages (cf. Spencer 1971, Garcia 1971, Judge 1993a and 1993b and Wakely 2000).

In the specific context of Gabon, which attained international sovereignty in 1960, the language in education policy has in practice been a continuation of the colonial regime policy (cf. Ndinga-Koumba-Binza 2005a: 66-70, 2005b: 139 and 2007: 108; Kwenzi Mikala 1990: 122-123). This is due to the fact that, as mentioned by Moyo (2002: 149), *after the attainment of independence, most African governments, for functional purposes in official circles, found it easier to retain their colonial masters' languages as official languages*.

Despite the constitutional amendment of 1994 to *protect and promote the national languages* (Constitution of the Republic of Gabon, 1994, Act 2, paragraph 8), the dispensations of the 1539 and 1922 decrees of the French colonial government are still in full implementation in the current Gabon's language policy and education system.

It has been shown that this key position of the French language, despite the fact that the language is now a Gabonese language per se, has caused seri-

ous damage to the survival of Gabonese native languages (cf. Idiata 2008, Mitchell 2004 and Ndinga-Koumba-Binza 2004). Thus, French has arisen to becoming a nationally-dominant Gabonese language at the expense of native languages due to a historically-kept favorable language policy.

3.3 A Language Variety with its Own Varieties

Mouloungui Nguimbyt (2002) has shown the existence of dialectal variations within the French language as it is spoken in Gabon. This is true in the sense that linguistic markers are found in the way Gabonese speak French (cf. Minko 2008): Respective Gabonese mother-tongues do have an influence on the French fluency of a number of Gabonese speakers. It is often popularly said that Fang speak French with Fang accent, Punu with Punu accent, Myene with Myene accent, Nzebi with Nzebi accent, etc. (cf. Moussirou Mouyama 1998, Itembo 1999).

However, varieties within Gabonese French are not limited to the influence of local mother-tongues in the speaking proficiency. Varieties for which this study is concerned are those of linguistic forms as sociolinguistically identified (cf. Mesthrie et al. 2009: 294, Stockwell 2007: 24), i.e. the acrolect, the mesolect and the basilect. The acrolect is referred to as the *prestige or standard variety (or lect) to which it is possible to compare with other lects* (Crystal 2008: 8). The mesolect is the *intermediate linguistic variety* (Crystal 2008: 302) whereas the basilect is the *most remote language variant from the prestige language* (Crystal 2008: 51). These variants have been identified in Gabonese French as seen in **Figure 6** by Mavoungou (2013a: 259).

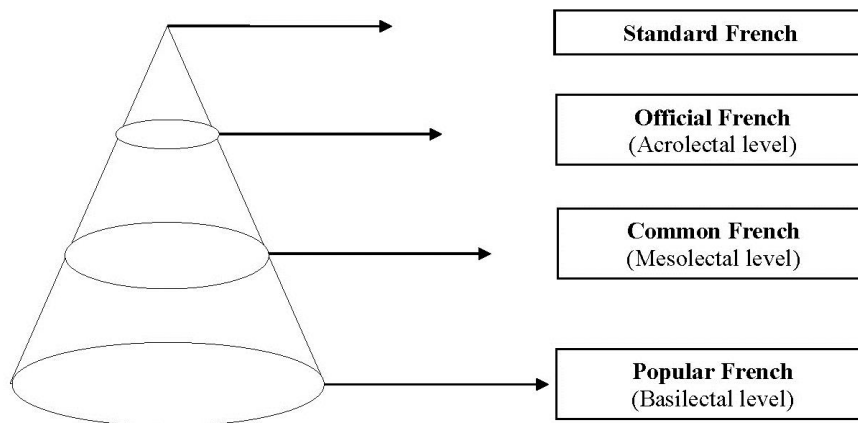


Figure 6: Gabonese French variant levels (Mavoungou 2013a: 259)

The view in this article does not agree with this stratification of the way the French language presents itself in Gabon. First, the term "official French" cannot be identified with a sociolect. In fact, what is known as "official language" in any country is not a language variety *defined on social (as opposed to regional) grounds, e.g. correlating with a particular social class or occupational group* (Crystal 2008: 440).

Second, it appears possible to make a distinction between upper mesolect and lower mesolect within Gabonese French. In fact, the concept of mesolect can cover a range of varieties. The upper mesolect is the closest variation to the acrolect and the lower is closer to the basilect (cf. Crystal 2008: 302). This distinction is suitable for understanding the differences within the mesolectal variety of Gabonese French.

In the context of Gabonese French, the acrolectal level is indeed standard French, also known as Parisian French. The variety is well used in Gabon in all official spheres including administration, education and business. It is best found in official written documents, various newspapers, academic essays and school textbooks. Administrative notes, minutes, reports as well as exam papers are required in standard French. This is one of the reasons why it is often said that Gabonese French is closer to standard French than a number of other French varieties in Africa (cf. Pambou 2011a, Ntsaga-Oyouni 1998). This has been indeed proven in the study by Mitchell (2004) who showed that most Gabonese adults and young adults perceive the form of French they speak as close to the standard form.

Furthermore, from the definition of Gabonese French by Mavoungou (2013a: 260) and from the contents of currently available Gabonese French dictionaries, it appears that what is often called as Gabonese French is merely the mesolectal form of the French language used in Gabon. It comprises the upper mesolectal level and the lower mesolectal level. The upper mesolectal form, as it is close to the acrolect, can be found in official circumstances, especially in public speeches. A form of language which late President Omar Bongo and his regime had often made use of in public circumstances (cf. Ndenguino Mpira 2005, Mebiame-Akono 2008a and 2008b). It was also the form of language used by the then Prime Minister in his obituary speech at the official funeral ceremony of Omar Bongo (cf. Mebiame-Akono 2009). The current President also makes use of the same linguistic strategies in the same circumstances (cf. Mebiame-Akono 2011, Engoung Nsi 2013). The upper mesolect is also the form of language used in most talk-shows on TV and radio (cf. Alves 1994, Artigues 1995) and a number of private newspapers (especially satirical newspapers). It is also found in various academic circumstances.

Table 1¹ below displays a few examples of the differences between the acrolect and the upper mesolect in Gabonese French (extracts from DB2010²). The first column contains words in the mesolectal form of Gabonese French. The second column presents words in the acrolectal form. The third and last column shows the translation of the words in formal English.

<i>Upper Mesolect</i>	<i>Acrolect</i>	
Aboyer. <i>Ce chiot aboie très fort.</i>	Japper. <i>Un chien aboie, un jeune chien n'aboie pas, mais il jappe.</i> (DB2010: 26)	<i>To yap</i>
Etre allergique à. <i>Les personnalités politiques sont allergiques aux critiques.</i> (DB2010: 19)	Ne pas apprécier. <i>Les personnalités politiques n'apprécient pas les critiques.</i> (DB2010: 19)	<i>To resent</i>
Bénéfique. <i>L'augmentation du prix du baril de pétrole est bénéfique pour l'économie gabonaise.</i> (DB2010: 26)	Favorable à/Avantageux pour. <i>L'augmentation du prix du baril de pétrole est favorable à l'économie gabonaise.</i> (DB2010: 26)	<i>Beneficial to</i>
Blessé grave/blessé léger. <i>On a enregistré six blessés graves et deux blessés légers lors de l'accident qui a eu lieu au P.K. 13.</i> (DB2010: 26).	Grièvement/légèrement blessé. <i>On a enregistré six personnes grièvement blessés et deux légèrement blessés lors de l'accident qui a eu lieu au P.K. 13.</i> (DB2010: 26).	<i>Seriously/slightly injured</i>
De rien du tout. <i>C'est une petite maison de rien du tout.</i> (DB2010: 45)	Sans importance. <i>C'est une petite maison sans importance.</i>	<i>Insignificant</i>

Table 1: Acrolectal vs. upper mesolectal differences in Gabonese French

The differences from one variety to the other — from the acrolect to the upper mesolect, from the upper mesolect to the lower mesolect and from the lower mesolect to the basilect — can be identified at four levels of the speech, i.e. at lexical, morphological, syntactical and at phonetic levels. At lexical level, there are lexical creations through neologisms, collocations as well as loanwords from African languages. The items below in **Table 2** are few examples of the lexicon of Gabonese French. Columns 1 to 4 respectively present the acrolectal, the upper mesolectal, the lower mesolectal and the basilectal lexica. Lastly, column 5 records the English translations. Data contained in this table were extracted variously from available dictionaries of Gabonese French.

Acrolect	Upper Mesolect	Lower Mesolect	Basilect	
Beignet	Gâteau	Gâteau farine	Bédoume	<i>fat cake</i>
Bidonville	Ghetto	Matiti	Mapane	<i>township</i>
Friperie	Marché	Moutouki	Assamsè	<i>Frippery</i>
Gâteau	Gâteau	Gâteau au four	Cake	<i>cake</i>
Sans importance	De rien du tout	De rien	Au balango	<i>insignificant</i>

Table 2: Lexical differences in Gabonese French varieties

At morphological and syntactical levels, there are also neologisms, but most morphological creations are from an influence of native African languages (cf. Italia 2006). **Table 3** below shows an example of the syntactic differences between the four levels of language in Gabonese French. The differences are shown from a single sentence extracted from a dictionary article by Mavoungou et al. (2014: 69). The basilectal sentence is presented in the dictionary as the headword of the dictionary article. The last row of the table below presents the sentence translation in formal English.

Acrolect	Ce ne sont pas des ragots.
Upper Mesolect	Ce ne sont pas des racontards.
Lower Mesolect	Ce n'est pas le kongossa.
Basilect	C'est pas on a dit.
	<i>This is not gossip</i>

Table 3: Syntactic differences in Gabonese French varieties

The influence of the native African language is also responsible of linguistic markers from pronunciation and prosodic structures, as detailed and analyzed in Minko (2008). For instance, in the difference between the acrolect and the upper mesolect, Dodo-Bounguendza (2010) indicates that the last consonant in the words **août** "August" (DB2010: 22) and **baril** "barrel" (DB2010: 25) are not pronounced in standard French. These consonants are ordinarily pronounced in Gabonese French upper mesolect.

Furthermore, unlike most previous studies (Boucher 1998, Bagouendi-Bagère Bonnot 2007, Moussounda Ibouanga 2008b and 2011, Mitchell 2009, Mouélé 2011 to mention a few), the current study makes a distinction between the basilectal form of Gabonese French and Toli-bangando. In fact, although Toli-bangando is considered as a social variety of Gabonese French, it rather bears the characteristics of slang than of a language basilect. In the distinction between basilect and slang, although they are both language varieties that are the most remote from the prestige variety, it is true that the basilect has more the status of common variety while slang is more *an in-group language for a particular set of people such as teenagers, army recruits, pop groups*, etc. (Richards and Schmidt 2002: 490). Thus, in places where a creole is spoken, the creole is the basilect of the language and it is often spoken by the maximum majority of the population. For instance, Irvine (2004) and Sand (1999) show that Jamaican creole is the basilect while Standard Jamaican English is the acrolect in Jamaican English. Likewise, Jamaican creole is different from Jamaican English slang (cf. Chang 2014, Patrick 1999).

The basilect of Gabonese French was brought to written form and litera-

ture by Ndong Mbeng (1992)³. In his well acclaimed novel, Ndong Mbeng (1992) mixes narration and description to report on life in the slums of Libreville, the capital city of Gabon. The narrative is given through a lexical, syntactic and stylistic structure that is proper to the populations of Libreville's shantytowns. The language variety that reads in Ndong Mbeng (1992)'s work ignores norms of standard French. Yet, the novel was acclaimed not only for its depiction of shantytowns' living conditions, but mainly for its use of the Gabonese French basilect, a language variety understandable by the population's majority, especially the people depicted in the novel (cf. Van den Avenne 2001 and 2004).

As for Toli-bangando, it is actually a jargon of teenagers and a certain youth of popular neighborhoods, as described by Moussounda Ibouanga (2008b and 2011). It is also the most common language variety used in Gabonese pop music. Dodo-Bounguendza (DB2013: 8), who acknowledges it as slang, indicates in the pretexts of his Toli-bangando dictionary, that the language variety leans on distortion, incomprehensibility, a peculiarity to young people and a trait marked by complication for the grown-ups. It is indeed an informal and original language invented by young Gabonese. It is mainly based on direct creation of words. This allows to conceive or to imagine words from scratch.

The language variety is also marked with extensive use of metaphors, semantic extensions and shifts and with words from various other languages than French, notably English, Spanish and native languages. For instance, the term *Toli-bangando* itself derived from two Gabonese native languages (*toli* "language" from Myene and *bangando* "caimans" from Punu) to mean "language of gangsters". From its meaning and from the description that can be made of it, Toli-bangando is more like Tsotsitaal (this glossonym also means "language of gangsters" in Afrikaans) in the South African context (cf. Slabbert and Myers-Scotton 1996, Satyo 2001 and Ntshangase 2002) as shown in the Tsotsitaal dictionary by Molamu (2013).

A few of characteristics of linguistic shifts and norm distortions in Toli-bangando are shown in the following illustrative sentence: *Coince moi mon bag, je vais d'abord check avec l'élément là* (Dodo-Bounguendza 2013: 39). Two English words are recorded in the sentence, i.e. *bag* and *check*. The meaning of the latter word is shifted to mean "to talk". The French words *coince* and *élément* also have meanings different from what they have in standard French. The latter word is borrowed from the military jargon where it means "an individual". Finally, the meaning of the sentence in formal English is: "Would you please carry my bag, I need first to talk with this person". Ultimately, the levels of variation indicated by Mavoungou (2013a: 259) in **Figure 6** further above do not reflect the actual sociolectal stratification in Gabonese French. **Figure 7** below is an attempt to schematize a more actual stratification with all identified sociolectal variants of Gabonese French.

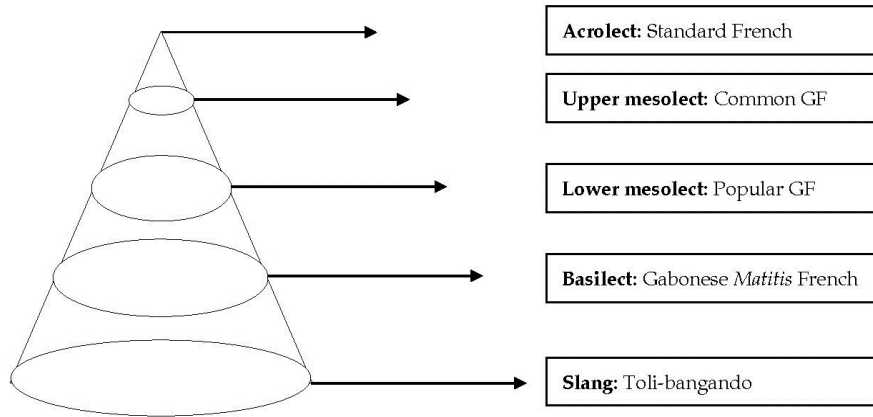


Figure 7: Sociolectal stratification of Gabonese French

The schema above has the advantage of displaying both the diastratic and the diaphasic variation of the Gabonese French, which contains varieties highly differentiated between the standardized and the popular poles (cf. van den Avenne 2004: 45). Subsequently, the distinction made by Mavoungou (2013a) between variant A and variant B, where variant A is supposedly the variety of France and variant B the Gabonese French, does not apply in the actual context of Gabon.

As a matter of fact, the distinction would not apply in any French-speaking country, because every country has indeed its own phonetic, lexical, morphological and syntactical particularities that diverge from Parisian French. It is for instance specifically the case in Belgium (Belgicism, e.g. *bourgmestre* "mayor" rather than Parisian French *maire*), Cameroon (e.g. *frère de famille* "cousin" instead of standard French *cousin*), Senegal (Senegalism, e.g. *essencerie* "petrol station" rather than Parisian French *station d'essence*) and Switzerland (e.g. *bordier* "neighbor" instead of standard French *voisin*).

One would note that even in France, there is not one variety of French as shown in France's French regional varieties dictionary by Rézeau (1999). Although, French is the national and official language in France with Parisian variety as the standard norm, the language still has many varieties which differentiate themselves not only regionally (north versus south of France, and west versus east of France) but also historically from the influences each variety gained from neighboring languages as well as from languages of ancient ethnic tribes that inhabited France long before French was formed as a language.

Thus, there should be no hierarchy such as variant A versus variant B, because there exists only one French language which does have a number of varieties not only within France, but also beyond the borders of France and overseas. It is the view of this paper that each French variety in a given country be legitimized, i.e. accepted as one of the various speech-forms of a particular

language following the codification of such a variant in a determined country (cf. Klinkenberg 2002).

Examples of such legitimization and codification exist in English where American English is a different norm from British English, and from South African English. For the specific case of French, one may notice that most codified French variants are worldwide recognized as specific varieties of the French language. Like for most major world languages (e.g. Arabic, Chinese, English, German, Spanish, etc.), the codified varieties can be found in the language repertoire of the Microsoft Office (Ms Office) applications package.

Figure 8 presents the French varieties codified and included in the Ms Office applications package for spelling and grammar checks. These are Belgian French, Cameroonian French, Canadian French, Congolese (DRC) French, Ivorian French, France's French (i.e. Parisian French), Haitian French, Luxembourg French, Malian French, Monaco French, Moroccan French, Reunion French, Senegalese French, Switzerland French and West Indies French.

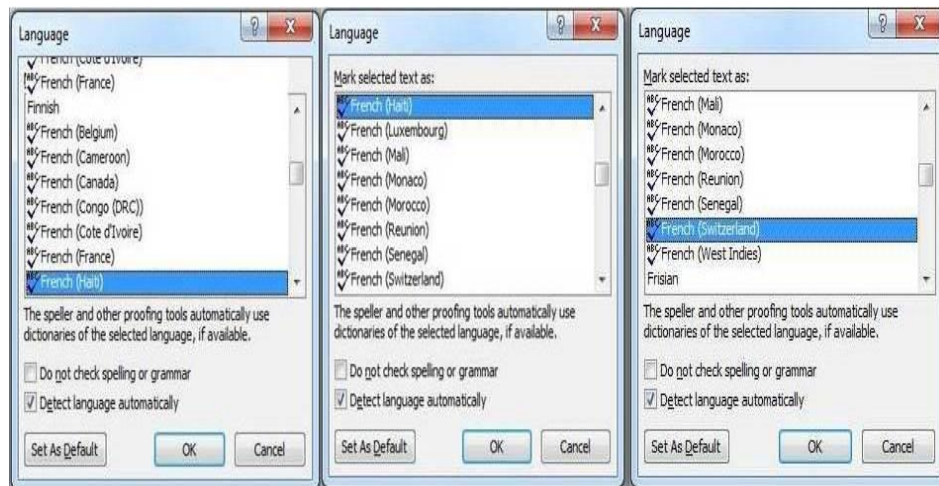


Figure 8: Codified varieties of French available on Microsoft Office

4. Grounding Gabonese French Lexicography

The previous two sections have tried to give a clear and actual definition of Gabonese French in terms of its sociolectal variants and its relationship with Parisian French. It is herein assumed that such a definition can be an important step towards Gabonese French lexicography, lexicographic activities (metalexigraphic research and dictionary production) which would have Gabonese French as sole object.

This section will now focus on few relevant matters related to metalexico-

graphical research and dictionary compilation in the theoretical inception of Gabonese French lexicography. In fact, in addition to the prospects and challenges for French lexicography in Gabon as outlined by Mavoungou (2013a: 268), it appears relevant to reflect on the purpose and objectives of Gabonese French lexicography and the necessary phases towards establishing Gabonese French lexicography as both a research focus and professional activity in the Gabonese context.

This section will thus be concerned with a brief outline on current trends in Gabonese lexicography since its first mention by Emejulu (2000) and inception by an important literature that followed suit (e.g. Nyangone Assam and Mavoungou 2000, Mavoungou 2001 and 2010, Ndinga-Koumba-Binza 2005b) and the suggestion of a strategic focus for Gabonese French lexicography. Prior to the above concerns, the section is introduced with a tentative definition of what it should be understood as Gabonese French lexicography.

4.1 What is to be called Gabonese French Lexicography?

In their attempt to determine the scope of lexicography as an independent discipline, Bergenholtz and Gouws (2012) suggest that lexicography is *the discipline dealing with theories about recently completed and also older existing dictionaries but also about future dictionaries as planned and produced by lexicographers* (Bergenholtz and Gouws 2012: 39).

A language-specific practice of this discipline is called the lexicography of that specific language. It can thus be understood that the lexicography of a language or a group of languages is the set of lexicographic traditions and activities (theoretical research and reference works compilation) conducted in or for that language or that group of languages. This is seen in the recent encyclopedia of lexicography (Gouws et al. 2013) where, for instance, the following chapter titles are found: *Lexicography of the Nguni languages* by Gauton (2013), *Lexicography of the Sotho languages* by Prinsloo (2013), *Lexicography of Shona* by Chabata (2013) and *Lexicography of Fang* by Afane Otsaga (2013) to mention just a few.

In the light of this understanding, what is to be called Gabonese French lexicography has then to be the set of theoretical and practical works done on the French language as it is spoken in Gabon within the research discipline defined above according to Bergenholtz and Gouws (2012: 39). It should also be herein said that Gabonese French lexicography is of course a component of Gabonese lexicography.

A comparison with other French-speaking African countries, and particularly in Central Africa and West Africa, shows that the inception of Gabonese French lexicography is a first of its kind in Francophone Africa. This is clearly observable in the survey of lexicographic activities in Central (cf. Mavoungou 2013b) and West Africa (cf. Mavoungou 2013c). In fact, Mavoungou (2013b and 2013c) shows no lexicographic activities for the French varieties spoken in these African countries. It should equally be noted that in most African countries sur-

veyed *these lexicographic reference products are translation dictionaries* (Mavoungou 2013b: 956) pairing with one or two colonial European languages (e.g. Dutch, English, French, German, Portuguese and Spanish). This however does not elude the fact, despite the absence of particular focus on lexicographic works in French, the language varieties have variously been codified in a number of these countries (cf. Boutin 2003, Ploog 2000 and 2001, Cissé 2005, Wamba and Noumssi 2003, Zang Zang 1998 and 2013, Mendo Ze 1999) as seen on **Figure 8** earlier in this paper.

4.2 A Note on Current Trends in Gabonese Lexicography

After more than fifteen years of activity, one would hardly argue that Gabonese lexicography is still "embryonic", as Ndinga-Koumba-Binza (2005b: 136; 2006a: 302) had once claimed. It has actually emerged into an important research and teaching discipline at Omar Bongo University. This is seen through the different survey works that have been made of Gabonese lexicography since its inception as a research discipline early in the 2000's.

The first survey was published by Nyangone Assam and Mavoungou (2000). It is a comprehensive account of lexicographic works and studies in Gabon. By the time of this survey most available dictionaries of Gabonese native languages were works compiled by religious missionaries and colonial administrators during the colonial era with reprints and/or new editions published by the Raponda-Walker Foundation. It is also learnt from this survey that most of these dictionaries *are bilingual dictionaries and biased towards French* (Nyangone Assam and Mavoungou 2000: 269).

Following the survey by Nyangone Assam and Mavoungou (2000), within the framework of defining a lexicographic plan for Gabon, Ndinga-Koumba-Binza (2005b) suggested a short review analysis of trends in Gabonese lexicography which he estimated as "embryonic" at the time. One of the major lessons learnt from Ndinga-Koumba-Binza (2005b)'s survey is the distinction between "earlier era" and "modern era" of dictionary production in Gabonese lexicography. According to Ndinga-Koumba-Binza (2005b: 138), the "earlier era" of Gabonese dictionary production is that of religious missionaries and colonial administrators, while the modern period *coincides with the completion of the first theses on metalexicographic planning of dictionaries in the Gabonese languages* (Ndinga-Koumba-Binza 2005b: 138) in the 2000's⁴. Ndinga-Koumba-Binza (2005b: 136) then outlined five development simultaneous phases of Gabonese lexicography. These phases are lexicographic training, metalexicographical accounting for dictionaries, circumscribing lexicography as an academic discipline and a career, formulating a coherent general framework for Gabonese lexicography, and updating reference works inherited from missionaries and colonial administrators.

For the present overview of the current trends in Gabonese lexicography, it can be appropriate to review these same phases into two aspects, i.e. lexicographic training and lexicographic domains of publications in Gabonese lexi-

cography. The aim of this short overview is mainly to answer the question of the place of Gabonese French in all that has been done or is being done in Gabonese lexicography.

4.2.1 Lexicographic training in Gabonese lexicography

Training is the one aspect where the strategic planning for Gabonese lexicography has been quite successful for a country that has just fewer than one million and half inhabitants. Following the ten Gabonese students who obtained their respective Ph.D. degrees in lexicography at Stellenbosch University under Professor Rufus Gouws between 2002 and 2010, lexicography is now one of the major disciplines taught from undergraduate to postgraduate levels in two departments at Omar Bongo University, namely the Department of Language Sciences and the Department of German Studies.

Although a Master program in lexicography has not yet been launched, there has been since 2005 an average of two to three students every year graduating in language sciences with a Master thesis topic related to lexicographic issues. At present, five other students are enrolled for doctoral degree with lexicography research topics at Omar Bongo University under Professor Mavoungou. **Figure 9** below shows a survey of lexicographic training in Gabon between 1999 (when the first Gabonese enrolled for lexicography studies at Stellenbosch University) and 2010 (when the two last Gabonese Ph.D. students graduated from Stellenbosch University). The survey was conducted by Ndinga-Koumba-Binza (2010) in the prospects for an objective assessment of the lexicographic plan for Gabon.

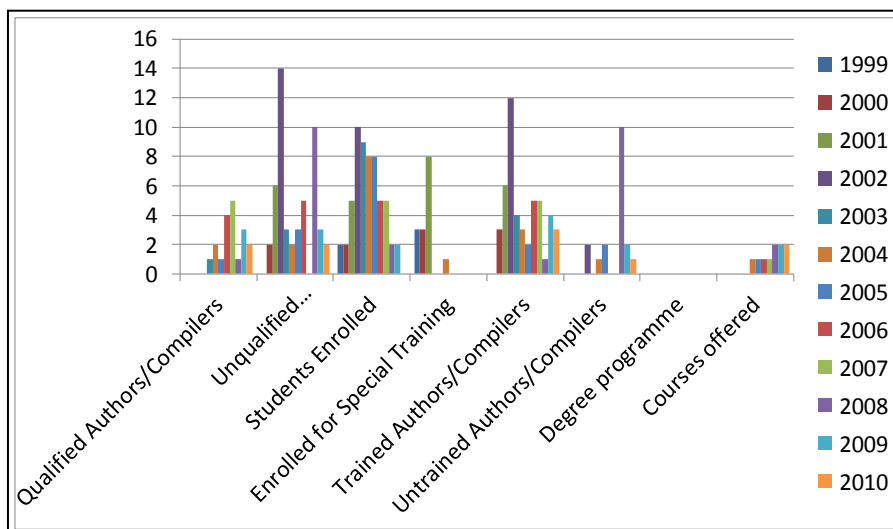


Figure 9: Survey of lexicographic training for Gabonese scholars

It can be seen from **Figure 9** above that the number of students enrolling for lexicography-related courses is increasing throughout the year while the number of lexicography modules offered is quite stagnant.

However, three recent developments allow hoping and expecting for new courses and more student enrollment and supervision in the near future in the Department of Language Sciences and the Department of German Studies at Omar Bongo University. First, both departments have been engaged into a process of restructuring in terms of study programs as from the year 2016. Second, the seven trained and qualified lexicographers who are the teaching staff in these two departments have been promoted to senior teaching roles: Apart from Professor Mavoungou who is an Associate Professor in both departments, there are now three lexicographers who are Senior Lecturers in the Department of Language Sciences, and three other lexicographers are Senior Lecturers in the Department of German Studies.

Third, the newly founded Research Centre for Germanic and Intercultural Studies⁵ (CREGI) headed by Professor Mavoungou within the Department of German Studies has recently established a division for lexicography research and courses. The division was launched at the occasion of the international colloquium on Gabonese dictionaries hosted by the CREGI at Omar Bongo University on 25–27 November 2013. The lexicography division of CREGI is indeed expected to contribute to increasing research outputs in Gabonese lexicography.

4.2.2 Lexicographic domains of publications

In terms of research outputs and dictionary compilation, Gabonese lexicography has experienced a fast-growing crop of publications since 2000 (cf. Ndinga-Koumba-Binza 2005b: 136). From the survey by Ndinga-Koumba-Binza (2010), it is learnt that domains of publications in Gabonese lexicography includes pedagogical lexicography (i.e. learners' lexicography), planning, bilingual lexicography, multilingual lexicography, monolingual lexicography, encyclopedic lexicography, corpora and Gabonese French.

It is shown in **Figure 10** below that the second top domain with the most research outputs in Gabonese lexicography is Gabonese French (16%), behind bilingual lexicography (27%), for the period between 1999 and 2010 in Ndinga-Koumba-Binza (2010)'s survey.

Two significant points can be drawn from **Figure 10**. First, although Gabon still lacks a formal degree program in lexicography, the crop of Gabonese lexicography literature is in constant progress. This is the consequence of not only the number of highly-trained lexicographers, but also of an affirmed trend to produce metalexigraphically-accounted dictionaries and to update with theoretical bases reference works stemming from the early era of Gabonese dictionary production (cf. Mavoungou 2001, Mihindou 2001, etc.). Second, despite the lack of an asserted intention to produce dictionaries for Gabonese

French, there has been an interest in lexicographic research on Gabonese French within the emerging Gabonese lexicography. This is an undoubtedly an important step towards a formal inception of Gabonese French lexicography.

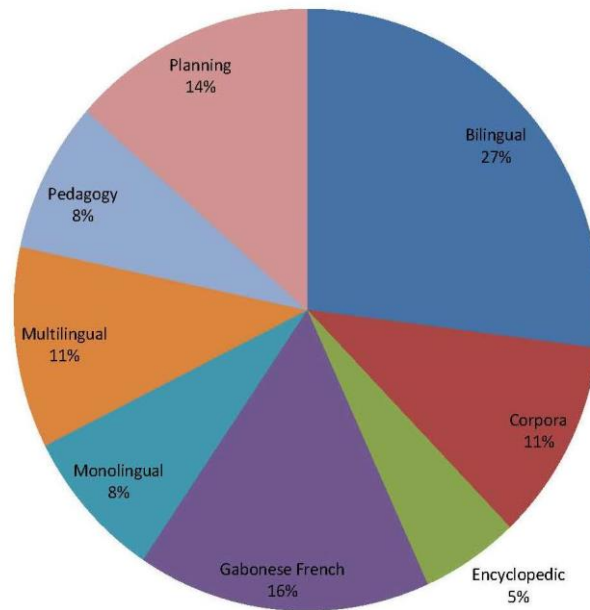


Figure 10: Survey of domains of publications in Gabonese lexicography

4.3 Strategic Focus for Gabonese French Lexicography

It is the view of the present study that the development of Gabonese French lexicography without a specific goal or a strategic theoretical framework might run the risk of some methodological inadequacy. It is for instance hard to find the genuine purpose of any of the currently available dictionaries of Gabonese French. Most of these dictionaries were made within what Ndinga-Koumba-Binza (2006a: 298) referred to as the *epistemology of the missionaries*⁶. According to Ndinga-Koumba-Binza (2006a: 298), the main point of the epistemological framework of the religious missionaries and colonial administrators who were the pioneers in scientific research in Gabon was to expose indigenous languages and cultures by means of linguistic and anthropological studies and dictionary compilation (most of them were not even acquainted with theoretical concepts and research methods of these disciplines). The purpose of such an epistemological framework was simply to make the indigenous languages and cultures known to their Western compatriots.

Equally, it is likely plausible to say that the current dictionaries of Gabonese French have no other aim than the exposition of some part of the Gabonese French lexicon (as described in most sociolinguistic studies) in lexicographic manners. A theoretically-conceived aim for Gabonese French lexicography is crucially needed in view of its scientific sustainability.

Such an aim can be conceived within what it should be called as the **strategic focus for Gabonese French lexicography**. It is actually a twofold focus. First, it will have to make a stand for only monolingual lexicography and monolingual dictionary production. One of the reasons for this choice is because the maximum majority of Gabonese research and dictionary production activities focus on bilingual lexicography with French as one of the members of the language pair. This will definitely help to avoid unnecessary duplications.

A second reason for such focus choice is that Gabonese French lexicography can also give itself the contribution to the codification of the French variety of Gabon (with its cultural and local inputs part of the French language as it is known worldwide) as a research objective with a societal implication. This will indeed help the Gabonese French lexicographers to step in the experiences of French-speaking African countries such as Cameroon (Zang Zang 1998, 2013), Senegal (Ndao 2002) and Cote d'Ivoire (Ploog 2000, 2001).

Second, as part of the strategic focus, Gabonese French lexicography will have to be a corpus driven lexicography. Thus, building a corpus of Gabonese French will be one of the most important steps in Gabonese French lexicography. The Gabonese French corpus will have to take into consideration the different sociolectal variants of the language the way it is spoken in Gabon. The dictionary production will have to stem not only from an annotated and clean corpus, but also from sound metalexigraphic research setting theoretical guidelines for all necessary types of French dictionaries in the Gabonese context.

Moreover, the strategic focus for Gabonese French lexicography should include the setting up of a lexicography unit for Gabonese French. In fact, the South African experience (cf. Gouws and Prinsloo 2005, Alberts 1996), or more particularly the Afrikaans lexicography (cf. Gouws 2013), has shown the importance of a language lexicography unit for the development of both lexicography research and dictionary production for a specific language.

Based on the South African experience, Emejulu (2001: 50-52) has suggested the establishment of lexicography units for Gabonese languages as a key factor for the development of Gabonese lexicography. Emejulu (2001: 50) suggests that lexicography units have to be made for Gabonese languages according to the 10 language-units identified by Kwenzi Mikala (1998: 217) from the 62 speech-forms (including languages and dialects) of the native Gabonese languages. Unfortunately, such a planning of lexicography units for Gabonese languages does not take Gabonese French into account. Meanwhile, the inception of Gabonese French lexicography will definitely take an important turning point with the launch of a lexicography unit for Gabonese French.

5. Conclusion

The current emerging Gabonese lexicography presents an interesting position for the French language, especially in terms of metalexicographical research. At the same time, while Gabon is experiencing a modern era of dictionary production in general, current Gabonese French dictionary production has solely focused on popular French and slang which are not the sole representatives of the Gabonese French variety. Gabonese French lexicography should target all the variants of the French language spoken in Gabon, including the acrolect.

Suggestions have been made in this paper to ground Gabonese French lexicography on a strategic focus. This strategic focus will have to include metalexicographic research and corpus building for all types of monolingual dictionaries production of Gabonese French as the main activities of a Gabonese French lexicography unit. As Lafage (1993: 215) puts it: *It seems to be increasingly accepted that the survival of French as an international language depends on its development in Africa.* This is where Gabonese French lexicography can play a significant role.

In conclusion, within the strategic planning for Gabonese lexicography, French should be as important as any Gabonese local language.

Notes

1. It should be noted that the lexicographical data contained in Table 1 are dictionary articles from the DB2010. The page number where the article was extracted is duly indicated. The headwords of these dictionary articles are genuine examples of Gabonese French which can variously be found in popular Gabonese newspapers, magazines, etc. (cf. Ondo-Mebiame 2008, Mindze M'Eyegehe 2001). However, a proper corpus of Gabonese French still has to be built. All English translations are our own in this paper.
2. This dictionary by Dodo-Boungouenda (2010) is an important resource book for distinguishing between the acrolect and the upper mesolect in Gabonese French. It identifies the shortcomings of Gabonese French and recommends the proper usage according to standard French. It is aimed at politicians, journalists, administration staff and academics.
3. Van den Avenne (2004: 45) argues that the language used in the narrative can be a mixed of Gabonese popular French and slang. However, by the time the work of Ndong Mbeng (who was 19 years old when his book got published) Toli-bangando, which was unnamed then, had not gotten the structural form known nowadays as the speech variety of the Gabonese youth and hip-pop singers. Toli-bangando indeed draws an important deal of lexicon from the Gabonese popular French.
4. The first two doctoral theses on metalexicographical planning of dictionaries for Gabonese languages are Mavoungou (2002b) and Afane Otsaga (2004). They were followed by eight other doctoral theses on various metalexicographical issues for Gabonese languages (cf. Mabika Mbokou 2006, Nyangone Assam 2006, Mihindou 2006, Ella 2007, Ekwa Ebanega 2007, Tomba Moussavou 2007, Saphou-Bivigat 2010 and Soami 2010).

5. Our translation from *Centre de Recherche en Etudes Germaniques et Interculturelles*.
6. Our translation from *l'épistémologie des missionnaires* (Ndinga-Koumba-Binza 2006a: 298).

Acknowledgments

The second author on this article is a holder of the NRF Research Career Advancement Fellowship. He is grateful to the **National Research Foundation** (NRF) in South Africa for this fellowship and financial support. The authors also wish to thank the anonymous reviewers and the audience of 2015 LSSA/SAALA/SAALT joint conference held in Potchefstroom, South Africa (24-26 June 2015) for insightful comments and questions. The authors alone are responsible for any particular view or shortcoming this article may contain.

References

Dictionaries and Encyclopedias

- 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.
- Crystal, D.** 2008. *A Dictionary of Linguistics and Phonetics*. Sixth edition. Oxford: Blackwell Publishing.
- DIT2009** = Ditougou, L. 2009. *On est ensemble: 852 mots pour comprendre le français du Gabon*. Libreville: Editions Raponda-Walker.
- DB2008** = Dodo-Bounguendza, E. 2008. *Dictionnaire des gabonismes*. Paris: L'Harmattan.
- DB2010** = Dodo-Bounguendza, E. 2010. *Diagnostic du français du Gabon. Guide pratique destiné aux journalistes, politiques, administratifs et universitaires*. Libreville: Les Editions Ntsame.
- DB2013** = Dodo-Bounguendza, E. 2013. *Dictionnaire du parler toli-bangando. Argot des jeunes gabonais*. Libreville: Les Editions Ntsame.
- 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/New York: Walter de Gruyter.
- Mavoungou, P.A. and B. Plumel.** 2010. *Dictionnaire yilumbu-français*. Libreville: Editions Raponda-Walker.
- Mavoungou, P.A., Moussounda Ibouanga, F. 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.
- Molamu, L.** 2003. *Tsotsitaal: A Dictionary of the Language of Sophiatown*. Pretoria: UNISA Press.
- Moussounda Ibouanga, F.** 2011. *Français du Gabon: approches sociolinguistiques et lexicographiques (le toli bangando)*. Paris: Editions Universitaires Européennes.
- Rézeau, P. (Ed.)**. 1999. *Variétés géographiques du français de France aujourd'hui: approche lexicographique*. Paris/Brussels: De Boeck and Larcier-Duculot.
- Richards, J.C. and R. Schmidt.** 2002. *Longman Dictionary of Language Teaching and Applied Linguistics*. Third Revised Edition. London/New York: Pearson Education.

Other Literature

- Afane Otsaga, T.** 2004. *The Standard Translation Dictionary as an Instrument in the Standardisation of Fang*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Afane Otsaga, T.** 2013. Lexicography of Fang. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 962-963.
- Alberts, M.** 1996. *Background Information on the Drafting of the National Lexicography Units Bill*. Report. Pretoria: DACST.
- Alves, M.** 1994. *La presse gabonaise en langue française: étude lexicale*. M.A. thesis. Paris: Université de Paris III-Sorbonne Nouvelle.
- Artigues, M.** 1995. *Participation à une étude des particularités lexicales du français parlé au Gabon*. M.A. thesis. Paris: Université de Paris III-Sorbonne Nouvelle.
- Bagouendi-Bagère, D.** 1999. *Etude des procédés de néologie lexicale dans le français parlé au Gabon*. DEA thesis. Aix-en-Provence: Université de Provence-Aix-Marseille I.
- Bagouendi-Bagère Bonnot, D.** 2007. *Le français au Gabon: représentations et usages*. Unpublished Ph.D. thesis. Aix-en-Provence: Université de Provence.
- Bergenholtz, H.J. and R.H. Gouws.** 2012. What is Lexicography? *Lexikos* 22: 31-42.
- Blanchon, J.A.** 1994. Gabon: Language Situation. Asher, R.E. (Ed.). 1994. *The Encyclopedia of Language and Linguistics*. Volume 3: 1341-1343. Oxford: Pergamon.
- Boucher, K.** 1997. *Créativité lexicale et identité culturelle du français au Gabon*. M.A. thesis. Paris: Université de Paris III-Sorbonne Nouvelle.
- Boucher, K.** 1998. *Langues et identité culturelle des jeunes Librevillois de 15 à 30 ans: une enquête de terrain*. DEA thesis. Paris: Université de Paris III-Sorbonne Nouvelle.
- Boucher, K.** 1999. Approche des représentations sociolinguistiques dans un groupe de jeunes Librevillois. *Le Français en Afrique* 13: 173-192.
- Boulard, G.** 1999. L'ordonnance de Villers-Cotterêts: le temps de la clarté et la stratégie du temps (1539-1992). *Revue Historique* 301(1): 45-100.
- Boutin, B.A.** 2003. La norme endogène du français de Côte d'Ivoire: mise en évidence de règles différentes du français de France concernant la complémentation verbale. *SudLangues. Revue électronique internationale de Sciences du langage* 2: 33-46.
- Boutin-Dousset, C.** 1989. *Matériaux pour un inventaire des particularités lexicales du français au Gabon*. M.A. thesis. Paris: Université de la Sorbonne Nouvelle.
- Chabata, E.** 2013. Lexicography of Shona. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 947-952.
- Chang, L.** 2014. *Biesik Jumiekan: Introduction to Jamaican Language*. Washington, DC: Chuu Wod.
- Cissé, M.** 2005. Langues, Etat et société au Sénégal. *SudLangues. Revue électronique internationale de Sciences du langage* 5: 99-133.
- Cohen, P.** 2003. L'imaginaire d'une langue nationale: l'Etat, les langues et l'invention du mythe de l'ordonnance de Villers-Cotterêts à l'époque moderne en France. *Histoire Épistémologie Langage* 25(1): 19-69.
- Couvert, C.** 1982. *La langue française en République gabonaise*. Paris: Haut Comité de la Langue Française.

- Dodo-Bounguendza, E.** 1999. Politique et organisation linguistiques pour l'éducation de base au Gabon. Chaudenson, R. and R. Renard (Eds). 1999. *Langues et développement*: 181-192. Paris: Didier Érudition.
- Dodo-Bounguendza, E.** 2008. *Des usages démocratiques des langues du Gabon*. Paris: L'Harmattan.
- Ekwa Ebanéga, G-M.** 2007. *A Microstructural Programme for Dictionaries in Fang*. Unpublished D. Litt. dissertation. Stellenbosch: Stellenbosch University.
- Ella, E.M.** 2007. *A Theoretical Model for a Fang–French–English Specialized Multi-volume School Dictionary*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Emejulu, J.D.** 2000. Lexicography, an Economic Asset in Multilingual Gabon. *Revue Gabonaise des Sciences du Langage / Gabonese Journal of Language Sciences* 1: 51-69.
- Emejulu, J.D.** 2001. Lexicographie multilingue et multisectorielle au Gabon: planification, stratégie et enjeux. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome 1*: 38-57. New York: Jimacs-Hillman Publishers.
- Emejulu, J.D.** 2002. Défis et promesses de la lexicographie intégrale dans les pays en développement. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome 2*: 366-381. New York: Jimacs-Hillman Publishers.
- Emejulu, J.D.** 2003. Challenges and Promises of a Comprehensive Lexicography in the Developing World: The Case of Gabon. Botha, W.F. (Ed.). *'n Man wat beur. Huldingsbundel vir Dirk van Schalkwyk*: 195-212. Stellenbosch: Bureau of the WAT.
- Engoung Nsi, S.P.** 2013. *Analyse comparative du discours de Omar Bongo et Ali Bongo: continuité et/ou rupture*. M.A. thesis. Besançon: Université de Franche-Comté.
- Fumaroli, M.** 1992. Le génie de la langue française. Nora, P. (Ed.). 1992. *Les lieux de mémoire. Volume 3: Les France. Book 3: De l'archive à l'emblème*: 911-973. Paris: Gallimard.
- Garcia, L.** 1971. L'organisation de l'instruction publique au Dahomey, 1894–1920. *Cahier d'études africaines* 11(41): 59-100.
- Gauton, R.** 2013. Lexicography of the Nguni Languages. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 911-929.
- Gouws, R.H.** 2013. Aspects of Afrikaans Lexicography. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 827-835.
- 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/New York: Walter de Gruyter.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PRESS.
- Idiata, D.F.** 2002. *Il était une fois les langues gabonaises*. Libreville: Éditions Raponda-Walker.
- Idiata, D.F.** 2008. Le français et les langues gabonaises, du partenariat au linguicide: une analyse des données des enfants tirées du contexte de la ville de Libreville. *Revue Gabonaise des Sciences du Langage* 3: 85-208.
- Idiata, D.F.** 2009. *Langues en danger et langues en voie d'extinction au Gabon*. Paris: L'Harmattan.
- Irvine, G.A.** 2004. A Good Command of the English Language: Phonological Variation in the Jamaican Acrolect. *Journal of Pidgin and Creole Languages* 19(1): 41-76.
- Italia, M.** 2000. *Morphosyntaxe verbale dans des corpus de locuteurs basilectaux au Gabon*. M.A. thesis. Aix-en-Provence: Université de Provence-Aix-Marseille I.

- Italia, M.** 2006. Le morphème *là* dans les variétés mésolectales et basilectales en français du Gabon. *Le Français en Afrique* 21: 281-290.
- Italia, M.** 2011. *Variation et variétés morphosyntaxiques du français parlé au Gabon*. Ph.D. thesis. Aix-en-Provence: Université Aix-Marseille.
- Italia, M. and A. Queffélec.** 2011. Gestion du verbe français chez des locuteurs gabonais analphabètes, peu et moyennant scolarisés. *Le Français en Afrique* 26: 85-96.
- Itembo, G.** 1999. *Normes endogènes du français, problèmes identitaires et institution scolaire à Libreville*. M.A. thesis. Libreville: University of Omar Bongo.
- Judge, A.** 1993a. Linguistic Legislation and Practice. Sampson, R. (Ed.). 1993. *Authority and the French Language. Papers from a Conference at the University of Bristol*: 63-73. Münster: Nodus Publikationen.
- Judge, A.** 1993b. French: A Planned Language? Sanders, C. (Ed.). 1993: *French Today, Language in its Social Context*: 7-26. Cambridge: Cambridge University Press.
- Klinkenberg, J.-M.** 2002. La légitimation de la variation linguistique. *L'Information grammaticale* 94: 22-26.
- Kwenzi Mikala, J.T.** 1990. Quel avenir pour les langues gabonaises? *Revue Gabonaise des Sciences de l'Homme* 2: 121-124.
- Kwenzi Mikala, J.T.** 1998. Parlers du Gabon: classification du 11.12.97. Raponda-Walker, A. (Ed.). 1998. *Les langues du Gabon*: 217-221. Libreville: Éditions Raponda Walker.
- Lafage, S.** 1993. French in Africa. Sanders, C. (Ed.). 1993: *French Today, Language in its Social Context*: 215-238. Cambridge: Cambridge University Press.
- Mabika Mbokou, L.** 2006. *A Model for the Macro- and Microstructure of a Yipunu–French School Dictionary*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Mabika Mbokou, L.** 2008. Le français langue maternelle! *CENAREST Infos* 4: 4.
- Mabika Mbokou, L.** 2012. A Survey of 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.
- Makanga Mboumba, J.** 2005. *Evaluation des compétences langagières en français, en milieu scolaire: cas des élèves de troisième générale à Port-Gentil*. DEA thesis. Strasbourg: Université Marc Bloch.
- Makanga Mboumba, J.** 2007. La langue française dans la cellule familiale mixte au Gabon. *Éducation et Sociétés Plurilingues* 22: 37-50.
- Mavoungou, P.A.** 2001. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise Tome I*: 160-183. New York: Jimacs-Hillman Publishers.
- Mavoungou, P.A.** 2002a. Vers un dictionnaire du français du Gabon. Emejulu, J.D. (Ed.). 2002: *Éléments de lexicographie gabonaise. Tome II*: 230-262. New York: Jimacs-Hillman Publishers.
- Mavoungou, P.A.** 2002b. *Metalexicographical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu–English–French*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Mavoungou, P.A.** 2010. *Lexicographie et confection des dictionnaires au Gabon*. Stellenbosch: SUN PRESS.
- Mavoungou, P.A.** 2011. Regard sur les onomastismes dans le français de Libreville et leur traitement lexicographique. *Mbaandza. Revue d'étude et d'analyse francophones* 1: 21-50.
- Mavoungou, P.A.** 2013a. Gabonese French Dictionaries: Survey and Perspectives. *Lexikos* 23: 255-272.
- Mavoungou, P.A.** 2013b. Lexicography of the Languages of Central Africa. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 953-962.

- Mavoungou, P.A.** 2013c. Lexicography of the Languages of Western Africa. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 964-968.
- Mbondzi, J.Y.** 1998. *La syntaxe du français des élèves du cours moyen de Libreville: cas de l'école publique Martine Oulabou*. M.A. thesis. Libreville: Omar Bongo University.
- Mebiame-Akono, P.** 2008a. Stratégies énonciatives et schématisation de soi dans une production discursive d'Omar Bongo Ondimba: analyse pragmatique. *Gabonica* 2(2): 148-167.
- Mebiame-Akono, P.** 2008b. Quelques aspects relationnels d'une production discursive d'Omar Bongo Ondimba: analyse pragmatique. *Itinérés. Itinéraires Pédagogiques et Médiations Scientifiques et Technologiques* 6(6): 33-48. Libreville: Editions du CENAREST.
- Mebiame-Akono, P.** 2009. Stratégies d'opacification énonciative lors d'une oraison funèbre: l'exemple d'une production discursive de Jean Eyeghe Ndong. Ondo-Mebiame, P. (Ed.). 2009. *Quel avenir pour les langues et culture du Gabon?*: 60-83. Libreville: Editions CUI-Gabon.
- Mebiame-Akono, P.** 2011. Procédés argumentatifs d'un discours politique: l'exemple d'une allocution d'Ali Bongo Ondimba. *Mbaandza 1. Revue d'étude et d'analyse francophones*: 169-200.
- Mendo Ze, G.** 1999. *Le français langue africaine: enjeux et atouts pour la francophonie*. Paris: Publisud.
- Mesthrie, R., J. Swann, A. Deumert and W.L. Leap.** 2009. *Introducing Sociolinguistics*. Second Edition. Edinburgh: Edinburgh University Press.
- Mihindou, G.-R.** 2001. Apports des missionnaires à la lexicographie gabonaise: Dictionnaires bilingues fang-français/français-fang; français-yipounou/yipounou-français; français-mpongwé. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 7-37. New York: Jimacs-Hillman Publishers.
- Mihindou, G.-R.** 2006. *A Theoretical Model for a Yipunu-English-French Explanatory Dictionary of Medical Terms*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Mindze M'Eyeghe, J.** 2001. *Approches de quelques particularités lexicales du français parlé au Gabon*. M.A. thesis. Libreville: Omar Bongo University.
- Minko, D.** 2008. Le marquage identitaire dans le français gabonais. *Synesgies Monde* 5: 159-164.
- Mitchell, R.** 2004. Les perceptions du français gabonais et la distribution des langues au Gabon. *Le Français en Afrique* 19: 177-188.
- Mitchell, R.** 2009. *A Study in Syntactic Variation: Mood Usage in Gabonese French*. Munich: Lincom Europa.
- Mouélé, M.** 2011. Les Racines bantu du français gabonais. *Mbaandza 1. Revue d'étude et d'analyse francophones*: 87-111.
- Mouloungui Nguimby, F.V.** 2002. *De la variation dialectale en français au Gabon*. Unpublished M.A. thesis. Libreville: Omar Bongo University.
- Moussirou Mouyama, A.** 1984. *La langue française au Gabon: Contribution sociolinguistique*. Unpublished Ph.D. dissertation Paris: Université René Descartes (Paris V).
- Moussirou Mouyama, A.** 1986. L'introduction de la langue française au Gabon. *Annales de l'Université Omar Bongo du Gabon: Série Lettres, Droit, Sciences et Médecine*: 55-63.
- Moussirou Mouyama, A.** 1998. Norme officielle du français et normes endogènes au Gabon. Calvet, L.-J. and M.-L. Moreau (Eds.). 1998. *Une ou des normes? Insécurité linguistique et normes endogènes en Afrique francophone*: 83-91. Paris: Didier Erudition.
- Moussounda Ibouanga, F.** 2008a. Pour une lecture des particularismes gabonais: cas de la ville de Mouila, faits de langues et identitaires. *Le français en Afrique. Revue du Réseau des Observatoires du Français Contemporain en Afrique* 23: 113-133.

- Moussounda Ibouanga, F.** 2008b. Le toli bangando: la variabilité de réalisation à Libreville et ses conséquences sur le français (du Gabon). *Revue Gabonaise des Sciences du Langage* 4: 135-152.
- Moyo, T.** 2002. Mother Tongues versus an Ex-colonial Language as Media of Instruction and the Promotion of Multilingualism: The South African Experience. *South African Journal of African Languages* 22(2): 149-160.
- Ndao, P.A.** 2002. Le français au Sénégal: une approche polynomique. *SudLangues. Revue électronique internationale de Sciences du langage* 1: 51-64.
- Ndenguino Mpira, H.** 2005. *La refondation: une analyse du discours*. M.A. thesis. Libreville: Omar Bongo University.
- Ndinga-Koumba-Binza, H.S.** 2004. *Le statut socio-politique du français et la promotion des langues locales au Gabon*. Paper presented at the 18th International Conference of the Association of French Studies in Southern Africa (AFSSA) held by the French Section of the Department of Modern Foreign Languages at the University of Stellenbosch. Stellenbosch, 8–11 September 2004.
- Ndinga-Koumba-Binza, H.S.** 2005a. Politique linguistique et éducation au Gabon: un état des lieux. *Journal of Education* 4(1): 65-78. Réduit: Mauritius Institute of Education.
- Ndinga-Koumba-Binza, H.S.** 2005b. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ndinga-Koumba-Binza, H.S.** 2006a. *Lexique Pove–Français/Français–Pove*, Mickala Manfoumbi: Seconde Note de Lecture. *Lexikos* 16: 293-308.
- Ndinga-Koumba-Binza, H.S.** 2006b. Demographic Profiles of Libreville and Lomé. Bekker, Simon and Anne Leildé (Eds.). 2006. *Reflections on Identity in Four African Cities*: 45-50. [South Africa]: African Minds.
- 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.** 2010. *Trends in Gabonese Modern Lexicography*. Paper presented at 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.** 2011. From Foreign to National: A Review of the Status of the French Language in Gabon. *Literator* 32(2): 135-150.
- Ndong Mbeng, H.F.** 1992. *Les matitis*. Saint-Maur: Sépia.
- Nsa Ndo, E.K.** 2010. *Les locutions figurées dans le français des jeunes librevillois de 15 à 30 ans: approche linguistique et sociolinguistique*. M.A. thesis. Libreville: Omar Bongo University.
- Nsafou, V.** 2010. *Les collocations dans le français des jeunes librevillois de 15 à 30 ans: approche linguistique et lexicologique*. M.A. thesis. Libreville: Omar Bongo University.
- Ntsaga-Oyouni, S.** 1998. *Approche de quelques traits du français parlé de jeunes Gabonais habitant la région parisienne*. DEA thesis. Paris: Université de Paris III-Sorbonne Nouvelle.
- Ntshangase, D.K.** 2002. Language and Language Practices in Soweto. Mesthrie, R. (Ed.). 2002. *Language in South Africa*: 407-418. Cambridge: Cambridge University Press.
- Nyangone Assam, B.** 2006. *Dictionaries as Teaching Instruments for Mother-tongue Education: The case of Fang in Gabon*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Nyangone Assam, B. and P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Ogden, J.** 1984. French in Gabon. *Contemporary French Civilization* 8: 339-348.
- Ompoussa, V.** 1998. *Problèmes de maîtrise de ponctuation chez les élèves du cours moyen élémentaire à Libreville: cas de l'Ecole d'Application de l'ENS/A*. M.A. thesis. Libreville: Omar Bongo University.

- 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* 1. *Revue d'étude et d'analyse francophones*: 75-86.
- Ondo-Mebiame, P.** 2008. La formation des néologismes en français gabonais: analyse d'un corpus tiré du journal *Le Scribouillard*. *Revue Gabonaises des Sciences du Langage* 3: 151-200.
- Ondo-Mebiame, P. and G.M. Ekwa Ebanéga.** 2011. Regard critique sur *On Est Ensemble: 852 mots pour comprendre le français du Gabon*. *Lexikos* 21: 337-358.
- Pambou, J.-A.** 1998. Le français au Gabon: une langue à multiples statuts. *iBoogha* 2: 127-149.
- Pambou, J.-A.** 2011a. Les hypercorrections dans le français pratiqué au Gabon. *Mbaandza* 1. *Revue d'étude et d'analyse francophones*: 127-167.
- Pambou, J.-A.** 2011b. *Les constructions prépositionnelles chez les apprenants de français langue seconde au Gabon: étude didactique*. Munich: Lincom Europa.
- Patrick, P.L.** 1999. *Urban Jamaican Creole: Variation in the Mesolect*. Amsterdam/Philadelphia: John Benjamins.
- Ploog, K.** 2000. La norme dans l'observation des normes abidjanaises: étude d'un continuum linguistique. *Lengas* 48: 103-128.
- Ploog, K.** 2001. Le non-standard entre norme endogène et fantasme d'unicité. L'épopée abidjanaise et sa polémique intrinsèque. *Cahiers d'études africaines* 163-164: 423-442.
- Prinsloo, D.J.** 2013. Lexicography of the Sotho Languages. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013: 929-947.
- Sand, A.** 1999. *Linguistic Variation in Jamaica: A Corpus-based Study of Radio and Newspaper Usage*. Tübingen: Gunter Narr Verlag.
- Saphou-Bivigat, G.** 2010. *A Theoretical Model for an Encyclopaedic Dictionary for the Gabonese Languages with Reference to Yilumbu*. Unpublished D.Litt. dissertation. Stellenbosch: Stellenbosch University.
- Satyo, S.C.** 2001. Kwaito-speak: A Language Variety Created for the Youth by the Youth. Ridge, E., S. Makoni and S. Ridge (Eds.). 2001. *Freedom and Discipline: Essays in Applied Linguistics from Southern Africa*: 139-148. New Delhi: Bahri Publications.
- Slabbert, S. and C. Myers-Scotton.** 1996. The Structure of Tsotsitaal and Isicamtho: Code-switching and In-group Identity in South African Townships. *Linguistics* 34: 317-342.
- Soami, L.S.** 2010. *Towards the Development and Application of Representative Lexicographic Corpora for the Gabonese Languages*. Ph.D. thesis. Stellenbosch: Stellenbosch University.
- Spencer, J.** 1971. Colonial Language Policies and Their Legacies. Sebeok, T.A. (Ed.). 1971. *Linguistics in Sub-Saharan Africa*. *Current Trends in Linguistics* 7: 559-569. The Hague/Paris: Mouton.
- Stockwell, P.** 2007. *Sociolinguistics. A Resource Book for Students*. Second Edition. London/New York: Routledge.
- Tomba Moussavou, F.** 2001. *Les réalisations des séquences de consonnes en français de Libreville*. Unpublished M.A. thesis. Libreville: Omar Bongo University.
- Tomba Moussavou, F.** 2007. *Metalexigraphic Criteria for a Monolingual Descriptive Dictionary Presenting the Standard Variety of Yipunu*. Ph.D. thesis. Stellenbosch: Stellenbosch University.
- Trudeau, D.** 1983. L'ordonnance de Villers-Cotterêts et la langue française: histoire et interprétation. *Bibliothèque d'Humanisme et Renaissance* 45(3): 461-472.
- Van den Avenne, C.** 2001. Un récit urbain plurilingue. "Les Matitis" d'Hubert Freddy Ndong Mbeng. *Plurilinguismes* 18: 49-66.

- Van den Avenne, C.** 2004. La position énonciative complexe d'un écrivain d'Afrique francophone: le cas d'Hubert Freddy Ndong Mbeng. *Glottopol. Revue de Sociolinguistique en Ligne* 3: 44-59. <http://www.univ-rouen.fr/dyalang/glottopl> (accessed on 28 August 2015).
- Van der Veen, L.** 2006. Gabon: Language Situation. Brown, K. (Ed.). 2006. *The Encyclopedia of Language and Linguistics. Volume 4*: 708-715. Second Edition. Amsterdam/Boston: Elsevier.
- Wakely, R.** 2000. Des yeux étrangers vous regardent: l'enseignement du français vu de l'extérieur. Vigner, G. (Ed.). 2000. *Documents pour l'histoire du français langue étrangère ou seconde*. Tome 25: *L'enseignement et la diffusion du français dans l'empire colonial français. 1815-1962*: 11-33. Paris: SIHFLES (Société internationale pour l'histoire du français langue étrangère ou seconde).
- Wamba, R.S. and G.M. Noumssi.** 2003. Le français au Cameroun contemporain: statuts, pratiques et problèmes sociolinguistiques. *SudLangues. Revue électronique internationale de Sciences du langage* 2: 1-20.
- Zang Zang, P.** 1998. *Le français en Afrique: normes, tendances évolutives, dialectalisation*. Munich: Lincom Europa.
- Zang Zang, P.** 2013. Codification et normalisation du français d'Afrique: enjeux et perspectives. *SudLangues. Revue électronique internationale de Sciences du langage* 19: 1-20.

Die rol van die US se Departement Afrikaans en Nederlands in die ontwikkeling van die Afrikaanse leksikografie onder Nederlandse invloed: Verklarende standaardwoordeboeke

Gerda Odendaal, *Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Suid-Afrika (godendaal@sun.ac.za)*

Opsomming: Hierdie artikel stel ondersoek in na die rol van die Universiteit Stellenbosch se Departement Afrikaans en Nederlands in die ontwikkeling van die Afrikaanse leksikografie deur hulle betrokkenheid by twee verklarende standaardwoordeboeke, te wete die *Verklarende Handwoordeboek van die Afrikaanse Taal* en *Nasionale Woordeboek*. Dit word gedoen teen die agtergrond van die invloed van Nederlands op die Afrikaanse leksikografie. Daar word veral gedui op die invloed van Nederlands op dié woordeboeke op sowel makro- as mikrostrukturele vlak. Daar word egter ook gedui op die geleidelike losbreek van die Nederlandse invloed en die bydrae van hierdie woordeboeke tot die ontwikkeling van die Afrikaanse leksikografie. Wanneer die ooreenstemmende uitgawes van dié woordeboeke met mekaar vergelyk word, word daar tot die gevolgtrekking gekom dat *Nasionale Woordeboek* minder Nederlandse invloed as die *Verklarende Handwoordeboek van die Afrikaanse Taal* blyk te toon. Die *Verklarende Handwoordeboek van die Afrikaanse Taal* het egter veral sedert die begin van die negentigerjare (nadat daar nie verdere uitgawes van *Nasionale Woordeboek* verskyn het nie) stadig maar seker begin om van Nederlandse invloed weg te breek, sodat die jongste uitgawe van dié woordeboek blyke gee van Afrikaans as 'n selfstandige taal wat nie langer op Nederlands hoef te steun nie.

Sleutelwoorde: UNIVERSITEIT STELLENBOSCH, DEPARTEMENT AFRIKAANS EN NEDERLANDS, AFRIKAANSE LEKSIKOGRAFIE, NEDERLANDSE INVLOED, NEERLANDISME, VERKLARENDE STANDAARDWOORDEBOEK, VERKLARENDE HANDWOORDEBOEK VAN DIE AFRIKAANSE TAAL, KERNWOORDEBOEK, NASIONALE WOORDEBOEK

Abstract: The Role of Stellenbosch University's Department of Afrikaans and Dutch in the Development of Afrikaans Lexicography: Development under Dutch Influence. This article investigates the role of Stellenbosch University's Department of Afrikaans and Dutch in the development of Afrikaans lexicography through their involvement in two standard descriptive dictionaries, the *Verklarende Handwoordeboek van die Afrikaanse Taal* en *Nasionale Woordeboek*. This is done against the background of the influence of Dutch on Afrikaans lexicography. The influence of Dutch on both the macro- and microstructures of these

dictionaries are investigated. A gradual movement away from Dutch influence and the contribution of these dictionaries to the development of Afrikaans lexicography are also indicated. When corresponding editions of the two dictionaries are compared, *Nasionale Woordeboek* seems to show less Dutch influence than *Verklarende Handwoordeboek van die Afrikaanse Taal*. However, since the beginning of the 1990s (after no new editions *Nasionale Woordeboek* appeared) *Verklarende Handwoordeboek van die Afrikaanse Taal* has slowly but surely started to free itself from Dutch influence. The result of this is that the most recent edition of the *Verklarende Handwoordeboek van die Afrikaanse Taal* gives evidence of Afrikaans as an independent language that no longer needs to rely on Dutch.

Keywords: STELLENBOSCH UNIVERSITY, DEPARTMENT OF AFRIKAANS AND DUTCH, AFRIKAANS LEXICOGRAPHY, DUTCH INFLUENCE, DUTCHISM, STANDARD DESCRIPTIVE DICTIONARY, VERKLARENDE HANDWOORDEBOEK VAN DIE AFRIKAANSE TAAL, KERNWOORDEBOEK, NASIONALE WOORDEBOEK

1. Inleiding

Die Universiteit Stellenbosch het in die twintigste eeu 'n belangrike bydrae gelewer tot die totstandkoming van die Afrikaanse leksikografie met die verskyning van die Nederlandse dosent, Nicolaas Mansvelt, se *Proeve van een Kaapsch-Hollandsch Idioticon* deurdat dit beskou is as die eerste selfstandige Afrikaanse woordeboek wat primêr op Afrikaanssprekendes gerig is en wat gedui het op die eiesoortigheid van Afrikaans as onafhanklike taal naas Nederlands (Gouws 1995: 48; Schoonheim 1998: 43; Gouws 2003: 79). 'n Nederlands-Afrikaans-stryd tussen die voorstanders van Nederlands, wat onder leiding van die Stellenbosse dosent, W.J. Viljoen, gepoog het om 'n vereenvoudigde Nederlandse spelling deur die publikasie van verskeie woordeboeke in te voer, en die voorstanders van Afrikaans, wat hulle onder die leierskap van nóg 'n Stellenbosser, J.J. Smith, beywer het om Afrikaans as onderrigmedium aan die Universiteit Stellenbosch erken te kry, het egter aan die begin van die twintigste eeu binne die Departement Afrikaans en Nederlands, maar ook in die breër Afrikaanse spraakgebied, posgevat. Dit het uiteindelik daartoe aanleiding gegee dat daar in ortografiese bereëling van Afrikaans teruggegryp is na Nederlands ten einde nie die voorstanders van Nederlands van die Afrikaanse ortografie te vervreem nie (vgl. Odendaal 2016). Die teruggryp na Nederlands sou uiteindelik ook neerslag vind in die leksikografiese vergestaltung van Afrikaans.

Een van die woordeboeke waarin die Nederlands-Afrikaans-stryd neerslag gevind het, was die *Woordeboek van die Afrikaanse Taal* (vgl. Odendaal, om te verskyn). Dié woordeboek, wat aanvanklik veronderstel was om 'n verklarende standaardwoordeboek te wees, het mettertyd tot 'n omvattende verklarende woordeboek ontwikkel, aangesien die eerste hoofredakteur, J.J. Smith, die "versameling van die Afrikaanse woordeskat" as "'n volksaak" en "die voltooiing van die eerste min of meer volledige Afrikaanse woordeboek" as "'n volksgebeurtenis" beskou het (Snijman 1964: 11; 1975: 3). Die gevolg was dat daar 'n merkbare leemte in die Afrikaanse leksikografie ontstaan het, naamlik 'n

gebrek aan 'n gesaghebbende verklarende standaardwoordeboek vir Afrikaans (Gouws en Ponelis 1992: 19). Alhoewel daar in 1936 'n deurbraak gemaak is met die verskyning van die *Verklarende Afrikaanse Woordeboek*,¹ sou die belangrike bydraes tot die standaard verklarende leksikografie eers later hulle verskyning maak, onder andere *Verklarende Handwoordeboek van die Afrikaanse Taal* (1965) (Gouws en Ponelis 1992: 20). Waar verklarende standaardwoordeboeke gewoonlik 'n belangrike standaardiseringsrol vervul, kon hierdie woordeboeke weens hulle vertraagde verskyning nie vir hierdie doel in Afrikaans aangewend word nie (Gouws en Ponelis 1992: 19). Hierdie vertraging het egter volgens Gouws en Ponelis (1992: 19-20) die gevolg gehad dat dit die kwaliteit van die verklarende standaardwoordeboeke verhoog het wat uiteindelik verskyn het, aangesien hierdie woordeboeke kon put uit 'n toenemende bewustheid van die studie van die Afrikaanse taalkunde. Ten spyte hiervan het ook die standaard verklarende woordeboeke egter nie die Nederlandse invloed vrygespring nie.

In hierdie derde artikel² oor die rol van die Universiteit Stellenbosch (voortaan US), maar in die besonder sy Departement Afrikaans en Nederlands, in die ontwikkeling van die Afrikaanse leksikografie onder Nederlandse invloed, word daar veral op verklarende standaardwoordeboeke gefokus. Twee woordeboeke waarin verskeie dosente van die Departement Afrikaans en Nederlands deel gehad het, word in hierdie artikel onder die loep geneem, te wete die *Verklarende Handwoordeboek van die Afrikaanse Taal* en *Nasionale Woordeboek*. In die bespreking van hierdie woordeboeke word daar nie soseer op dié woordeboeke as gebruikersvriendelike naslaanbronne gefokus nie, maar eerder op die invloed van Nederlands op die woordeboeke, die wegbreek van Nederlandse invloed, asook hulle bydrae tot die Afrikaanse standaardleksikografie.

2. Verklarende Handwoordeboek van die Afrikaanse Taal

2.1 HAT 1

Teen die draai van die 1950's besluit die Raad van Beheer oor Die Afrikaanse Woordeboek om gelyklopend met die WAT ook aan 'n minder omslagtige verklarende woordeboek te werk (Odendal 2006: 281). In 1965 verskyn die eerste uitgawe van die *Verklarende Handwoordeboek van die Afrikaanse Taal* (voortaan HAT) onder redakteurskap van P.C. Schoonees (voormalige hoofredakteur van die WAT), C.J. Swanepoel (wat aan Stellenbosch gegradeer het) en S.J. du Toit (beide voormalige WAT-redaksielede), en C.M. Booysen (Schoonees, Swanepoel, Du Toit en Booysen 1965: 'n Woord vooraf). Hierdie woordeboek, wat "op dieselfde grondslag opgestel [is] as *Die Woordeboek van die Afrikaanse Taal* (WAT)", is teen 1 067 bladsye "heelwat kleiner in omvang en beperkter van opset as WAT", aldus Schoonees et al. (loc.cit.). Alhoewel die US se Departement Afrikaans en Nederlands waarskynlik nie met hierdie uitgawe 'n direkte verbintenis met HAT gehad het nie, het die redaksie wel "professore van die Universiteit van

Stellenbosch" in die voorwoord uitgesonder as medewerkers "wat hulp verleen het by die vasstelling van definisies" (Schoonees et al. loc.cit.). *HAT* se band met die *WAT* impliseer ook 'n redelike mate van samewerking tussen die *HAT* en die *US* (vgl. Odendaal, om te verskyn). Dié samewerking sou egter later meer prominent raak met R.H. Gouws se toetrede tot die redaksie van *HAT* (sien 2.4).

Die siening dat Nederlands "die onmisbare bron [is] waaruit Afrikaans moet put, oral waar voorlopig nie deur eie taalmiddels in die behoeftes kan voorsien word nie" wat aan die begin van die twintigste eeu in Afrikaanse taalegeledere posgevat het (Le Roux, Malherbe en Smith 1917: iv; vgl. Odendaal 2016: 265), het na die leksikografie oorgevloei. Ponelis (1992: 81) dui daarop dat die Afrikaanse leksikografie "ontwikkel [het] uit die vertaling van Nederlandse woordeboeke". Dat ook *HAT* 1 se Nederlandse onderrok plek-plek uithang, aldus Van Houwelingen en Carstens (1998: 3), blyk daaruit dat *HAT* "sowel mikrostruktureel as makrostruktureel sterk by Nederlands aanleun" (Van Houwelingen en Carstens 1998: 2). Vergelyk in hierdie verband die volgende lys Neerlandistiese³ lemmas of lemmas met 'n Neerlandistiese betekenis wat in *HAT* 1 opgeneem is met die frekwenter Afrikaanse vorme daarnaas:⁴

abonneer ("inteken", bv. op 'n koerant), *abonnement* ("intekening/subskripsie"), *agternaam* ("van"), *benieu* ("nuuskierig maak"), *benieud* ("nuuskierig"), *bigotterie* ("skynheiligheid"), *bil* ("boud"), *bloesem* ("bloeisel"), *blootshoof(s)* ("kaalkop"), *blootsvoet(s)* ("kaalvoet"), *briewebesteller* ("posbode/-man"), *briewebus* ("posbus"), *felisiteer* ("gelukwens"), *felisitasie* ("gelukwensing"), *fiets* (ww.) ("fietsry"), *haag* ("heining"), *hals* ("nek"), *heg* ("heining"), *houthakker* ("houtkapper"), *internaat* ("kosskool"), *invasie* ("inval"), *invektief* ("skel(d)woord"), *invitasie* ("uitnodiging"), *jarig* ("verjaar"), *kameleon* ("verkleurmannelietjie"), *kapper* ("haarkapper"), *kloset* ("toilet"), *kneding* (handeling van "knie"), *knee(d)baar* ("kniebaar"), *kus* (s.nw./ww.) ("soen"), *moedervlek* ("geboortemerke/-vlek"), *onbesuis* ("roekeloos/onstuimig/wild"), *origineel* ("oorspronklik"), *pardon* (tw.) ("ekskuus"), *pinda* ("grondboontjie"), *serius* ("ernstig"), *sielig* ("droewig"), *skavuit* ("skelm/skurk"), *spaarpot* ("spaarbussie"), *stopkontak* ("muurprop/kragpunt"), *swyneboel* ("vuil huishouding"), *tarwe* ("koring"), *toonladder* ("toonleer"), *twyg* ("dun takkie"), *visite* (s.nw.) ("besoek")

Behalwe *kus* (s.nw.), *skavuit* en *tarwe*, is al die bogenoemde lemmas ongeëtiketteerd opgeneem. Dit impliseer volgens Gouws (1989: 147, 201) dat hierdie lemmas (wat ongeëtiketteerd opgeneem is) korrekte en aanvaarde vorme in Standaardafrikaans is en skep sodoende die indruk by woordeboekgebruikers dat daar geen beperking op die gebruik van hierdie lemmas is nie. *Kus*, *skavuit* en *tarwe* is onderskeidelik met die stilistiese etikette *w.g.* (weinig gebruiklik) en *verh.* (verhewe) gemerk. Eersgenoemde, wat by *skavuit* en *tarwe* voorkom, dui daarop dat "'n woord wat wel dwarsoor die taalgebied bekend is" maar dat "die frekwensie om een of ander rede laag is", terwyl laasgenoemde, wat by *kus* voorkom, daarop dui dat die lemma bo die neutrale vlak van die standaardtaal lê (Odendaal en Gouws 2005: xix, xv). Alhoewel temporele of herkomsetiket-

ontbreek wat die woordeboekgebruiker daarop wys dat dit hier eerder om verouderde of Neerlandistiese lemmas of betekenis gaan, word die gebruiker wel daarop bedag gemaak dat hierdie lemmas afwyk van die standaardtaal (Gouws 1989: 147).

Behalwe Neerlandistiese lemmas, is daar ook Neerlandistiese betekenisonderskeidings in *HAT 1* opgeneem. Vergelyk in hierdie verband die volgende woordeboekartikels in *HAT 1*:

bloot, b.nw. en bw. **1.** Kaal, naak, onbedek: *Met blote hoof, voete.* **2.** Sonder bedekking: *Onder die blote hemel. Met die blote swaard*, uit die skede getrek. *Met die blote oog*, sonder verkyker. **3.** Sonder meer, eenvoudig: *'n Blote gissing.*

gi-raf', (-s, -fe; -fie). **1.** Groot, geelbruin Afrikaanse dier met donkerbruin, onreëlmatige vlekke, baie lang nek en bene en gesplete kloue; kameelperd. **2.** Giraf. Naam van 'n sterrebeeld; *Camelopardalis*.

ste'ker, (-s; -tjie). **1.** Iemand wat steek. **2.** Voorwerp, werktuig waarmee gesteeek, wat ingesteeek [kan] word, o.a. 'n luns, 'n elektriese (in)steekkontak e.d.

In die eerste voorbeeld is "kaal" as eerste betekenisonderskeiding by *bloot* sonder 'n leksikografie etiket opgeneem om daarop te dui dat dit hier om 'n Neerlandistiese betekenisonderskeiding gaan (vgl. Van Houwelingen en Carstens 1998: 7). Die indruk word geskep dat dit sowel die gebruiklikste as frekwentste betekenisonderskeiding in Afrikaans is; iets wat deur Van Houwelingen en Carstens (loc.cit.) se studie weêrlê is. Alhoewel die gebruikliker, Standaardafrikaanse *kameelperd* as sinoniem by die eerste betekenisonderskeiding van *giraf* aangebied word, word die lemma *giraf* volledig bewerk terwyl *HAT 1* by die lemma *kameelperd* slegs die sinoniemdefinisie "giraf" aanbied. Die indruk wat hierdeur by die gebruiker geskep word, is dat *giraf* die gebruikliker, Standaardafrikaanse vorm in Afrikaans is, terwyl dit (of eerder die gebruikliker spelvariant *giraffe*) Standaardnederlands verteenwoordig (De Boer 2015; Den Boon en Hendrickx 2015). Die tweede betekenisonderskeiding wat by *steker* aangebied word, stem eerder ooreen met die Afrikaanse *kragprop*. Hierdie inskrywing skep weer die indruk dat die gebruiker hier met 'n gebruiklike betekenisonderskeiding in Afrikaans te make het deurdat daar geen leksikografiese etikette aangebring is nie. Lemmas soos *braaf*, *eventueel* en *fok* is monosemies hanteer met slegs die Neerlandistiese betekenis (vgl. Carstens 1995a: 152), sonder die inagneming van die addisionele betekenisonderskeidings of alternatiewe betekenis wat in Afrikaans ontwikkel het, naamlik "dapper", "uiteindelik" en die vulgêre betekenis van *fok* onderskeidelik:

braaf, b.nw. en bw. Gaaf, goed, deugszaam: *Sy brawe ouers. 'n Brawe Hendrik, Maria*, seun, dogter wat [kastig] baie voorbeeldig is.

e·ven·tu·eel', b.nw. en bw. Moontlik, gebeurlik: *Eventuele klagtes kan by my ingedien word.*

fok, ww. (gefok). Minder gewoon vir *teel*.

Carstens (1995a: 150) dui in haar bespreking van *HAT 3* daarop dat daar ook "in die definisies van talle leksikale en subleksikale lemmas argaïese (dikwels Nederlandse) [...] woorde behou" is. Vergelyk in hierdie verband die volgende woordeboekartikels in *HAT 1*:

eer'lik, b.nw. en bw. [...] **2.** Wat die waarheid spreek, wat die feite nie verdraai nie; opreg, betroubaar [...].

kleur'be·leid. Beleid ten opsigte van mense van 'n ander huidskleur.

stra'le·krans, stra'le·kroon. Krans, kring van lig om die hoof van goddelike, hemelse persone, op 'n skildery bv.; halo; oureool [...].

Die woorde *spreek, huidskleur* en *hoof* wat in bogenoemde definisies gebruik is, kon eerder met hulle frekwenter, Standaardafrikaanse ekwivalente *praat, velkleur* en *kop* vervang word.

Van Houwelingen en Carstens (1998: 10) dui verder daarop dat baie van die kotelesinligting wat in *HAT 3* aangetref word direkte vertalings van kotelesinligting uit die *Van Dale Groot Woordenboek der Nederlandse Taal* (1982) blyk te wees. Dieselfde geld *HAT 1*. Vergelyk in hierdie verband die poëme by *sielig, skaats* (s.nw.) en *spaarpot* in *HAT 1* met die poëme by *zielig, schaats* (s.nw.) en *spaarpot* in die *Van Dale Groot Woordenboek der Nederlandse Taal* (1961) (Kruyskamp 1961; vgl. Van Houwelingen en Carstens 1998: 10):

<i>HAT 1 (1965)</i>		<i>Van Dale 8 (1961)</i>	
Lemma	Poëem	Lemma	Poëem
sielig	Daar sielig uitsien.	zielig	Er zielig uitzien.
skaats	Die skaatse aanbind.	schaats	De schaatsen aanbinden.
spaarpot	Sy het 'n vet spaarpot.	spaarpot	Zij heeft een vette spaarpot.

Laastens bevat *HAT 1* ook Neerlandistiese kollokasies en uitdrukkings. Vergelyk in hierdie verband die volgende inskrywings by die lemmas *begaan* en *insien* (vgl. Ponelis 1992: 81; Luther 2010: 677):

be·gaan¹, ww. (—). **1.** Gaan, loop op: 'n *Steil pad begaan. Die begane grond, deel van gebou gelyk met die grond. [...]*

in'sien. I ww. (ingesien). [...] **4.** Beoordeel, waarneem: *Hy sien alles so duister in. [...]*

Die kollokasie *begane grond* wat by die eerste betekenisonderskeiding van *begaan* aangebied word, is eerder 'n Standaardnederlandse uitdrukking (Neder-

landse Taalunie 2015; De Boer 2015; Den Boon en Hendrickx 2015) wat in Afrikaans vervang kan word met *grondverdieping* of *grondvlak*. Die voorbeeldmateriaal wat by die vierde betekenisonderskeiding van *insien* aangebied word, is in werklikheid 'n Nederlandse uitdrukking wat beteken "om die ergste te verwag" (Den Boon en Hendrickx 2015). In bogenoemde woordeboekartikels is daar egter geen aanduiding deur byvoorbeeld 'n leksikografiese etiket dat dit hier om Neerlandistiese betekenisonderskeidings gaan nie.

2.2 HAT 2

In 1971 word F.F. Odendal, ook 'n voormalige redaksielid van die *WAT* en 'n Stellenbosch-alumnus, tóe verbonde aan die Randse Afrikaanse Universiteit, genader om die hoofredakteurskap van *HAT* oor te neem. Nadat Odendal die woordeboek, wat reeds deur die vorige redakteurs hersien en uitgebrei is, verder hersien, aangepas en uitgebrei het tot sowat 1 378 bladsye, verskyn die tweede uitgawe agt jaar later in 1979. Met die hersiening van hierdie uitgawe is daar veral gefokus op die korreksie van foute en die verfyning van definiense in bestaande woordeboekartikels, asook die byvoeging van lemmas "wat [...] algemeen in gebruik is" en die toevoeging van koteksinligting. Hiervoor is daar tot 'n mate geput uit die sesde uitgawe van die *WAT* (*kla-kol*). (Odendal 1979: Voorwoord by die tweede uitgawe; 2006: 283)

Volgens Odendal (2006: 283) is daar in hierdie uitgawe "argaïese woorde (en betekenis) geskrap, veral ook woorde en betekenis wat eerder Nederlands as Afrikaans was" (Odendal loc.cit.). Wanneer die lemmalys in 2.1 in ag geneem word, blyk dit egter dat daar in hierdie opsig in werklikheid nie veel sedert die vorige uitgawe verander het nie. Slegs een van die lemmas in hierdie lys is geskrap, naamlik *pinda*, terwyl daar ook weer ander Neerlandistiese lemmas toegevoeg is, soos *belendend* ("aangrensend") en *stekker* ("kragprop") (vgl. Ponelis 1992: 81; Van Houwelingen en Carstens 1998: 5). Daar is wel 'n aantal leksikografiese etikette bygevoeg wat die woordeboekgebruiker daarop bedag maak dat hierdie lemmas afwyk van Standaardafrikaans (Odendal 2006: 285; vgl. Gouws 1989: 201). By *bil* is die temporele etiket *veroud.* (verouderd) aangebring om daarop te dui dat die lemma "nog (dikwels) raakgeloop word maar tog in 'n mate in onbruik geraak het" (Odendal en Gouws 2005: xix). *Haag* en *internaat* is met die stilistiese etiket *w.g.* (weinig gebruiklik) gemerk en *twyg* met *dig.* (digterlik), wat daarop dui dat, alhoewel die lemma nie vreemd aandoen nie, dit meer algemeen voorkom in die werk van skrywers as in gewone spreektaal (Odendal en Gouws 2005: xvi). Saam met die stilistiese etiket *nie alg.* (nie algemeen), word die herkomstetiket *F.* (Frans) by die lemma *visite* aangebring. Eersgenoemde dui daarop dat die lemma 'n beperkte gebruik geniet (Odendal en Gouws 2005: xviii), waarskynlik aangesien daar 'n gebruikliker Afrikaanse ekwivalent bestaan ("besoek"). Laastens is die etiket *onpers.* (onpersoonlik) by die lemma *benieu* aangebring. Buiten die feit dat die afkorting *onpers.* in *HAT 2* self opgeneem is, word daar nie in enige voortekste aangedui

wat met hierdie etiket bedoel word nie; ook nie in latere uitgawes nie.

Die Neerlandistiese betekenisonderskeidings by *bloot*, *steker*, *braaf* en *eventueel* is behou sonder dat die Afrikaanse betekenis of betekenisonderskeidings bygevoeg is of as die gebruiklikste betekenis aangedui is. By *braaf* is daar selfs verder gegaan deurdat die volgende opmerking aan die woordeboekartikel toegevoeg is:

Braaf het nie in Afr. die betekenis van "dapper" nie.

Hierdeur is die Neerlandistiese betekenis nog meer eksplisiet gemaak. Ook by *fok* is die Afrikaanse betekenis steeds nie toegevoeg nie, maar die lemma met sy Neerlandistiese betekenis is wel uit die woordeboek verwyder.

Dat Odendal (2006: 283) inderdaad "omsigtig te werk gegaan" het "veral met die oog op ouer taalgebruikers", blyk daaruit dat daar nog baie Neerlandistiese lemmas en betekenisonderskeidings in *HAT 2* behou is.

2.3 *HAT 3*

Die "nuwe, verbeterde, uitgebreide en omgewerkte derde uitgawe" verskyn in 1994 — vyftien jaar ná die verskyning van *HAT 2*. Dat daar in hierdie vyftien jaar daaraan gewerk is om 'n woordeboek daar te stel wat op leksikografiese beginsels geskoei is, blyk reeds uit die "Voorwoord by die derde uitgawe". *HAT 3* verteenwoordig volgens Beyer (2005: 4) "'n waterskeiding in gebruiksheid in dié woordeboek", aangesien Odendal (1994: Voorwoord by die derde uitgawe) in die voorwoord "verantwoording doen aan sy gebruikers met betrekking tot sy doelstellings en [...] van sy werkswyse". Sodoende word *HAT 3* "die eerste Afrikaanse woordeboek wat geprobeer het om 'n behoorlike uiteensetting te gee van sy werkswyse en die teoretiese beginsels waarop dit berus", aldus Odendal (2006: 286).

So word *HAT* vir die eerste keer pertinent as 'n standaardwoordeboek onderskei deurdat dit "nie 'n volledige beeld en behandeling van die Afrikaanse woordeskat kan of wil gee nie", maar "die algemeenste woorde van Standaardafrikaans" opneem (Odendal 1994: Voorwoord by die derde uitgawe). Die onderskeiding as standaardwoordeboek word ook bevestig deur die feit dat Odendal (loc.cit.) hom in die samestelling van *HAT 3* gehou het aan die "rigsnoer van wat [...] as Standaardafrikaans beskou word", naamlik die *Afrikaanse woordelys en spelreëls* soos saamgestel deur "die hoogste gesag oor Afrikaans", die Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. Gegewe die eksplisiete verbintenis met Standaardafrikaans in *HAT 3*, sou 'n mens verwag dat talle van die Neerlandismes wat nog aanwesig was in vorige uitgawes in hierdie uitgawe geskrap sou word. Dat Odendal (loc.cit.) "aan die konserwatiewe kant [ge]sondig" het, blyk daaruit dat slegs *toonladder* geskrap is uit die lys Neerlandistiese lemmas wat in 2.1 aangebied is.

As rede voer Odendal (loc.cit.) aan dat hierdie lemmas na sy mening "nog in genoeg geskrifte of in die taal van ouer sprekers voorkom".

Ten spyte daarvan dat daar nog heelwat Neerlandistiese lemmas in *HAT 3* voorkom, word die gebruik van hierdie lemmas beperk deurdat daar in hierdie uitgawe meer aandag geskenk is aan die toevoeging van leksikografiese etikette (Odendal loc.cit.; 2006: 285; vgl. Gouws 1989: 201). Die volgende stilistiese etikette is byvoorbeeld toegevoeg waar daar voorheen geen etikette was nie:

Lemma	Leksikografiese etiket(te)
abonneer/abonnement	w.g.
belendend	w.g.
benieud	w.g.
bloot	ietwat deftig
blootshoof(s)	deftig
felisiteer/felisitatie	deftig
hals	deftig
heg	deftig
houthakker	verouderend
invasie	w.g., deftig
invektief	w.g., deftig
invitasie	deftig
kameleon	w.g.
kloset	veroud.
moedervlek	veroud.
origineel	w.g.
pardon	w.g., deftig
serieus	deftig
sielig	deftig

Ander lemmas se leksikografiese etikette is uitgebrei, byvoorbeeld *benieu* is nou ook as *w.g.* (weinig gebruiklik) geëtiketteer, *internaat* as *deftig*, *twyg* as *veroud.* (verouderd) en *visite* as *deftig*. Daarbenewens word *haag* nou eerder as *verouderd* as *weinig gebruiklik* beskou en *kus* as *deftig* en *digterlik*. Hoewel die Neerlandistiese betekenisonderskeiding by *steker* nog voorkom, is daar in *HAT 3* nou 'n stilistiese etiket by die Neerlandistiese betekenisonderskeiding byvoeg om daarop te dui dat dit "nie as neutraal beskou kan word en dus vrylik gebruik kan word nie" (Odendal 1994: Voorwoord by die derde uitgawe). Alhoewel Odendal (loc.cit.) in die "Voorwoord" dui op die probleme wat hy ondervind het met die aanbring van die leksikografiese etikette en beken dat daar gevalle kon "deurglip" waarby 'n mens etikette sou verwag, het Odendal deur die ruimer insluiting van leksikografiese etikette in *HAT 3* aan "een van die woordeboek se sentrale opdragte van taalbeskrywing" voldoen, aldus Ludwig (1982, in Gouws 1989: 32). Leksikografiese etikette word egter nie net effektief in die

sentrale teks toegepas nie, maar ook deeglik in die "Voorwoord" verantwoord, aldus Beyer (2005: 31). Dit verteenwoordig volgens Beyer (2005: 6) ook goeie leksikografiese praktyk dat die funksie en aanwending van die leksikografiese etikette in 'n voorteks getiteld "Gebruiklikste etikette en afkortings" uiteengesit is.

Voorts is die koteksinligting by die frekwenter derde betekenisonderskeiding van *bloot* uitgebrei:

- bloot**, b.nw. en bw. **1.** (*ietwat deftig*) Kaal, naak, onbedek: *Met blote hoof, voete.*
2. Sonder bedekking: *Onder die blote hemel. Met die blote swaard, uit die skede getrek. Met die blote oog, sonder verkyker.* Vgl. ONTBLOOT. **3.** Sonder meer, eenvoudig: *'n Blote gissing. Sommer uit blote nuuskierigheid vra. Dit was blote toeval dat ek hom gesien het. Die blote feit dat jy so 'n wilde bewering maak, bewys jou onwetenskaplikheid, bevooroordeeldheid.*

Braaf en *eventueel* wat in die vorige uitgawes nog monosemies en met Neerlandistiese betekenis hanteer is, is in *HAT 3* uitgebrei met die toevoeging van Afrikaanse betekenisonderskeidings. *Braaf* kan nou ook "dapper" beteken, terwyl *eventueel* nou "uiteindelik" kan beteken. By die lemma *insien* (ww.), waar die Nederlandse uitdrukking "Hy sien alles so duister in" as voorbeeldmateriaal aangebied is, word die betekenisonderskeiding waarin hierdie uitdrukking voorkom nou as *weinig gebruiklik* geëtiketteer. Gebruikers word in hierdie uitgawe selfs direk "gewaarsku" teen die gebruik van Neerlandismes deurdat die volgende opmerking by die lemma *aanry* aangebring is:

OPM.: *Aanry* in die bet. "met iemand bots terwyl jy ry", "in 'n ongeluk betrokke wees" is 'n Neerlandisme wat vermy moet word.

In die vorige uitgawes is hierdie Neerlandistiese betekenisonderskeiding ongeëtiketteer (m.a.w. as Standaardafrikaans) aangebied.

Die voorbeeldmateriaal is ook enigermate "verafrikaans" deurdat Odendal sitate uit die Afrikaanse literatuur toegevoeg het waar voorheen slegs of geen poëme aangebring was (Odendal 2006: 286). Enkele voorbeelde hiervan is sitate uit die werk van D.J. Opperman (vgl. *blikkies*), Antjie Krog (vgl. *duf*, betekenisonderskeiding 1), Wilma Stockenström (vgl. *fluks*, betekenisonderskeiding 2), M.E.R. (vgl. *gemeen*, betekenisonderskeiding 3c), C.L. Leipoldt (vgl. *graf*, betekenisonderskeiding 1a), Boerneef (vgl. *hiernatoe*), en Elisabeth Eybers (vgl. *inkrimp*, betekenisonderskeiding 1b) wat in *HAT 3* aangebring is.

Die groter mate van leksikografiese verantwoordbaarheid waarna daar in hierdie uitgawe gestreef is, blyk ook uit Odendal (1994: Voorwoord by die derde uitgawe) se benadering ten opsigte van deskriptiwiteit teenoor preskriptiwiteit in *HAT 3*. Odendal (loc.cit.) se besluit om in *HAT 3* "onomwonde" aan te dui "as 'n sekere woord, vorm of betekenisonderskeiding nie aanvaarbaar is nie", strook met die meer preskriptiewe funksie van standaard verklarende woordeboeke (vgl. Gouws 1989: 70). Tog dui Carstens (1995a: 164) daarop dat daar tot 'n groot mate "gebreek [is] met die elitistiese preskriptiwiteit en

ongesonde purisme wat Afrikaans so lank genuilband het". Behalwe die opname van enkele woorde uit die omgangstaal soos *bodder*, *laitie*, *skedonk*, *skiewie*, *stres*, *tjopper* en *woema*, is "nie alles wat maar na Engels lyk of ruik, [...] geban nie", byvoorbeeld *blues*, *Country*, *cowboy*, *establishment*, *folk*, *jop*, *shandy*, *soul*, *steak*, *tjop* en *walkman* (Odendal loc.cit.; Carstens 1995a: 155).

2.4 HAT 4

Kort ná die verskyning van *HAT 3* in 1994 sluit R.H. Gouws, ook 'n Stellenbosch-alumnus en dosent aan die Departement Afrikaans en Nederlands, hom by Odendal as redakteur van *HAT* aan (Odendal en Gouws 2000: v; Odendal 2006: 287). Gouws, wat sedert 1980 taalkunde en leksikografie aan die US se Departement Afrikaans en Nederlands doseer, was teen hierdie tyd reeds 'n gerekende leksikograaf op sowel plaaslike as internasionale vlak (Odendal en Gouws loc.cit.; Odendal loc.cit.; Botha, in Botha, Mavoungou en Nkomo 2013: ix). As praktiese leksikograaf was hy betrokke by verskeie woordeboekprojekte, en as teoretiese leksikograaf het hy die leksikografie as vakgebied in Suid-Afrika gevestig en was hy ook reeds aktief betrokke op die internasionale leksikografiese verhoog (Odendal en Gouws loc.cit.; Odendal loc.cit.; Botha loc.cit.; Swanepoel, in Botha, Mavoungou en Nkomo 2013: Agterblad). Volgens Carstens (1994: 252-253) het die Afrikaanse metaleksikografie 'n hoogtepunt bereik met die publikasie van Gouws se leksikografiese handleiding *Leksikografie* in 1989 deurdat dit aanleiding gegee het tot 'n toename in die publikasie van leksikografiese artikels. Nog 'n dosent van die US se Departement Afrikaans en Nederlands wie se naam genoem word as iemand wat gehelp het om leemtes in *HAT 4* uit te wys en aanbevelings vir verbeterings gedoen het, is A.E. Feinauer (Odendal en Gouws loc.cit.).

Een van Gouws se eerste take was om 'n elektroniese weergawe van *HAT* op CD-ROM beskikbaar te stel. Die *ELHAT*, soos die elektroniese *HAT* genoem is, is in 1998 bekendgestel en het "met 'n besliste toegevoegde waarde die terrein van die elektroniese leksikografie betree". (Odendal en Gouws loc.cit.; Odendal loc.cit.) In 2000 verskyn die vierde uitgawe van *HAT*. Alhoewel hierdie uitgawe nie uitvoerig hersien is nie, is daar wel 'n aantal argaïese lemmas uit die woordeboek verwyder (Odendal en Gouws loc.cit.; Odendal 2006: 288). Dit sluit in *abonneer/abonnement*, *belendend*, *bigotterie* en *haag*. Waar verouderde, Neerlandistiese woorde nie geskrap is nie, is leksikografiese etikette weereens toegevoeg of aangepas ten einde op die beperkte gebruik van hierdie lemmas te dui. So kry *agternaam* en *bloesem* wat voorheen ongeëtiketteerd was nou onderskeidelik die stilistiese etikette *w.g.* (weinig gebruiklik) en *deftig*, terwyl *houthakker*, wat in *HAT 3* nog as *verouderend* geëtiketteer is, nou *verouderd* is. 'n Interessante geval is *benieu* en *benieud* wat, in stede van *weinig gebruiklik* (soos in *HAT 3*), nou verander is na *deftig*. Die Neerlandistiese betekenisonderskeiding by *braaf* ("Gaaf, goed, deugszaam") is in hierdie uitgawe as *verouderend* geëtiketteer, terwyl *fok* weer opgeneem is, maar hierdie keer met sy Afrikaanse (vulgêre)

betekenis. Die lemma *kleurbeleid* (en daarmee saam sy Neerlandistiese definiens) is ook geskrap, terwyl die poëem "Die begane grond" nou as *weinig gebruiklik* geëtiketteer is.

Dat daar, soos in die vorige uitgawe, enigsins wegbeweeg is van 'n streng preskriptiewe benadering, blyk volgens Beyer (2005: 3) daaruit dat daar in die makrostruktuur nie slegs standaardtaalitems opgeneem is nie, maar ook "woorde wat algemeen in die spreektaal voorkom" (Odendal en Gouws loc.cit.). Dit sluit direkte en verafrikaanste leenwoorde uit Engels in, byvoorbeeld *cool*, *sorry*, *oukei* en *orraait* (Odendal en Gouws loc.cit.). Dat hierdie lemmas egter met die etikette *geselst.* (geselstaal) en *E.* (Engels) gemerk is, dui daarop dat dit veral tot die informele register beperk is, hoewel dit as "nuttig, algemeen, gebruiklik" beskou word (Odendal en Gouws 2000: xvii-xviii). *HAT* 4 vervul dus "'n deskriptiewe funksie deurdat die leksikon van **Standaardafrikaans** beskryf word" (Beyer loc.cit.; My beklemtoning — G.O.), maar ook deur erkenning te gee aan die omgangstaal "van 'n groot groep sprekers van die standaardvariëteit van Afrikaans" (Odendal en Gouws 2000: v). Dié woordeboek is egter steeds preskriptief deurdat daar duidelike uitsprake gemaak word oor die standaardheid, al dan nie, van sekere lemmas (Beyer loc.cit.). Tog dui die opname van spreektaalvorme enigermate op 'n erkenning van Afrikaans as 'n selfstandige, lewenskragtige taal wat nie langer op Nederlands hoef te steun nie (sien 2.1), maar wat nou "deur eie taalmiddels" in sy sprekers se behoeftes kan voorsien.

Soos in *HAT* 3 is die aantal sitate in woordeboekartikels uitgebrei ten einde "'n beeld van die werklike gebruik van Afrikaans te gee". In hierdie uitgawe is daar veral gefokus op sitate uit ligte literatuur en literatuur deur jonger skrywers, om sodoende verder erkenning te gee aan die lewenskragtigheid van Afrikaans. (Odendal en Gouws loc.cit.)

2.5 *HAT* 5

In 2005, veertig jaar ná die verskyning van die eerste *HAT*, word die vyfde uitgawe gepubliseer (Odendal 2006: 280). Alhoewel *HAT* 5, soos die vorige uitgawe, nie grondig hersien is nie, is daar volgens Botha (2005: 15) tog "heelwat waarde toegevoeg tot die inhoud van die vierde uitgawe". In die redakteurs se volgehoue poging om die makrostruktuur van *HAT* so verteenwoordigend moontlik te maak van die volle Afrikaanse spraakgemeenskap en daardeur uiting te gee aan die selfstandigheid en lewenskragtigheid van Afrikaans (vgl. 2.4), is daar weer 'n aantal (veral Engelse) leenwoorde en woorde uit ander variëteite as Standaardafrikaans opgeneem (Odendal en Gouws 2005: vii; Odendal 2006: 289). Voorbeelde hiervan is die lemmas *baai* (tw.), *baasterd/bastard*, *babasit*, *babasitter*, *babawagter*, *gapjaar*, *gazebo*, *gender*, *globalisasie/globalisering*, *gumbootdans*, *happening*, *jackpot*, *mechanic*, *samie*, *SMS* en *vaalseun* wat opgeneem is (vgl. Botha loc.cit.). Daar is eweneens erkenning gegee aan die invloed van ander inheemse Afrikatale op Afrikaans deur die insluiting van 'n aantal herkomstetikette wat

Afrikaans se band met dié tale illustreer, byvoorbeeld *So*. (Sotho), *Xh*. (Xhosa) en *Z*. (Zulu) wat onderskeidelik by die lemmas *kgotla*, *Masakhane* en *kleza/klêza* aangebring is.

Die meer deskriptiewe benadering tot opname wat in *HAT 5* gevolg is, is volgens Botha (loc.cit.) waarskynlik te danke aan die feit dat die keuse van lemmas in hierdie uitgawe vir die eerste keer gebaseer is "op 'n verteenwoordigende, uitgebreide en gebalanseerde elektroniese korpus" (Odendal 2006: 288). Die gebruikmaking van 'n Afrikaanse korpus is eweneens tekenend van die ontwikkeling van Afrikaans tot selfstandige taal — daar hoef nie meer op 'n gevestigde standaardtaal (soos Nederlands) en sy leksikografiese produkte gesteun te word om Afrikaans op leksikografiese gebied uit te brei nie (sien 2.1), aangesien Afrikaans oor voldoende middele beskik om self 'n verteenwoordigende standaard daar te stel.

Vervolgens is daar in hierdie uitgawe weer Neerlandistiese lemmas soos *skaavit* en *stekker* verwyder, terwyl *blootsvoet(s)*, *onbesuis* en *spaarpot* onderskeidelik met die stilistiese etikette *deftig*, *verouderend* en *deftig*, en *w.g.* (weinig gebruiklik) gemerk is. By *eventueel* word die Afrikaanse betekenisonderskeiding ("uiteindelik") soos oorgeneem uit Engels, nou eerste aangebied, terwyl die Neerlandistiese betekenisonderskeiding ("moontlik, gebeurlik") nou as *Neerlandisties* en *weinig gebruiklik* geëtiketteer is. By *stralekrans/stralekroon* is die woord *hoof* wat sedert die eerste uitgawe in die definieens voorkom uiteindelik met *kop* vervang. In *HAT 5* word daar selfs 'n stappie verder gegaan as in vorige uitgawes, deurdat gebruikers by sommige lemmas deur middel van leksikografiese etikette en in teksbokse gewaarsku word teen die gebruik van Neerlandismes. Vergelyk in hierdie verband die leksikografiese etiket *Neerlandisties* by *eventueel* soos uitgewys hierbo en by die lemmas *geluiddemper*, *geluiddig* en *geluidsgrens*, die etiket *Neerlandisme* by die lemma *dagorde*, die etiket *Ndl.* (Nederlands) by die lemmas *geneswyse* en *weder-*, asook die etiket *Meer Ndl.* by die lemma *geneë*.⁵ Die opmerking wat gebruikers sedert *HAT 3* teen die Neerlandistiese betekenisonderskeiding van die lemma *aanry* waarsku, is nou selfs meer prominent gemaak deurdat dit in 'n teksboks geplaas is:

Aanry in die bet. "met iemand bots terwyl jy ry", "in 'n ongeluk betrokke wees" is 'n Neerlandisme wat vermy moet word.

Gebruikers word in hierdie uitgawe as 't ware weer aktief "gewaarsku" teen die gebruik van Neerlandismes, wat enigermate daarop dui dat Afrikaans 'n onafhanklike taal naas Nederlands is wat eie ontwikkeling ten opsigte van sy leksikon ondergaan het en nie meer onder die invloed van Nederlands staan nie.

Ná die verskyning van *HAT 5* tree Odendal na meer as dertig jaar uit as redakteur van *HAT* (Odendal 2006: 289). Gouws bly egter betrokke as konsultant en nasienredakteur. Twee jaar later stel die uitgewer vir die eerste keer in *HAT* se geskiedenis 'n voltydse redakteur, die leksikograaf J. Luther, aan. (Luther, Pheiffer en Gouws 2015: vii) In 2009 verskyn die derde druk van *HAT 5* onder

die verkorte titel *Handwoordeboek van die Afrikaanse Taal*. Hiermee saam verskyn daar ook 'n nuwe CD-ROM-weergawe van *HAT 5*, die *e-HAT*, met verbeterde spel- en setfoute onder redaksie van Gouws en Luther. (Gouws en Luther 2009: Voorwoord by dié elektroniese uitgawe)

2.6 HAT 6

In 2012 word nog 'n voltydse leksikograaf, F. Pfeiffer, deur *HAT* se uitgewers, Pearson, aangestel met die oog op "'n omvattende herbewerking en uitbreiding van die *HAT*" (Luther, Pfeiffer en Gouws 2015: vii). So verskyn daar in 2015, 'n halfeeu ná die verskyning van die eerste *HAT*, 'n sesde uitgawe van die *Handwoordeboek van die Afrikaanse Taal*.

In *HAT 6* word daar, selfs meer as in vorige uitgawes, 'n deskriptiewe benadering gevolg. Alhoewel hierdie woordeboek "wat spelling en skryfwyse betref, op enkele uitsonderings na die leiding van die *Afrikaanse woordelys en spelreëls*" volg, is hierdie standaardwoordeboek se hoof funksie nie voorskrywing nie, maar om "die verskuiwende standaardvariëteit van Afrikaans" te beskryf, aldus Luther, Pfeiffer en Gouws (2015: vii-viii, ix). So is daar in sekere gevalle van die *Afrikaanse woordelys en spelreëls* afgewyk deur die opname van alternatiewe spel- en skryfwyses in *HAT 6*, selfs al word dit nie in die tiende uitgawe van die *Afrikaanse woordelys en spelreëls* erken nie (Luther, Pfeiffer en Gouws 2015: ix). Soos met die vorige uitgawe is dié deskriptiewe benadering tot 'n groot mate moontlik gemaak deur die elektroniese teksargiewe en woordkorpusse waarop *HAT 6* gebaseer is (Luther, Pfeiffer en Gouws 2015: viii). In die samestelling van *HAT 6* het die redakteurs hulle egter nie alleen op geskrewe standaardtaalbronne beroep nie, maar ook op gesproke en informeler taalgebruik oor die radio, op televisie en in die sosiale media en het hulle ook insae in die werk van verskeie taalberoepslui (bv. joernaliste, teksredakteurs, skrywers en vertalers) gehad (Luther, Pfeiffer en Gouws 2015: vii-viii). Die redakteurs het dus weer uit Afrikaans se eie middele geput om dié woordeboek saam te stel en het dit nie meer noodsaaklik geag om op Nederlands te steun nie (sien 2.1).

Dat *HAT 6* inderdaad "die grootste hersiening sedert die verskyning van die derde uitgawe" verteenwoordig, blyk reeds uit die makrostruktuur van die woordeboek (Luther, Pfeiffer en Gouws 2015: vii). 3 228 nuwe lemmas is opgeneem (Luther, Pfeiffer en Gouws loc.cit.), onder andere *avatar*, *bed-en-ontbyt-plek*, *dim sum*, *feesvoos*, *Greenpeace*, *homofobie*, *internetverslawing*, *Jozi*, *kaderontplooiing*, *kamp* (b.nw./bw.), *kitsskare*, *linkerbreinmens*, *metroman/-seksueel*, *naanbrood*, *opkikkertoets*, *paggeld*, *reeksverkragter*, *selfie*, *selfdood*, *slimfoon*, *tjoeftjaf*, *vetdruk* en *werklikheidsprogram/-reeks/-televisie*. Ten einde 'n volledige beeld te gee "van hoe Afrikaans vandag gebruik word", is daar in *HAT 6* se makrostruktuur dan ook, soos in vorige uitgawes, leksikale items uit ander variëteite van Afrikaans ingesluit (Luther, Pfeiffer en Gouws 2015: vii-viii). 'n Variëteit van Afrikaans wat besondere aandag in *HAT 6* geniet, is Moslemafrikaans; dit is "geloofs- en

kultuurverwante woorde van Maleise en Arabiese herkoms [...] uit die skryftaal van Moslem-Afrikaanssprekendes" (Taalkommissie 2009: ix). Voorbeelde hiervan is *abdas*, *batja*, *djannat*, *ëpoewasa*, *fadjer*, *garaam*, *insja'Allah*, *kanalla*, *koenoet*, *miembar*, *nabi*, *oestaad*, *popgloër*, *popgloëry*, *rasoel*, *salaah* en *tartiep* (vgl. Hugo 2015: 13). Volgens Pheiffer, een van HAT 6 se redakteurs (in Morris 2015), is HAT 6 ook een van die eerste Afrikaanse publikasies wat tot 'n groter mate erkenning gee aan Kaaps deur die opname van leksikale items uit dié variëteit. Dit sluit lemmas soos *gattas*, *kappityt*, *piemp*, *sharp*, *sjarrap*, *tiep* en *uitstiek/ytstiek* in (Morris 2015). Hugo (2015: 13) is egter van mening dat Kaaps as variëteit nog afgeskeep word in HAT. 'n Groetvorm soos *awe* wat deesdae al hoe algemener is, ontbreek byvoorbeeld nog. Daar is ook 5 365 nuwe betekenisonderskeidings in HAT 6 opgeneem (Luther, Pheiffer en Gouws 2015: vii), byvoorbeeld *afpak* ("iemand van jou probleme/ellendes vertel"), *duidelik* (b.nw.) ("goed; reg; mooi; lekker"), *gevaarlik* ("baie"), *tik²* (ww.) ("metamfetamien gebruik") en *venster* ("rame op 'n rekenaarskerm"). Die groter mate van erkenning wat daar in HAT 6 aan die variëteite van Afrikaans gegee word, spreek van 'n omarming van "eie taal-goedere" van Afrikaans (teenoor die omarming van die "vreemde", Nederlandse, waardeur die eerste uitgawes gekenmerk is).

Aangesien HAT veral op "eietydse taalgebruik" fokus, is daar meer as 3 000 ouer woorde en uitdrukkings geskrap (Luther, Pheiffer en Gouws loc.cit.). 'n Mens sou dus verwag dat die Neerlandistiese lemmas in hierdie uitgawe aansienlik verminder is. Wanneer die Neerlandistiese lemmas of lemmas met 'n Neerlandistiese betekenis in 2.1 weer in oënskou geneem word, blyk dit wel die geval te wees. Lemmas wat sedert die vorige uitgawe uit dié lys geskrap is, is *agternaam*, *benieu*, *benieud*, *bil*, *bloesem*, *blootshoof(s)*, *blootsoet(s)*, *internaat*, *invitasie*, *kneeding*, *knee(d)baar*, *onbesuis* en *twyg*. 'n Verdere groep lemmas uit hierdie lys wat wel in HAT 6 behou is, is óf vir die eerste keer van leksikografiese etikette voorsien, óf die leksikografiese etikette is uitgebrei — meestal om aan te dui dat hierdie lemmas 'n lae gebruiksfrekwensie het of verouderend/verouderd is:

Lemma	Etiket(te) HAT 5	Etiket(te) HAT 6
bloot	ietwat deftig	deftig, verouderend
briewebus	–	deftig, verouderend
hals	deftig	deftig, verouderend
heg	deftig	deftig, verouderend
jarig	–	nie alg., verouderend
spaarpot	w.g.	w.g., veroud.
swyneboel	–	neerh., w.g., veroud.
visite (s.nw.)	F., nie alg., deftig	Fr., deftig, w.g.

In ander gevalle is 'n kruisverwysing na 'n lemma bygevoeg wat 'n gebruikliker, Afrikaanse vorm verteenwoordig:

Lemma	Etiket(te) HAT 5	Etiket(te) HAT 6
briewebesteller	–	veroud. →POSBODE
felisiteer / felisitasie	deftig	deftig, veroud. →GELUKWENS
kameleon	w.g.	soöl., w.g. →VERKLEURMANNETJIE
moedervlek	veroud.	veroud. →GEBORTEMERK
pardon (tw.)	w.g., deftig	deftig, w.g. →EKSKUUS
steker	–	elektr. →KRAGPROP
stopkontak	–	elektr. →KRAGPUNT
tarwe	w.g.	landb., w.g. →KORING

Dit geld ook die Neerlandistiese betekenisonderskeiding (1) by *giraf*:

gi·raf *s.nw.* [~s; ~fie] 1 (soöl.) →KAMEELPERD 2 Giraf (*astron.*) naam van die sterrebeeld Camelopardalis. [...]

Ander lemmas se leksikografiese etikette is onveranderd gelaat sedert die vorige uitgawe:

Lemma	Etiket(te) HAT 5	Etiket(te) HAT 6
fiets (ww.)	–	–
houthakker	veroud.	veroud.
invasie	w.g., deftig	w.g., deftig
invektief	w.g., deftig	w.g., deftig
kapper	–	–
kloset	veroud.	veroud.
kus (s.nw./ww.)	deftig, digt.	deftig, digt.
naak	deftig, veroud.	deftig, veroud.
origineel	w.g.	w.g.
serieus	deftig	deftig
sielig	deftig	deftig

Uit die hele lys van Neerlandismes, is slegs *fiets* ("fietsry") en *kapper* ("haarkapper") steeds ongeëtiketteerd opgeneem. Dié lemmas se gebruikliker, Standaardafrikaanse sinonieme (bo tussen hakies aangedui) vorm wel deel van die semantiese kommentaar. Neerlandismes wat nog in die vorige uitgawe in definiense en voorbeeldmateriaal teenwoordig was, is in HAT 6 óf geskrap, óf aangepas. Die kollokasie *begane grond* en die uitdrukking [*iets*] *duister insien* is geskrap. Vergelyk ook in hierdie verband die vervanging van "spreek" met "praat" en "aanbind" met "aantrek" by die lemmas *eerlik* en *skaats* onderskeidelik:

eer·lik *b.nw., bw.* [~e, ~er, die ~ste] 1 wat die waarheid praat; wat die feite nie verdraai nie: [...].

skaats ■ *s.nw.* 1 [~e] een van 'n paar spesiale skoene met aan die onderkant skerp lemme (ysskaatse) of wieletjies (rolskaatse) waarmee 'n mens oor ys of op 'n harde oppervlak kan beweeg: *jou skaatse aantrek*. [...].

Die poëem "Sy het 'n vet spaarpot" wat voorheen by die lemma *spaarpot* voorgekom het, is nou uitgebrei tot "Ná jare se geduldige spaar het sy nou sy nou 'n vet spaarpot vir 'n oorsese reis" en die poëem "Daar sielig uitsien" kom steeds by die lemma *sielig* voor. Ten spyte van laasgenoemde, beskou Hugo (2015: 13) hierdie "suiwering", soos hy dit noem, saam met die opname van nuwe lemmas as die grootste aanwinst van *HAT 6*.

Alhoewel baie ouer woorde uit *HAT 6* verwyder is, is hierdie woorde nie heeltemal verlore vir Afrikaanssprekendes nie. Die woorde is behou in 'n aanlyn "moederwoordeboek", die nuwe *HAT Aanlyn* op die internet (Luther, Pheiffer en Gouws 2015: vii-viii). Hierdie "lewende" woordeboek, wat saam met die vrystelling van *HAT 6* bekendgestel is, is nie 'n blote duplikaat van die gedrukte weergawe nie, maar word sedert sy ontstaan gereeld bygewerk ten einde reg te laat geskied aan Afrikaans as lewende, ontwikkelende taal (Luther, Pheiffer en Gouws 2015: viii). Volgens Luther (2016) word die *HAT Aanlyn* se inhoud nou omtrent daaglik aangevul en bygewerk.

Wanneer die agterteks in *HAT 6*, 'n lys met erkennings wat die name van persone bevat uit wie se werk sitate geneem is vir opname in *HAT 6*, in oënskou geneem word, blyk dié woordeboek se band met die US se Departement Afrikaans en Nederlands. Die lys name sluit 'n hele aantal voormalige en huidige dosente van die US se Departement Afrikaans en Nederlands in, onder andere Willem Anker, Jan F.E. Celliers, Johan Combrink, Meyer de Villiers, Rufus Gouws, J.C. Kannemeyer, Willem Kempen, W.E.G. Louw, D.J. Opperman, F.A. Ponelis, Merwe Scholtz, Lina Spies, Marlene van Niekerk, Dirk van Schalkwyk, Dorothea van Zyl en Louise Viljoen (Luther, Pheiffer en Gouws 2015: 1615-1616).

Hugo (2015: 13) verwys na *HAT 6* as "die heel beste en grootste verklarende handwoordeboek in Afrikaans". Desnieteenstaande die aanvanklike steun op Nederlands, het die *Handwoordeboek van die Afrikaanse Taal* hom oor die verloop van jare losgemaak van Nederlandse invloed en 'n voorloper op die gebied van die Afrikaanse verklarende leksikografie geword.

3. Nasionale Woordeboek

3.1 Kernwoordeboek van Afrikaans

Slegs twee jaar ná die verskyning van die *Verklarende Handwoordeboek van die Afrikaanse Taal*, verskyn daar in 1967 die *Kernwoordeboek van Afrikaans* (voortaan *KvA*) onder redakteurskap van J. Smuts, L.C. Eksteen, en die Stellenbosse dosent M. de Villiers. De Villiers was van 1947 tot 1952 senior lektor in Nederlands en Afrikaans aan Stellenbosch, in 1953 word hy aangestel as professor in

Nederlandse taalwetenskap en Germaanse filologie, en sedert 1975 tot met sy aftrede in 1978 beklee hy die Jan Marais-leerstoel in Afrikaans (Breuer 2016). Met die verskyning van *Kernwoordeboek* het De Villiers reeds verskeie boeke oor die Afrikaanse en Nederlandse grammatika gepubliseer, onder andere *Woordsoorte, werkwoorde en tye* (1948), *Afrikaans simplified for English-speaking students* (1950), *Moderne Afrikaanse taalstudie volgens die leerplanne van die junior sekondêre kursus: Afrikaans as eerste taal, Deel II: Standerds VII en VIII* (1955), *Nederlands vir Suid-Afrika* (1956), *Nederlands en Afrikaans* (1960), *Afrikaanse klankleer: Inleiding tot die fonetiek en fonologie* (1958, sedert die derde druk in 1965 gewysig en die titel verander na *Afrikaanse klankleer: fonetiek, fonologie en woordbou*).

Gegewe De Villiers se bemoeienis met Nederlands, sou dit nie verbasend wees indien *KvA* ook uitvoerige blyke gee van Nederlandse invloed nie. As 'n mens ter wille van vergelykbaarheid die lys Neerlandistiese lemmas in *HAT 1* (sien 2.1) met die lemmas in *KvA* vergelyk, blyk dit dat ses van die Neerlandistiese lemmas wat ongeëtiketteerd in *HAT 1* opgeneem is, nie in *KvA* opgeneem is nie, te wete *agternaam* ("van"), *bil* ("boud"), *briewebus* ("posbus"), *kneding* (handeling van "knie"), *pinda* ("grondboontjie") en *swyneboel* ("vuil huishouding"). 'n Groot aantal is egter, net soos in *HAT 1*, ongeëtiketteerd opgeneem:

abonneer ("inteken", bv. op 'n koerant), *abonnement* ("intekening/subskripsie"), *bloesem* ("bloeisel"), *blootshoof(s)* ("kaalkop"), *blootsvoet(s)* ("kaalvoet"), *briewebesteller* ("posbode/-man"), *felisiteer* ("gelukwens"), *felisitasie* ("gelukwensing"), *fiets* (ww.) ("fietsry"), *hals* ("nek"), *heg* ("heining"), *houthakker* ("houtkapper"), *internaat* ("kosskool"), *kapper* ("haarkapper"), *kloset* ("toilet"), *knee(d)baar* ("kniebaar"), *moedervlek* ("geboortemerk/-vlek"), *onbesuis* ("roekeloos/onstuimig/wild"), *pardon* (tw.) ("ekskuus"), *sielig* ("droewig"), *spaarpot* ("spaarbussie"), *stopkontak* ("muurprop/kragpunt"), *toonladder* ("toonleer")

Daar is hierbenewens ook Neerlandistiese lemmas in *KvA* opgeneem wat nie in *HAT 1* voorkom nie (vgl. 2.1), byvoorbeeld *parasjuut* ("valskerm") en *pindakaas* ("grondboontjebotter"), of wat eers in *HAT 2* opgeneem is, soos *belendend* ("aangrensend") en *stekker* ("kragprop"). Anders as in *HAT 1* waar slegs drie lemmas met leksikografiese etikette gemerk is (sien 2.1), is daar heelwat van die Neerlandistiese lemmas in *KvA* wat met die stilistiese etiket *ong.* (ongewoon) gemerk om daarop te dui dat daar vir die betrokke lemma "'n meer gewone of gebruiklike woord bestaan" (De Villiers, Smuts en Eksteen 1967: Toeligting vir die gebruiker):

benieu ("nuuskierig maak"), *benieud* ("nuuskierig"), *bigotterie* ("skynheiligheid"), *haag* ("heining"), *invasie* ("inval"), *invektief* ("skel(d)woord"), *invitasie* ("uitnodiging"), *jarig* ("verjaar"), *kameleon* ("verkleurmanneltjie"), *kus* (s.nw./ww.) ("soen"), *origineel* ("oorspronklik"), *serius* ("ernstig"), *skavuit* ("skelm/skurk"), *tarwe* ("koring"), *twyg* ("dun takkie"), *visite* (s.nw.) ("besoek")

Laasgenoemde is ook van die stilistiese etiket *plegt.* (plegtig) voorsien. Soos *HAT 1* bevat *KvA* verder ook Neerlandistiese betekenisonderskeidings:

bloot (blote) b.nw. 1. *kaal, onbedek.* Met die blote hoof. Die blote lyf. 2. *oop, sonder bedekking.* Die blote hemel, grond. 3. *sonder hulpmiddel.* Met die blote oog. 4. *eenvoudig. sonder meer, slegs.* 'n Blote gissing. Die blote feit.

'steker, 'stekker (-s) s.nw. 1. *steekkontak* 2. *Werktuig waarmee gestee word; priem.*

Die Neerlandistiese betekenisonderskeiding "kaal" word, soos in *HAT 1*, sonder 'n leksikografiese etiket as eerste betekenisonderskeiding by *bloot* aangebied. Die indruk word weer geskep dat dit 'n gebruikliker en frekwenter betekenisonderskeiding in Afrikaans verteenwoordig (sien 2.1). By *steker/stekker* word daar selfs verder gegaan as in *HAT 1* deurdat die Neerlandistiese betekenisonderskeiding eerste aangebied word. In *KvA* is daar eweneens lemmas monosemies hanteer met slegs hulle Neerlandistiese betekenis:

eventu'eel (-tuele) b.nw. *moontlik, gebeurlik; moontlikerwys.* Eventuele skade, besware. Afkorting ev. **eventuali'teit.**

fok (ge-) ww. (ong.) *teel.* **'fokkery.**

gi'raf (-fe, -s) s.nw. *dier met 'n ontsaglike lang nek en voorbene, wat in Afrika voorkom.*

Alhoewel die Afrikaanse (vulgêre) betekenisonderskeiding van *fok* nie in *KvA* opgeneem is nie, is daar wel met 'n leksikografiese etiket aangedui dat die Neerlandistiese betekenis wat wél opgeneem is, ongewoon is in Afrikaans. By die lemma *braaf* is die meer Afrikaanse betekenisonderskeiding egter opgeneem, hoewel dit nog as anglisisties gemerk is:

braaf (brawe; brawer, -ste) b.nw. (ong.) 1. *gehoorsaam, soet* 2. *deugsam, fatsoenlik, eerbaar* 3. (angl.) *dapper.*

Die Nederlandse invloed op mikrostrukturele vlak blyk ook uit die teenwoordigheid van Neerlandismes in die definisies en voorbeeldmateriaal. Vergelyk in hierdie verband *posbode* waar *briewebesteller* as sinoniemdefinisie aangebied word en die volledige bewerking dan by *briewebesteller* aangebied word. Die indruk wat hierdeur geskep word, is dat die Neerlandistiese *briewebesteller* die frekwenter vorm is.⁶

'posbode (-s) s.nw. *briewebesteller*

'briewebesteller (-s) s.nw. *posbeampste wat briewe aflewer.*

Dit is egter interessant om daarop te let dat die woord *posbode* in *briewetas* se definisies gebruik word en nié *briewebesteller* nie. Vergelyk ook die gebruik van

hoof en *huid* in onderstaande definiense, asook die kollokasie *begane grond* as deel van *begaan* se vierde betekenisonderskeiding:

'hoendervleis s.nw. 1. *vleis van 'n hoender* 2. *onwillekeurige sametrekking van bloedvate in die menslike huid en die spiertjies van die hare veroorsaak deur koue, vrees ens.*

'stralekrans (-e) s.nw. *kring van strale soos gedink en afgebeeld om die hoof van goddelike of heilige wesens.*

be'gaan II (-gane) b.nw. 1. *bekommerd, besorg.* – wees oor die toekoms. 2. *met meegevoel vervul.* Hy is – oor die armes. 3. *ingenome, geesdriftig.* – wees met 'n saak. 4. *Weer op begane grond wees, weer op vaste grond.* Die begane grond, *grondverdieping, natuurlike grondoppervlakte.* Begane weë, *gebaande paaie.* **begaandheid** (by 1, 2, 3).

KvA het ook, soos sommige artikelinskrywings in *HAT 1*, inskrywings wat skynbaar direkte vertalings van inskrywings in die *Van Dale Groot Woordenboek der Nederlandse Taal* (1961) is. Vergelyk in hierdie verband die Neerlandistiese lemma *luk* se artikel:

<i>KvA</i> (1967)	<i>Van Dale 8</i> (1961)
luk (ge-) ww. <i>goed uitval, tot stand kom, geluk, slaag.</i> Dit sal nie – nie.	luk'ken , (lukte, is gelukt), (onoverg.) 1. <i>gelukken, goed uitvallen, wel slagen, tot stand komen: [...] – dat zal niet lukken, niet gaan, niet mogelijk zijn, niet toegestaan worden; [...].</i>

Vervolgens blyk die invloed van Nederlands ook uit die agterteks "Spelling: enkele basiese beginsels". In die uiteensetting van die ortografiese beginsels van Afrikaans word daar naamlik telkens melding gemaak van die Afrikaanse ortografie se band met Nederlands (De Villiers, Smuts en Eksteen 1967):⁷

Die Afrikaanse spelling (soos die Nederlandse) toon in verskeie gevalle die tekortkominge van die Latynse alfabet vir ons klankweergawe. (640)

In die eerste plek is daar die feit dat 'n vyftig jaar gelede besluit is om ter wille van die historiese en lewende verband met Nederlands nie onnodig van dié taal se spelling af te wyk nie. Die bogenoemde gevalle is 'n voortsetting van die Nederlandse tradisies, en daarby kom nog die gebruik van *y* en *ei*, en van *f* en *v* onderskeidelik. In Nederlands was *y* en *ei* op 'n tyd aparte klanke en *f* en *v* is dit nou nog ten dele, maar in Afrikaans dui *y* en *ei* dieselfde klinker, en *f* en *v* dieselfde medeklinker aan, bv. *fyn* en *feit*; *fort* en *voor*. (640)

'n Tweede besondere spellingverskynsel (ook 'n tradisie in Nederlands) is om *d* vir die klank *t* en *b* vir die klank *p* te skryf [...]. (640)

In Afrikaans, soos in Nederlands en Duits, word woorde dikwels aanmekaar geskryf [...]. (643)

Uit die voorafgaande is dit duidelik dat die Nederlandse invloed ook tot *KvA* deurgedring het. Tog wil dit lyk asof daar in sekere opsigte in *KvA* probeer is om die Nederlandse invloed te temper deurdat minder Neerlandistiese lemmas opgeneem is of deur meer gebruiksheidsleiding te gee, byvoorbeeld deur die opname van meer leksikografiese etikette.

3.2 Nasionale Woordeboek

Slegs vier jaar na die verskyning van *KvA* verskyn 'n tweede uitgawe van dié woordeboek in 1971, maar nou onder die nuwe titel *Nasionale Woordeboek* (voortaan *NW*) met die subtitel *Afrikaanse woordverklaring*. Oor die verloop van die volgende sestien jaar sou daar nog vyf verdere uitgawes onder dié titel in 1977, 1979, 1981, 1985 en 1987 verskyn. Teen die tyd dat *NW* 6 in 1985 verskyn, het De Villiers se kollega aan die US se Departement Afrikaans en Nederlands, R.H. Gouws (sien 2.4), hom as mederedakteur by De Villiers, Smuts en Eksteen aangesluit.

Wat die Nederlandse invloed in *NW* betref, is daar in die sestien jaar nie veel gedoen om die Nederlandse invloed verder te temper nie. Al die Neerlandistiese lemmas wat in *KvA* opgeneem is, kom steeds met dieselfde leksikografiese etikette in *NW* 7 voor. Enkele uitsonderings is *belendend* wat sedert *NW* 3 met die etiket *ong.* (ongewoon) gemerk is, en *heg* wat sedert *NW* 6 met dié etiket gemerk is. Daar is egter ook enkele gevalle waar daar sprake van 'n terugwaartse stap in die rigting van vernederlandising blyk te wees. In *NW* 2 is daar in die agtertekst "Spelling: enkele basiese beginsels" selfs meer verwysings na Nederlands bygevoeg (De Villiers, Smuts en Eksteen 1971):

Om dié moeilikheid te bowe te kom, het ons taal-voorouers, die Nederlanders, mettertyd 'n stelsel ontwerp wat ons ook volg t.o.v. die tekort aan vokaalletters. (645)

Afgesien hiervan los die Nederlanders en ons die tekortkominge van die alfabet op deur letterkombinasies wat heeltemal ander klanke verteenwoordig as die afsonderlike letters [...]. (646)

Die lemmas *haag*, *invitasie* en *kus* ("soen") wat in *KvA* met die etiket *ong.* gemerk is, onderskeidelik in *NW* 6 en *NW* 7 weer ongeëtiketteerd opgeneem. Net so word die Neerlandistiese betekenisonderskeidings by *bloot* ("kaal, onbedek") en *steker/stekker* ("steekkontak") in *NW* 7 steeds eerste aangebied, terwyl *eventueel* en *giraf* steeds monosemies hanteer word met slegs hulle Neerlandistiese betekenis ("moontlik, gebeurlik" en "kameelperd" onderskeidelik). Hoewel *fok* se meer Afrikaanse betekenis nie bygevoeg is nie, is dié lemma met sy Neerlandistiese betekenis wel uit *NW* 7 verwyder. *Braaf* se Afrikaanse betekenisonderskeiding ("dapper") is in *NW* 7 steeds as anglisisme gemerk. Die Nederlandse invloed in die definisies en voorbeeldmateriaal wat in *KvA* teenwoordig was in die lemmas *posbode*, *hoenderoleis*, *stralekrans*, *begaan* en *luk* se artikels is ook

steeds aanwesig in *NW 7*. Ondanks die feit dat Neerlandismes skynbaar beter hanteer is in *KvA* as in *HAT 1*, is daar nie veel gedoen om die Nederlandse invloed oor die verloop van die sewe uitgawes van *NW* verder te verminder nie.

In *NW 5* is daar wel sprake van 'n meer deskriptiewe benadering deurdat daar in die "Toeligting vir die gebruiker" gemeld word dat, hoewel verskeie betekenisonderskeidings as anglisismes gemerk is, gebruik daarvan nie noodwendig afgekeur word nie (De Villiers, Smuts en Eksteen 1981: Toeligting vir die gebruiker).

4. Samevatting

As 'n mens *NW 7* met die uitgawe van *HAT* vergelyk wat gedurende dieselfde tyd in omloop was (dit is *HAT 2*), vaar *NW 7* beter as sy eweknie wat die hantering van Nederlandse invloed betref. Vergelyk in hierdie verband die volgende tabel met Neerlandistiese lemmas in *HAT 2* en *NW 7* onderskeidelik:

Lemma	HAT 2 (1979)	NW 7 (1987)
abonneer (ww.)/abonnement (s.nw.)	✓	✓
agternaam	✓	✗
belendend	✓	ong.
benieu	onpers.	ong.
benieud	✓	ong.
bigotterie	✓	ong.
bil	veroud.	✗
bloesem	✓	✓
bloot	✓	✓
blootshoof(s)	✓	✓
blootsvoet(s)	✓	✓
briewebesteller	✓	✓
briewebus	✓	✗
felisiteer (ww.)/felisitasie (s.nw.)	✓	✓
fiets (ww.)	✓	✓
haag	w.g.	✓
hals	✓	✓
heg	✓	ong.
houthakker	✓	✓
internaat	w.g.	✓
invasie	✓	ong.
invektief	✓	ong.
invitasie	✓	✓
jarig	✓	ong.
kameleon	✓	✓

kapper	✓	✓
kloset	✓	✓
kneding	✓	x
knee(d)baar	✓	✓
kus (s.nw./ww.)	verh. (s.nw.)	✓
moedervlek	✓	✓
onbesuis	✓	✓
origineel	✓	ong.
parasjuut	x	✓
pardon (tw.)	✓	✓
pinda	x	x
pindakaas	x	ong.
serieus	✓	ong.
sielig	✓	✓
skavuit	w.g.	ong.
spaarpot	✓	✓
steker	✓	✓
stekker	✓	✓
stopkontak	✓	✓
swyneboel	✓	x
tarwe	w.g.	ong.
toonladder	✓	✓
twyg	dig. t.	ong.
visite (s.nw.)	F., nie alg.	ong., plegt.

Uit bogenoemde is dit duidelik dat NW 7 meermale Neerlandistiese lemmas óf nie opgeneem het nie, óf van 'n leksikografiese etiket voorsien het waar dit nie in HAT 2 gedoen is nie.

Dit sou sekerlik vergesog wees om NW se enigszins beter hantering van Neerlandismes toe te skryf aan die Stellenbosse dosent, M. de Villiers, se betrokkenheid by dié woordeboek. Dit is myns insiens egter nie onwaarskynlik dat De Villiers se bemoeienis met en kennis van Nederlands (soos bv. weerspieël in sy publikasie van die boek *Nederlands en Afrikaans*; sien 3.1) aan hom 'n sensitiviteit ten opsigte van Nederlands se verhouding tot en invloed op Afrikaans gegee het wat uiteindelik oorgespoel het in die samestelling van NW nie. Dit is byvoorbeeld ook interessant om daarop te let dat die Departement Afrikaans-Nederlands, soos wat die US-departement waarvan De Villiers 'n personeellid was sedert 1952 bekend gestaan het,⁸ in 1979 herbenoem is na die Departement Afrikaans en Nederlands. Dit lyk dus asof daar gedurende die tyd wat NW verskyn het 'n groter bewustheid van Afrikaans as selfstandige taal náas Nederlands (en nie ondergeskik aan nie) in die Departement Afrikaans en Nederlands ontstaan het.

Later, ná die gerekende leksikograaf, R.H. Gouws se toetrede tot die redaksie van *HAT*, het hierdie bewustheid van Afrikaans as selfstandige en lewenskragtige taal oorgespoel na hiërdie standaardwoordeboek. Nadat *NW* ná die 1980's nie verdere uitgawes beleef het nie, kon *HAT* put uit die "Goue Eeu van die leksikografie", soos wat Gouws (1989: 24) na die tagtigerjare verwys. Die "ongekende oplewing" in "die teoretiese en praktiese leksikografie" wat hom in dié tydperk afgespeel het (Gouws loc.cit.) en onder andere gekenmerk is deur die publikasie van Gouws se "epogmakende werk" *Leksikografie* in 1989 (aldus Carstens 1994: 252), het ongetwyfeld sedert die negentigerjare 'n verrykende invloed op *HAT* gehad, deurdat dié woordeboek hom stadig maar seker van Nederlandse invloed begin losmaak het. Die gevolg is dat die jongste uitgawe van dié woordeboek blyke gee van Afrikaans as 'n selfstandige taal wat nie langer op Nederlands hoef te steun nie.

5. Slot

Die Nederlands-Afrikaans-stryd tussen die voorstanders van Nederlands en die voorstanders van Afrikaans wat aan die begin van die twintigste eeu binne die Departement Afrikaans en Nederlands, maar ook in die breër Afrikaanse spraakgebied, posgevat het, het 'n onmiskenbare invloed op die Afrikaanse leksikografie gehad, deurdat daar in die leksikografiese vergestaltung van Afrikaans telkens teruggegryp is na Nederlands (vgl. Odendaal 2016 en Odendaal, om te verskyn). Ten spyte daarvan dat die standaardwoordeboeke *HAT* en *NW*, waarby personeellede van die Departement betrokke was, heelwat later as hulle vertalende eweknieë verskyn het en hulle dus kon put uit die toenemende bewustheid van die studie van die Afrikaanse taalkunde en ander leksikografiese produkte, kon dié woordeboeke ook nie die Nederlandse invloed waarmee hulle voorgangers belas is (vgl. Odendaal 2016 en Odendaal, om te verskyn), vryspring nie. Soos wat die Departement Afrikaans en Nederlands se naamsverandering in die laat-sewentigs aantoon, het daar egter 'n veranderde ingesteldheid ten opsigte van Nederlands se verhouding tot Afrikaans in die tweede helfte van die vorige eeu ingetree wat uiteindelik ook 'n invloed gehad het op die hantering van Neerlandismes in standaardwoordeboeke. Die gevolg is dat hierdie woordeboeke 'n geleidelike losmaak van Nederlands toon en dat *HAT* vandag blyke gee van Afrikaans as 'n selfstandige taal wat nie langer op Nederlands hoef te steun nie, maar op eie leksikografiese bene kan staan.

Eindnote

1. Hierdie woordeboek kan volgens Gouws en Ponelis (1992: 20) nie as 'n standaardwoordeboek beskou word nie, aangesien dit te beperk in omvang is.
2. Die artikel vorm deel van 'n omvattender projek van die Dekaan van die Fakulteit Lettere en Sosiale Wetenskappe van die Universiteit Stellenbosch oor die geskiedenis van die Fakulteit. Dit is die derde artikel in 'n beoogde reeks artikels oor die rol van die Universiteit Stellenbosch se Departement Afrikaans en Nederlands in die ontwikkeling van die Afrikaanse leksiko-

grafie teen die agtergrond van Afrikaans se stryd teen Nederlands. Die artikel bied nie 'n volledige oorsig oor dié Departement se rol in die ontwikkeling van die Afrikaanse leksikografie nie, maar fokus in die besonder op dosente van die Departement se betrokkenheid by die verklarende standaardwoordeboeke *Verklarende Handwoordeboek van die Afrikaanse Taal* en *Nasionale Woordeboek*.

3. Met "Neerlandisties" word in hierdie artikel bedoel "wat die Nederlandse model navolg en in stryd is met die taaleie", terwyl 'n "Neerlandisme" op enige sodanige verskynsel in Afrikaans dui, byvoorbeeld 'n lemma, betekenisonderskeiding, uitdrukking, ensovoorts (vgl. Botha 2014).
4. Die lys is gedeeltelik oorgeneem uit Ponelis (1992: 80-81), Carstens (1995a: 154), Van Houwelingen en Carstens (1998: 5-8), Luther (2010: 677) en Luther (2015) en bygewerk uit *HAT 1* (Schoonees et al. 1965). Al die lemmas in die lys is gekontroleer aan die hand van die *Dikke Van Dale Online* (Den Boon en Hendrickx 2015) en *Van Dale Online Hedendaags Nederlands* (De Boer 2015). Behalwe *invektief* wat as formeel en *zielig* wat as informeel geëtiketteer is in *Hedendaags Nederlands*, kom al die lemmas of hulle spesifieke betekenis ongeëtiketteer in *Dikke Van Dale* en *Hedendaags Nederlands* voor, wat daarop dui dat hulle as deel van die Nederlandse standaardtaal beskou word (vgl. Gouws 1989: 147, 201).
5. Die onderskeid tussen hierdie drie etikette word nêrens in die buitetekste belig nie.
6. In *HAT 1* kry beide *briewebesteller* en *posbode* volledige bewerkings sonder 'n kruisverwysing na die ander lemma.
7. Die verbintenis van die Afrikaanse ortografie met Nederlands is in ooreenstemming met die grondbeginsels van die Afrikaanse spelling soos uiteengesit in die 1964-uitgawe van die *Afrikaanse woordelys en spelreëls* (vgl. Odendaal 2016: 267). Vergelyk byvoorbeeld die volgende aanhalings uit die 1964-AWS (Taalkommissie 1964):
"Die Afrikaanse spelling wil [...] rekening hou met die spelling wat vandag amptelik in Nederland erken word" (1)
"As ons alleen met die Afrikaanse uitspraak rekening gehou het, sou daar geen rede gewees het om in ons spelling tussen ei en y en tussen f en v te onderskei nie. [...] om kultuurhistoriese redes is besluit om in hierdie geval nie van Nederlands af te wyk nie" (1-2)
8. Dié departement het sedert sy ontstaan in 1922 bekend gestaan as die Departement Nederlands en Afrikaans.

Bronnelys

Woordeboeke

- Botha, W.F. (Hoofred.).** 2014. *Elektroniese WAT, A-SKOOI*. CD-ROM. Stellenbosch: Buro van die WAT.
- De Boer, T. (Hoofred.).** 2015. *Dikke Van Dale Online 2015. Hedendaags Nederlands*. Van Dale Uitgevers. <https://www.vandale.nl/zoeken/zoeken.do>.
- Den Boon, C.A. en R. Hendrickx (Hoofreds.).** 2015. *Dikke Van Dale Online 2015*. Van Dale Uitgevers. <https://www.vandale.nl/zoeken/zoeken.do>.
- De Villiers, M., J. Smuts en L.C. Eksteen.** 1967. *Kernwoordeboek van Afrikaans*. Kaapstad: Nasou.
- De Villiers, M., J. Smuts en L.C. Eksteen.** 1971. *Nasionale Woordeboek*. Tweede uitgawe. Kaapstad: Nasou.

- De Villiers, M., J. Smuts en L.C. Eksteen.** 1977. *Nasionale Woordeboek*. Derde, hersiene en uitgebreide uitgawe. Kaapstad: Nasou.
- De Villiers, M., J. Smuts en L.C. Eksteen.** 1979. *Nasionale Woordeboek*. Vierde, hersiene en uitgebreide uitgawe. Kaapstad: Nasou.
- De Villiers, M., J. Smuts en L.C. Eksteen.** 1981. *Nasionale Woordeboek*. Vyfde, hersiene en uitgebreide uitgawe. Kaapstad: Nasou.
- De Villiers, M., J. Smuts, L.C. Eksteen en R.H. Gouws.** 1985. *Nasionale Woordeboek*. Sesde, hersiene en uitgebreide uitgawe. Kaapstad: Nasou.
- De Villiers, M., J. Smuts, L.C. Eksteen en R.H. Gouws.** 1987. *Nasionale Woordeboek*. Sewende, hersiene uitgawe. Kaapstad: Nasou.
- Gouws, R. en J. Luther.** 2009. *e-HAT 2009*. CD-ROM. Kaapstad: Pearson Education SA.
- Kruyskamp, C.** 1961. *Van Dale Groot Woordenboek der Nederlandse Taal*. Agste, omvattend herbewerkte en vermeerderde druk. 's-Gravenhage: Martinus Nijhoff.
- Luther, J., F. Pheiffer en R.H. Gouws (Reds.)**. 2015⁶. *Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson.
- Odendal, F.F. (Hoofred.)**. 1979². *Verklarende Handwoordeboek van die Afrikaanse Taal*. Johannesburg/Kaapstad: Perskor.
- Odendal, F.F. (Hoofred.)**. 1994³. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.
- Odendal, F.F. en R.H. Gouws (Reds.)**. 2000⁴. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.
- Odendal, F.F. en R.H. Gouws (Reds.)**. 2005⁵. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson Education South Africa.
- Schoonees, P.C., C.J. Swanepoel, S.J. du Toit en C.M. Booysen.** 1965. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Klerksdorp/Johannesburg/Pretoria: Voortrekkerpers.

Ander bronne

- Beyer, H.L.** 2005. Standaardwoordeboek en kernwoordeboek: 'n Vergelykende evaluering. *Lexikos* 15: 1-37.
- Botha, W.** 2005. Nuwe HAT modern, leier in klas. *Die Burger*, 31 Oktober 2005: 15.
- Botha, W., P. Mavoungou en D. Nkomo (Reds.)**. 2013. *Festschrift: Rufus H. Gouws*. Stellenbosch: SUN PReSS.
- Breuer, R.** 2016. Meyer de Villiers. Stellenbosch Writers. <http://www.stellenboschwriters.com/devilliersme.html> (16 Mei 2016).
- Carstens, A.** 1994. *Verklarende Afrikaanse Woordeboek as spieël van normverplasing*. *Lexikos* 4: 249-281.
- Carstens, A.** 1995a. 'n Kritiese beskouing van HAT3. *Lexikos* 5: 138-165.
- Carstens, A.** 1995b. Ideologiese normverplasing en die Afrikaanse handwoordeboek van die negentigerjare. *Literator* 16(1): 13-30.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 1995. Ou wyn in nuwe sakke: 'n Metaleksikografiese herwaardering van enkele komponente in Mansvelt se *Idioticon*. Bosman, N. (Red.). 1995. *'n Man van Woorde*: 42-58. Pretoria: J.L. van Schaik.
- Gouws, R.H.** 2003. Oor patriotte en ander leksikografiese vernuwers. Botha, W. (Red.). 2003. *'n Man wat beur*: 71-85. Stellenbosch: Buro van die WAT.

- Gouws, R.H. en F.A. Poneis.** 1992. Issues in the Development of Afrikaans Lexicography. *Suid-Afrikaanse Tydskrif vir Taalkunde*. Supplement 12: 1-44.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hugo, D.** 2015. Daar is nog baie skop in dié 'antie'. *Rapport Weekliks*, 9 Augustus 2015: 13. <http://www.netwerk24.com/Vermaak/Boeke/Daar-is-nog-baie-skop-in-die-antie-20150809> (30 April 2016).
- Le Roux, T.H., D.F. Malherbe en J.J. Smith.** 1917. *Afrikaanse woordelys en spelreëls*. Bloemfontein: Het Volksblad-drukkerij.
- Luther, J.** 2010. A Language on the Back Foot. The Afrikaans Lexicographer's Dilemma. EURALEX. Euralex 2010 Proceedings. http://www.euralex.org/elx_proceedings/Euralex2010/054_Euralex_2010_3_LUTHER_A%20language%20on%20the%20back%20foot_The%20Afrikaans%20lexicographers%20dilemma.pdf (2 Junie 2016).
- Luther, J.** 2015. HAT 50 jaar. Lesing gelewer by die African Association for Lexicography (AFRILEX) se 20ste Internasionale Konferensie, 6–8 Julie 2015, Universiteit van KwaZulu-Natal, Durban, Suid-Afrika.
- Luther, J.** 2016. [Persoonlike korrespondensie per e-pos]. 12 Mei.
- Morris, M.** 2015. Kyk hoe lyk Afrikaans nou. *Sunday Argus*, 2 Augustus 2015: 17. <http://www.iol.co.za/news/kyk-hoe-lyk-afrikaans-nou-1894635> (7 Mei 2016).
- Nederlandse Taalunie.** 2015. Taalunieversum. Taaladvies. Gelijkvloers/begane grond/beneden-verdieping. http://taaladvies.net/taal/advies/vraag/1021/gelijkvloers_begane_grond_benedenverdieping/ (2 Mei 2016).
- Odendaal, G.** 2016. Die rol van die US se Departement Afrikaans en Nederlands in die ontwikkeling van die Afrikaanse leksikografie: Die Nederlands-Afrikaans-stryd gedurende die aanvangsjare. *Tydskrif vir Geesteswetenskappe* 56(1): 257-276.
- Odendaal, G.** Om te verskyn. Die rol van die US se Departement Afrikaans en Nederlands in die ontwikkeling van die Afrikaanse leksikografie onder Nederlandse invloed: *Woordeboek van die Afrikaanse Taal*.
- Odendal, F.F.** 2006. HAT veertig jaar — 'n persoonlike oorsig. *Lexikos* 16: 280-289.
- Poneis, F.A.** 1992. Standaardafrikaans in oorgang. Webb, V.N. (Red.). 1992. *Afrikaans ná apartheid: 69-89*. Pretoria: J.L. van Schaik.
- Schoonheim, F.** 1998. Een studie naar aspecten van de invloed van de Nederlandse lexicografie op Afrikaanse woordenboeken. Ongepubliseerde M.A.-verhandeling. Stellenbosch: Universiteit van Stellenbosch.
- Snijman, F.J. (Hoofred.).** 1964. *U woorde, u woordeboek*. Stellenbosch: Raad van Beheer oor Die Afrikaanse Woordeboek.
- Snijman, F.J.** 1975. U woorde, u woordeboek. *Lantern* 25(1): 2-9.
- Taalkommissie (van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns).** 1964. *Afrikaanse woordelys en spelreëls*. Sewende, verbeterde uitgawe. Kaapstad: Tafelberg.
- Taalkommissie (van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns).** 2009. *Afrikaanse woordelys en spelreëls*. Tiende, verbeterde en omvattend herbewerkte uitgawe. Kaapstad: Pharos Woordeboeke.
- Van Houwelingen, F. en A. Carstens.** 1998. "Nederlandismes" in HAT³. *Literator* 19(2): 1-12.

A Critical Analysis of Multilingual Dictionaries

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

Abstract: This article evaluates the lexicographic value of multilingual dictionaries. Dictionaries covering three or more languages spoken in South Africa are taken as a case in point. An attempt will be made to reflect on their merits and shortcomings as reference works and learning tools but the focus will be on presumed shortcomings in the macro and micro structures of such dictionaries with special attention to the lemmatisation of common words, quality of the dictionary articles and consistency in presentation.

Keywords: MULTILINGUAL DICTIONARIES, MACRO STRUCTURE, MICRO STRUCTURE, AFRICAN LANGUAGES, THEMATIC ORDERING, ALPHABETIC ORDERING

Opsomming: 'n Kritiese evaluering van veeltalige woordeboeke. In hierdie artikel word die leksikografiese waarde van veeltalige woordeboeke geëvalueer. Woordeboeke wat drie of meer tale wat in Suid-Afrika gepraat word leksikografies bewerk, dien as voorbeeld. Daar sal gepoog word om hulle sterk punte en tekortkominge as naslaanbronne en aanleerhulpmiddels te omskryf. Klem sal gelê word op veronderstelde tekortkominge in die makro- en mikrostrukture van sulke woordeboeke met spesifieke verwysing na die lemmatisering van algemeen gebruikte woorde, kwaliteit van die woordeboekartikels en konsekwentheid in die aanbieding.

Sleutelwoorde: MEERTALIGE WOORDEBOEKE, MAKROSTRUKTUUR, MIKROSTRUKTUUR, AFRIKATALE, TEMATIESE ORDENING, ALFABETIESE ORDENING

1. Introduction

This paper evaluates the lexicographic value of multilingual dictionaries. Dictionaries covering three or more languages spoken in South Africa are taken as a case in point. An attempt will be made to reflect on their merits and shortcomings as reference works and learning tools. The focus will be on presumed shortcomings in the macro and micro structures of such dictionaries with special attention to the lemmatisation of common words, quality of the dictionary articles and consistency in presentation. The Multi-Language Dictionary (MLD), Multilingual Illustrated Dictionary (MID), Concise Multilingual Dictionary (CMD), and the Oxford Junior Primary Dictionary for Southern Africa (OJPD) will be analysed as typical examples. Reference will also be made to The English-Afrikaans-Northern Sotho-Tswana Aid (EANTA).

It has to be stated at the outset in terms of Gouws et al. (2014: 25) that no single dictionary can be everything for everyone.

Among the merits of these dictionaries count the impressive presentation

of specific everyday themes such as 'my body', 'family and friends', 'clothing', 'my home', thus in most cases a complementary thematic approach in which a number of typical situations in everyday life are presented with or without pictorial illustrations and the relevant words linked to an alphabetical section where they are presented as treated or un-treated lemmas.

Such a thematic approach naturally reflects appropriate contextualisation of lexical items that belong together which would otherwise have been scattered all over the dictionary if only an alphabetical ordering was followed. Furthermore, most of these dictionaries are appealing to the eye, attractive to use and encourage learning and dictionary use. So, for example, giving a picture of the face of a person indicating all the different parts in seven languages, encourages language learning, especially for young learners. Such dictionaries usually also score high marks in terms of comparison between different languages. The user can see at a glance how the translation equivalents in the other languages compare to his/her source and target language knowledge.

In contrast, however, it will be argued that the multilingual model underpinning the compilation of multilingual dictionaries is problematic on many levels, resulting in products of inferior lexicographic quality. This is aggravated by various degrees of inconsistency, the lack of cohesion in the treatment of the different languages and even many basic or sloppy errors in the dictionaries. Problematic aspects to be explored are (a) poor and inconsistent covering of lexical items, (b) impoverished/skeleton dictionary articles: lack of sense distinction, lack of parity between translation equivalents, absence of examples of usage, incomplete translation equivalent paradigms, etc. and (c) risk of misleading the user through inappropriate text/speech production or communicative guidance. A brief outline of the contents and layout of the dictionaries will first be presented as a basis for the discussion.

The preferred reference to the languages in this article is to give the names of the languages in the specific language, i.e. Afrikaans, English, Sepedi, Setswana, Sesotho, isiZulu and isiXhosa. The comparative references in MID, MLD, CMD, OJPD and EANTA are given in table 1.

	Afrikaans	English	Sepedi	Setswana	Sesotho	IsiZulu	IsiXhosa
MID	Afrikaans	English	Sepedi	Setswana	Sesotho	IsiZulu	IsiXhosa
MLD	Afrikaans	English	N(orthern) Sotho	Tswana	Sesotho	Zulu	Xhosa
CMD	Afrikaans	English/Engels	Northern Sotho/ Noord-Sotho	Tswana	Southern Sotho/ Suid-Sotho	Zulu/ Zoeloe	Xhosa
OJPD	Afrikaans	English	North Sotho	Setswana	South Sotho	-	-
EANTA	Afrikaans	English	Northern Sotho	Tswana	-	-	-

Table 1: Names of the languages used in MID, MLD, CMD, OJPD and EANTA

2. Dictionary versus text book as reference source

It could be argued that most multilingual dictionaries enable thematic as well as alphabetic lookup and viewed as such, this is a strong point. MID, EANTA and MLD attempt to strike a good balance between thematic and alphabetic lookup by means of a system of cross-referencing between the thematic sections and the alphabetical stretches. Using multilingual dictionaries such as MID, MLD and EANTA, be it through thematic or alphabetical lookup, also entails an element of "reading the dictionary". A substantial part of the learning potential of such dictionaries is to not only look up words in a specific theme, e.g. the human body, but then also to learn the translation equivalents for the other body parts. So, an issue to be considered at the outset for multilingual dictionaries is the question as to what extent the average user is interested in *reading* a dictionary versus *consulting* a dictionary for e.g. the meaning of a specific word. Getting maximum benefit from multilingual dictionaries implies/presumes a fair amount of reading. Martin, *Pharos Groot Woordeboek. Afrikaans en Nederlands* (ANNA) attempts to make a case for the value of reading a dictionary for recreation in addition to the standard procedure of consulting a dictionary.

Laaste maar nie die minste nie is 'n geamalgameerde woordeboek ook 'n leeswoordeboek: 'n woordeboek waarin 'n mens nie slegs vertaalekwivalente soek of verskille vind nie, maar een waarin jy ook net kan lees vir ontspanning. (ANNA: 23)

[Last, but not least, an amalgamated dictionary is also a dictionary for reading: a dictionary where you not only search for translation equivalents or contrasts, but one which one can read for recreation.]

This assumption, however, should be supported by actual user studies and falls outside the scope of this article.

3. Macrostructural considerations

3.1 Dictionary components of multilingual dictionaries

MID presents its table of contents in the seven languages covered by the dictionary, i.e. English (Contents), isiZulu (Okuqukethwe), Sesotho (Dikahare), isiXhosa (Isiqulatho), Setswana (Diteng), Afrikaans (Inhoud) and Sepedi (Diteng). It is not clear on what basis the sorting order for these languages was determined. So, for example, it would have been better to group the related African languages together, i.e. the so-called *Sotho languages* (Sepedi, Sesotho, Setswana) together and the Nguni languages (isiZulu and isiXhosa) together as has been done in MLD and CMD. (The other four official African languages of South Africa Siswati, isiNdebele, Tshivenda and Xitsonga are not included in MID.) A specific colour has been allocated to each language, e.g. purple for isiZulu, yellow for Setswana, green for isiXhosa, etc. which is applied throughout in a consistent manner. The user guide is also presented in these seven languages. This is followed by 13 theme sections, i.e. My body, Family and friends, Clothing, My home, Our community, Transport, Communication, Economy, Nature and us, Food and drink, Our spare time, South Africa and Useful words. The 13th

theme, Useful words, is not an illustrated one but consists of 10 categories, i.e. 'to be', 'to have', Articles, Pronouns, Question words, Prepositions, Conjunctions, Adverbs etc., Adjectives and Verbs, in the seven languages. Each of these 13 themes is presented in the seven languages and each theme is subdivided into 10 sub-themes, e.g. for Theme 1: The body, Face, Upper body, Lower body, Hand, Foot, The skin, Inside the body, Feelings and Phrases. Thus 13 themes x 10 sub-themes x 7 languages = 91 topics. The 13 numbered themes are followed by an index where the words used in the thematic sections are given in alphabetical order with a page number referring to the illustration and set of words, e.g. *air traffic control tower* 127 with reference address a picture of the tower with translations into the seven languages on this page. The suggested strategy given in the user's guide is to look up the word in the alphabetical index to find the page reference number and then to turn to the given page to find the word in the seven languages. So, for example departing from the index sections *mouth, nose, nostril, leihlo, molomo, nko, impumlo, ukhophe, umlomo*, all refer to page 16 where the picture of a face is given indicating the mouth, nose, nostril, etc. with translations in all seven languages as indicated in Figure 1.

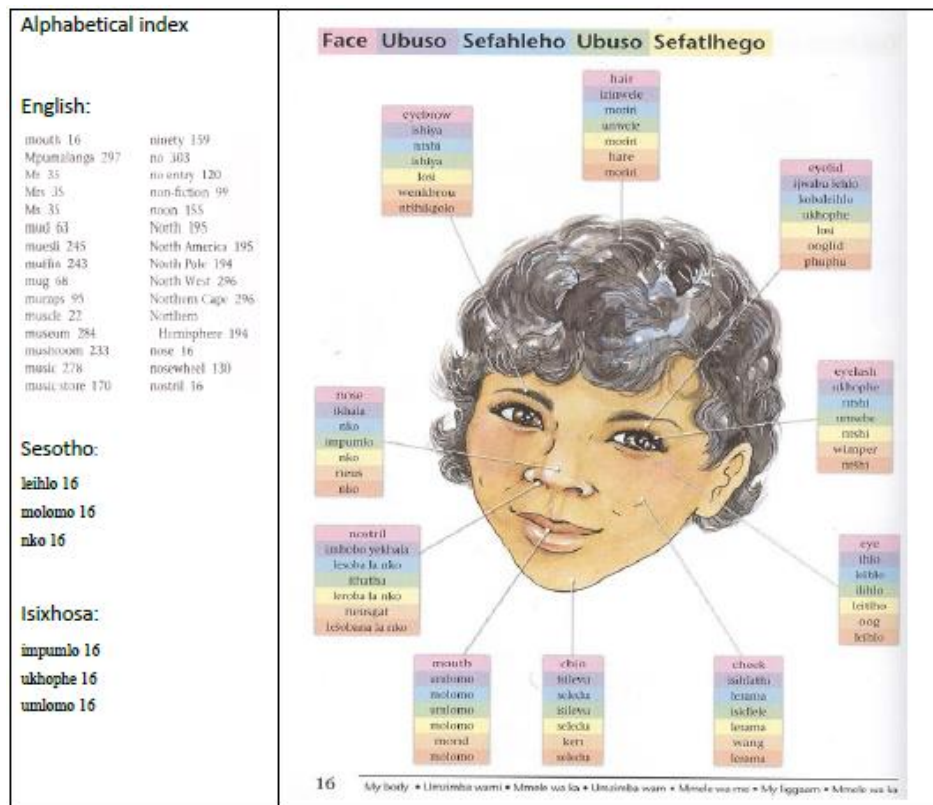


Figure 1: Index and thematic pictorial treatment of Face in MID

MLD's cover page indicates that it covers English, Afrikaans, "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu". In the table of contents it explains the colour keys for the different languages, how to use this book, know your grammar, visual dictionary followed by a section with English lemmas as the source language (L1) followed by Afrikaans, "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu", then a section Afrikaans as L1 followed by English, "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu" and then sections where the African languages lemmas in turn are given as L1, i.e. "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu" with English and Afrikaans as target languages in each case. Consider the treatment of the stretch *abacus – accident* in Figure 2.

English	Afrikaans	N Sotho	Sesotho	Tswana	Xhosa	Zulu
abacus (n)	telraam	mmadiphetana	mmadibolwana	kgathatalama	isibali	i-abakhusi
abandon (v)	verlaat	-tlogela	-nyahlatsa	-tlogela	-ncama	-yeka
abattoir (n)	slagplaas	bohlabelo	selagapalo	bothabelo	isilarha	amadela
abdomen (n)	buik	mpa	mpa	mpa	isisu	isisu
abduct (v)	ontvoer	-thopa	-kwetela	-utswa	-thwala	-thwala
ability (n)	bekwaamheid	bokgoni	bokgoni	kgoni	ubuchule	ukwazi
able (a)	bekwaam	-kgona	-kgona	-kgonang	-kwazi	-nokwazi
abnormal (a)	abnormaal	-sa tlwaelegago	-sa tlwaelehang	-sa tlwaegang	-ngaqhelekanga	-ngavamile
aboard (adv)	aan boord	ka gare ga	-palame	-palameng	-ngaphezulu	emkhunjini
abode (n)	woonplek	bodulo	bodulo	legae	indawo	isikhundla
abolish (v)	afskaf	-fediša	-fedisa	-khutisa	-bhangisa	-susa
aboriginal (n)	oorspronklike inwoner	semetafa	motala	monnafa	inzalwane	umuntu wondabu
abort (v)	aborteer		-nyopisa	-senyega	-phanza	-khipha isisu
abortion (n)	aborsie		nyopiso	tshenyegelo ya mpa	impanza	ukuphuma isisu
abortive (a)	misluk		-nyopiteng	-senyegetsweng	-phanzileyo	-gejile
about (adv)	rond	mo le mola	ka-	gomela	-ngokujikeleyo	nga-
about (prep)	oor		ka-	ka ga	nga-	phezu kwa-
above (prep)	bo	godimo ga	-hodima	mo godimo	-phezulu	enhla kwa-
abscess (n)	verswering	sekaku	seso	sekaku	igqitha	ithumba
abscess (n)	afwesigheid	tlhokego	boziyo	tlhokafalo	ubungabikho	ukungabikho
absent (a)	afwesig	-hlokega	-ba siyo	-se yong	-ngekhoyo	-ngekho
absent-minded (a)	ingedagte	-ya le megopolo	-bua ka pelo	-bua ka pelo	-zungubala	-alukile
absolute (a)	absoluut		-hle	tota	kanye	ngempela
absolutely (adv)	absoluut	ruri	ho hang	tota	-ngenene	ngempela
absorb (v)	absorbeer	-hupa	-monya	-monyela	-funxa	-munca
abstain (v)	onthou	-ila	-sesefa	-ila	-yeka	-yeka
abstract (a)	afgetrokke		-sa bonweng	-kgopolo	-ngaphathekiyo	-okusengqondweni
absurd (a)	onsinnig	-segišago	-hloka kelelo	mo go se nang	-nobuhiba	-ngasile
abundance (n)	oorloed	bontši	nala	monono	ubuninzi	isibhidi
abundant (a)	oorvloedig	-ntši	-ngata	-tlatlalo	-ninzi	-xhaphakile
abuse (n)	misbruik		tlatlapo	tiriso e e sa siamang	ukusebenzisa kakubi	ukoniwa
abuse (v)	misbruik	-senya	-tlatlapa	-dirisa e seng ka fa siamang	-sebenzisa kakubi	-ona
acacia (n)	akasia	mohlare wa akasia	leoka	mmafu	umnga	umkhamba
academy (n)	akademie	aketemi	akhademi	maithuto	isikolo senkcubeko	imfundo ephakeme
accelerate (v)	versnel	-fa makhura	-akofisa	-okofatsa	-khawulezisa	-sheshisa
accelerator (n)	brandstofpedaal	sefamakhura	seakofisi	seokofatso	umcephe	isipuni
accent (n)	aksent	tobeletšo	tokodiso	kgateletšo	uhlobo lokuthetha	isigcizelelo
accept (v)	aanvaar	-amogela	-amohela	-amogela	-amkela	-amukela
acceptable (a)	aanvaarbaar	-amogelega	-amohelaha	-amogelesegang	-amkerekileyo	-amukelekayo
acceptance (n)	aanvaarding	kamogelo	kamohelo	kamogelo	uwamkelo	ukwamukela
accident (n)	ongeluk	kotsi	kotsi	kotsi	ingozi	ingozi

Figure 2: Layout and treatment of *abacus – accident* in MLD

Finally, MLD has a section devoted to Phrases with subthemes such as Time, Household, Travel, etc.

In contrast to MID a better grouping of the languages, is given, i.e. Sotho languages together and Nguni languages together. Under "Visual dictionary" MLD lists The human body, The house, The Office, The Car and Sport as main themes with subthemes, e.g. under Sport soccer, rugby and cricket are listed. Its stated aim is to help the user to achieve "a basic level of understanding in the seven major languages of South Africa" by means of (a) a visual dictionary, (b) a 5,000-word dictionary that translates English and Afrikaans into five African languages.

CMD does not offer any thematic sections but it covers English, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, Afrikaans.

A

English	Xhosa	Zulu	Northern Sotho	Southern Sotho	Tswana	Afrikaans
abandon	-lahla	-yeka	-tlogela	-nyahlatsa	-tlogêla	verlaat
abate	-damba	-nciphisa	-okobala	-kokobela	-reba	bedaar
abdomen	isisu	isisu	mpa	mpa	mpa	maag
abduct	ukuba umntu	-thwala	-thopa	-kwetela	-utswa	skaak
abide	-hlala	-hlala	-dula	-dula	-nna	bly
ablaze	-vutha	-vutha	-tuka	-tukang	tukang	in vlamme
abode	ikhaya	ikhaya	bodulo	bodulo	legaë	verblyfplek
above	ngaphezulu	phezu kwa-	godimo	hodimo	godimo	bo
abridge	-finyeza	-thothanisa	-khutsofatša	-kgutshufatsa	-khutswafatsa	verkort
absent	-ngekho	-ngekho	-hlokega	-ba siyo	-se yōng	afwesig
absent, to be	ukungabukho	ukuba ngabikho	go hlokega	ho ba siyo	-tlhōkēgwa, go	afwesig te wees, om
accede	-vuma	-vuma	-dumela	-dumela	-dumela	toege
accident	ingozi	ingozi	kotsi	kotsi	kōtsi	ongeluk

Figure 3: Layout and treatment of *abandon* – *accident* in CMD

The dictionary consists of two sides, i.e. English followed by "Xhosa" "Zulu", "Northern Sotho" "Southern Sotho", "Tswana" and Afrikaans and a reverse side printed upside down (i.e. the user has to flip over the dictionary and start from the other end for Afrikaans as L1 followed by "Xhosa", "Zoeloe", "Noord-Sotho", "Suid-Sotho", "Tswana" and English. In the introduction section of CMD this macrostructural arrangement is praised as "English and Afrikaans word lists bound into one handy volume, with their inverse juxtapositions providing each of the books with equal recognition and prominence". However, unlike MLD, this "equality" is not carried through for the African languages since they are not listed as L1 nor are indexes provided to look up African language words as has been done in MID, see Figure 1 column 1 as an example of Sesotho and isiXhosa indexed lemmas with reference to the page where treatment is offered. No thematic pages or user guides are given but an approximately two-page mini-grammar is attempted for each of the African languages treated. Tables for days of the week, months of the year and numbers are given for English as L1 translated into the other languages. In the mini-grammars an effort is made to give grammatical information for "Xhosa", "Zulu", "Northern Sotho", "Southern Sotho" and "Tswana". It is clear that no attempt was

made to bring the data on a par for the other languages. Different issues and illustrations receive attention for the different languages. So, for example, for isiZulu, in addition to pronunciation guidance, detailed attention is given to the noun class system, the concordial system, pronouns and the verb but for isiXhosa, Sepedi, Sesotho, Setswana such important guidance in terms of the nominal class system is absent — the entire sections deal only with pronunciation.

OJPD's user's guide under the heading "to the teacher", states the fact that the dictionary gives the meanings of more than 1,500 words and highlights five characteristics of the dictionary, i.e. (1) the alphabet appears at the top of each page, (2) the defining vocabulary is controlled, (3) each word is used in an example sentence, (4) pictures help children to understand the meaning and (5) mother tongue translations help with meaning. The A-Z section with English as L1 represents more than 90% of the space allocation in the dictionary. Relatively few schematic illustrations are given, approximately one on every second page in the A-Z section. Five themes are presented in the form of schematic illustrations at the end of the dictionary, i.e. A Girl, A Boy, A Classroom, A City Street and A Kitchen, but these are given only in English. The user has to look up the English word in the alphabetical section. Tables are given for Days of the week and Months of the year with translations in "North Sotho", "South Sotho", Setswana and Afrikaans. This is followed by a Numbers theme 1 (one) to 50 (fifty) and 1st (first) to 50th (fiftieth) but not translated into the four target languages. Finally another range of theme pages of schematic illustrations of Animals and young animals, Inserts, Fruit and vegetables is given, also in English only.

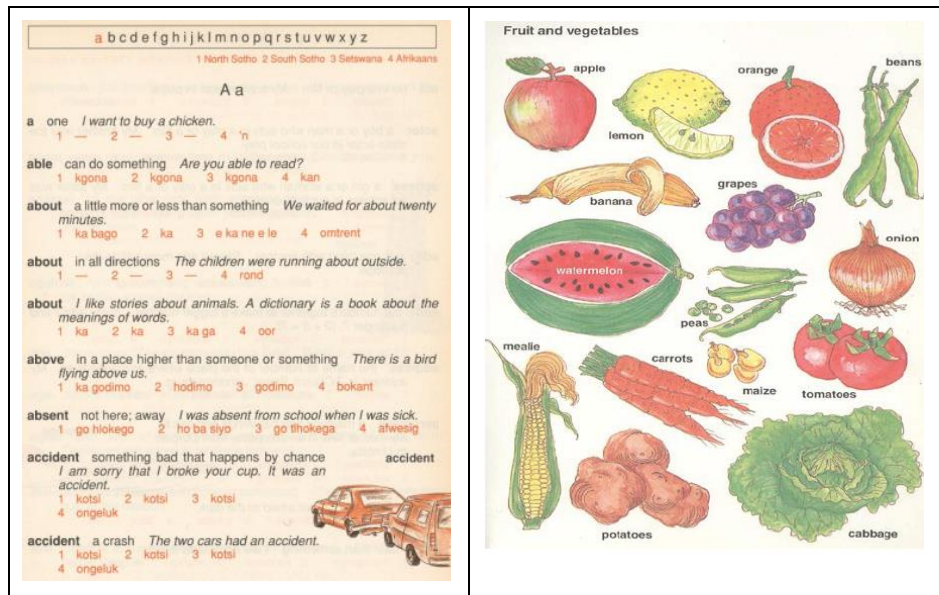


Figure 4: An extract from the alphabetical and thematic sections in OJPD

Minor mistakes noticed is that the English words for *days* and *months* are not correctly aligned with the translation table and that numbers are not translated, cf. Figure 5.

Days of the week					Numbers			
	North Sotho	South Sotho	Setswana	Afrikaans				
Sunday	Lamorena	Sontaha	Letsatsi la Tshipi	Sondag	1	one	1st	first
Monday	Mosupologo	Mantaha	Mosupologo	Maandag	2	two	2nd	second
Tuesday	Labobedi	Labobedi	Labobedi	Dinsdag	3	three	3rd	third
Wednesday	Laboraro	Laboraro	Laboraro	Woensdag	4	four	4th	fourth
Thursday	Labone	Labone	Labone	Donderdag				
Friday	Labohlano	Labohlano	Labothlano	Vrydag				
Saturday	Mokibelo	Moqebelo	Lamathatso	Saterdag				

Figure 5: Days of the week and Numbers in OJPD

OJPD is inconsistent in terms of reference to the Sotho languages. For Sepedi and Sesotho it uses the English terms North Sotho and South Sotho (which are linguistically speaking questionable) but refers to Setswana instead of Tswana which would linguistically be aligned with the terms North(ern) and South(ern) Sotho.

3.2 Coverage of lexical items

It is a reasonable expectation of the user to find the word (s)he is looking for in the dictionary.

A good dictionary is one in which you can find the thing you are looking for preferably in the very first place you look. (Haas 1962)

Dictionaries for which the lemma lists are restricted to a relatively small number, e.g. 5,000 lemmas, naturally focus on the most common words of the language. The question is whether dictionaries such as MID, MLD, CMD and OJPD reflect a reasonable coverage of common words in the language and if they provide the required items in the comment on form and comment on semantics for the user. Formulated differently, for most users to "find the thing" in terms of Haas, simply means to find the lemma in the dictionary and especially the meaning of the word represented by the lemma sign, in a user-friendly way.

In MID the alphabetical indexes reflect the number of lemmas treated in the dictionary. For each language approximately 3000 words are listed with cross-reference to the page number where a picture is given and the translation equivalents in the other languages. cf. Figure 1 above.

MLD has more than 400 pages for the alphabetical section. This approach is user-friendly in the sense that the user can use any of the seven languages as direct access for lookup. There is no need, as in the case of MID, to consult an

index list in order to be cross-referred to the correct L1 lemma and page number to look it up, or as in the case of CMD only being able to look up Afrikaans and English words.

English/Afrikaans/Northern Sotho/Sesotho/Tswana/Xhosa/Zulu	50
Afrikaans/English/Northern Sotho/Sesotho/Tswana/Xhosa/Zulu	137
Northern Sotho/English/Afrikaans	225
Sesotho/English/Afrikaans	273
Tswana/English/Afrikaans	321
Xhosa/English/Afrikaans	369
Zulu/English/Afrikaans	417

Figure 6: The scope of translation equivalents given and page numbers for each L1 language in MLD

However, it drastically reduces the available dictionary space per section to approximately 86 pages for English; 87 for Afrikaans, 47 for "Northern Sotho", 47 for Sesotho, 48 for "Tswana", 47 for "Xhosa", and 47 for "Zulu". Furthermore, even though the African languages are in turn given lemma status as L1, only English and Afrikaans translation equivalents are given and no equivalents in the other African languages. Not giving translation equivalents in the other African languages enables the use of double columns per page which results in 50% space saving, but it means that the treatment of African languages as L1 is not on a par with English and Afrikaans.

Dividing the dictionary into seven alphabetical sections as in Figure 6 reduces the potential dictionary space for lemmas from 24,000 (400 pages x 60) lemmas per language to approximately 5,000 lemmas per language in order to give each language the opportunity to be lemmatised as L1. The question is whether an index system as in MID with cross reference to an English as L1 section would not have been a better option. This would be less user-friendly because Afrikaans and African language words will have to be looked up via alphabetical index lists but would enable a treated lemma list of e.g. 20,000 English L1 lemmas instead of the current 5,000 or less per language. Alphabetical index lists also require dictionary space but approximately six times less because indexed words can be presented in six to seven columns per page as was done in MID.

In CMD approximately 60 lemmas are given per page and a total of 155 pages could in principle render a lemma list of approximately 9,000. However, due to the fact that English and Afrikaans are presented in two separate sections as L1, this potential is divided by two i.e. the lemma list for English is reduced to circa 4,500.

For OJPD approximately 180 pages x 10 lemmas per page equals 1800. This is fairly limited and it is unlikely that the user will find the words he/she is looking for.

In all of these dictionaries the number of lemmas can be regarded as very limited and an attempt has been made to see to what extent they succeeded in lemmatising the most common words of the language. The 136 most frequently used words in the alphabetical stretch A, indicated with a three-star frequency band in the *Macmillan English Dictionary* (MED) was used as a measuring instrument. These words are compared to the most frequently used words in the Pretoria English Internet Corpus (PEIC) and to the lemma lists of MLD, OJPD, CMD and MID in Table 2.

MED ***	PEIC	MLD	OJPD	CMD	MID					
and	32101	✓	✓	✓	✓					
as	7794		✓	✓	✓					
at	6384	✓	✓	✓	✓					
all	3549	✓	✓	✓						
an	3060		✓							
any	1659	✓	✓							
about	1569	✓	✓		✓					
after	1463	✓	✓	✓	✓					
army	1399	✓		✓						
against	1072	✓	✓		✓					
also	684	✓	✓							
among	560		✓	✓						
again	519	✓	✓							
away	483		✓	✓						
always	448	✓	✓							
another	442	✓	✓							
almost	405	✓	✓	✓						
already	395		✓							
above	387	✓	✓	✓	✓					
able	342	✓	✓							
age	341	✓	✓	✓						
afterwards	333	✓		✓						
August	304	✓	✓		✓					
attack	303	✓	✓							
ancient	289	✓								
according	278									
authority	276	✓								
along	259	✓	✓		✓					
act	253		✓	✓	✓					
art	243	✓								
anything	217		✓							
account	210	✓		✓						
alone	207	✓	✓	✓						
answer	183	✓	✓	✓						
around	174	✓			✓					
attention	165		✓							
across	157	✓	✓	✓	✓					
ask	147	✓	✓	✓						
ago	134		✓							
admit	129	✓								
attempt	122	✓		✓						
although	121	✓		✓						
arrival	121	✓			✓					
April	119	✓	✓		✓					
airport	110	✓	✓		✓					
action	105									
afternoon	105	✓	✓		✓					
air	100	✓	✓	✓						
assistance	94	✓								
add	88	✓	✓	✓						
assembly	88	✓								
afraid	85	✓	✓	✓	✓					
amount	85		✓							
author	85	✓			✓	✓				
ability	84	✓								
appearance	83	✓								
approach	83	✓			✓					
altogether	81									
allow	79	✓	✓							
appear	79	✓			✓					
argument	79	✓	✓							
active	73	✓								
arm	66	✓	✓	✓	✓					
absence	63	✓								
accept	62	✓								
affair	59	✓								
ahead	59	✓								
absolutely	57	✓								
agree	55	✓	✓							
alive	54	✓	✓							
angry	52	✓	✓	✓	✓					
artist	49									✓
assume	47									
anybody	46	✓	✓							
apply	43	✓								
article	43	✓								
accident	42	✓	✓	✓						
avoid	41	✓								
animal	40	✓								
address	39	✓	✓	✓						
attend	38									
apparent	37									
arrangement	37	✓								
afford	36	✓								
apparently	36									
adopt	35	✓								
audience	35									✓
actually	33	✓								
apart	33									
appeal	32	✓								
assist	32	✓			✓					
association	31									
annual	27									
application	27	✓			✓					
appointment	27	✓	✓							

available	27	✓				
actual	26					
abroad	22					
anywhere	22	✓	✓			
arise	22	✓				
arrive	22	✓	✓	✓		
aspect	22					
adequate	19	✓				
agreement	19	✓				
affect	18	✓				
angle	18	✓				
arrange	18	✓				
attitude	18	✓				
access	17					
activity	16	✓				
appoint	15	✓		✓		
agent	14	✓		✓		
approve	14	✓				
approval	13	✓				
attract	13	✓				
aunt	13	✓	✓	✓	✓	✓
average	13	✓				
accuse	11	✓			✓	
aim	11	✓	✓			
associate	11					
alternative	8					
area	8	✓				
autumn	8	✓	✓	✓	✓	✓
attractive	7	✓				
achieve	6	✓				
argue	6	✓	✓	✓		
anyone	5				✓	
agency	4					
academic	3					
award	3					
aware	3					
achievement	2	✓				
actor	2	✓	✓	✓	✓	✓
analysis	2	✓				
anyway	2					
adult	1	✓			✓	✓

Table 2: MED three-star words compared to the PEIC, MLD, OJPD, CMD and MID

There are 136 three-star words in the alphabetic stretch A in MED. MLD has 102 (75.0%), OJPD 54 (39.7%), CMD 42 (30.9%) and MID 25 (18.4%) of these MED three-star words as treated lemmas. In PEIC the top 5,000 words occur with a frequency of 200 or more. 33 of these words are also three-star words in MED (indicated in bold in Table 2) and were also compared with the lemma lists of MLD, OJPD, CMD and MID. Of these 33 words OJPD lemmatised and treated 26 (78.8%), MLD 25 (75.8%), CMD 15 (45.5%) and MID 9 (27.3%).

Alphabetic stretch A	MLD	OJPD	CMD	MID
MED *** words	75.0%	39.7%	30.9%	18.4%
33 PEIC	75.8%	78.8%	45.5%	27.3%
Size of dictionary	495	198	150	365

Table 3: Size of MLD, OJPD, CMD and MID in relation to the overlap with MED three-star words and the PEIC

These percentages of overlap of MLD, OJPD, CMD and MID with MED three-star words and the PEIC are schematically illustrated in Figure 7.

From Figure 7 it is clear that none of the dictionaries captured more than 80% of the most frequently used words and e.g. that MID captured less than 30% of such words.

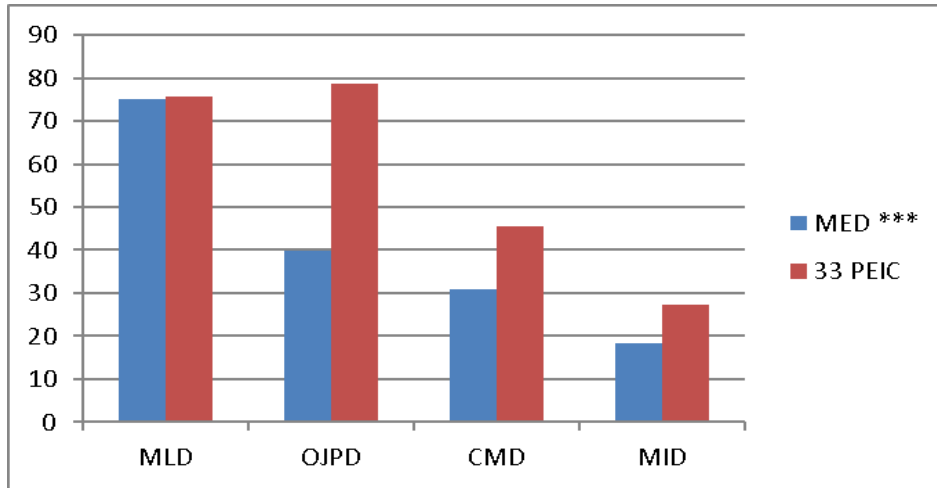


Figure 7: Size of MLD, OJPD, CMD and MID in relation to the overlap with MED three-star words and the PEIC

The percentages in Table 3 were then calculated as a percentage of the size of each dictionary and are illustrated in Figure 8.

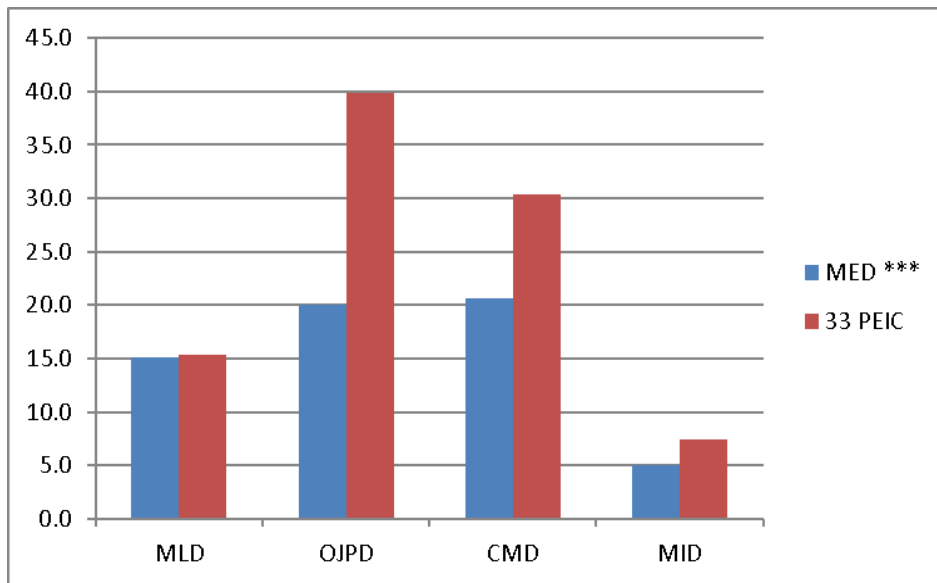


Figure 8: Percentage of top frequencies in MLD, OJPD, CMD and MID in relation to the size of the dictionaries

A slightly different situation emerged. OJPD and CMD scored higher than the other dictionaries because they achieved their percentages given in Table 3 in dictionaries consisting of less than 200 pages. MLD however, although it captured 75% of the frequently used words, scored worse because it is a much bigger dictionary, i.e. more than twice the size of OJPD and CMD. MID's scores remained low, i.e. a low percentage of frequently used words captured, as well as the fact that it is a relatively bigger dictionary than OJPD and CMD. Formulated differently: the four dictionaries did not do well in terms of capturing the most frequently used words, especially so for the two bigger dictionaries, MLD and MID. For example, the omission of common words such as *as, an, among, away, already* in MLD is less satisfactory for a relatively big dictionary, i.e. one consisting of 495 pages. Thus one would expect from the relatively bigger dictionaries, i.e. MLD and MID to reflect more of the most frequently used words.

4. Microstructural considerations

4.1 Comment on form

The root of the perceived deficiencies of multilingual dictionaries on the microstructural level is what could be referred to as impoverished or skeleton dictionary articles. Impoverished dictionary articles are an almost unavoidable result of the multilingual model because in dictionaries treating up to *seven* languages, there is hardly any space available beyond the listing of a single translation equivalent for each lemma. As indicated in Figure 3, on a single page in CMD more than 60 lemmas are treated for six languages, i.e. presenting approximately 400 articles per page. In CMD and MLD seven columns have to be fitted in on every page to reflect the seven languages thus limiting the space for the dictionary article to a default of 20 characters with the occasional use of a second line.

Such severe space limitation results in a number of deficiencies and even a combination of deficiencies ranging from mere omissions and inconsistencies to mistakes and the ultimate deficiency: misguidance of the user.

First, no other item besides the lemma is given in the comment on form. In some cases, comment on form is given in an inconsistent way or an incorrect label is given.

CMD employs a strategy of inserting a comma and guiding word(s) after the lemma. These guiding words which follow the comma are intended to inform the user that they should be used in front of the lemma. (CMD's use of leadword versus main word is not clear).

In order to facilitate arranging the English and Afrikaans leadwords in alphabetical order, for quick and easy reference, a comma (,) has been inserted — where thought necessary — after a leadword to indicate that the words following such comma must be utilised in front of the leadword to arrive at the correct shade of meaning of the main word. (CMD: v)

English	Xhosa	Zulu	Northern Sotho	Southern Sotho	Tswana	Afrikaans
abandon	-lahla	-yeka	-tlogela	-nyahlatsa	-tlogêla	verlaat
abate	-damba	-nciphisa	-okobala	-kokobela	-reba	bedaar
abdomen	isisu	isisu	mpa	mpa	mpa	maag
abduct	ukuba umntu	-thwala	-thopa	-kwetela	-utswa	skaak
abide	-hlala	-hlala	-dula	-dula	-nna	bly
ablaze	-vutha	-vutha	-tuka	-tukang	-tukang	in vlamme
abode	ikhaya	ikhaya	bodulo	bodulo	legaê	verblyfplek
above	ngaphezulu	phezu kwa-	godimo	hodimo	godimo	bo
abridge	-finyeza	-thoanisa	-khutsofatša	-kgutshufatsa	-khutswafatsa	verkort
absent	-ngekho	-ngekho	-hlokega	-ba siyo	-se yông	afwesig
absent, to be	ukungabukho	ukuba ngabikho	go hlokega	ho ba siyo	-tlhókêgwa, go	afwesig te wees, om
accede	-vuma	-vuma	-dumela	-dumela	-dumela	toege
accident	ingozi	ingozi	kotsi	kotsi	kôtsi	ongeluk
accompany, to	ukukhapha	-phelezela	-felegetša	ho felehetsa	-bulêdisa, go	vergesel
...						
account, to give an	ukubalisela	ukubalisisa	go hlaloša	ho hlalosa	supa tlôtlô, go	verslag te gee, om
account, give an	-balisela	-balisisa	-hlaloša	-hlalosa	supa tlôtlô	verslag gee
account, an	umlandu	i-akhawunti	tšhupamolato	akhaonte	tshupatlôtlô	'n rekening
accumulate (n)	-fumba	-andisa	-kgoboketša	-bokella	-koêla	ophoop
accuse	-tyhola	-beka icala	-bega	-qosa	latofatsa	beskuldig
accused, the	ummangalelwa	ummangalelw	mmeğiwa	moqosuwa	mosêkisiwa	beskuldigde, die
...						
acid (n)	iasidi	i-asidi	sedilana	esiti	êsêtê	suur (nw)
...						
act, an	isenzo	isenzo	tiro	ketso	tirô	'n daad
act, to	ukwenza	ukwenza	-dira	ho etsa	-dira, go	doen, om te
...						
afraid, be	-oyika	-esaba	-boifa	-tshaba	-bôifa	bang wees

Table 4: A selection of treated lemmas in CMD

The indication of part of speech, or the absence thereof, is problematic. For example, for the first ten lemmas, *abandon* – *absent* in Table 4, no part of speech guidance is given. The verb *accumulate* is incorrectly indicated as a noun by means of a part of speech label (*n*). There are three lemmas for *account*, all pertaining to a noun. One is followed by a comma plus "an" which could be interpreted as indicating its part of speech as a noun and which is indeed in line with all the nominals given as translation equivalents.

The approach of inserting a comma followed by so-called guide word(s) leads to many inaccuracies, inconsistencies, uncertainties and misinterpretations. Most users are likely to regard the use of the *,an* and *,to* conventions as implicit indicators of part of speech as nominal versus verbal, e.g. in the articles of *act, an* versus *act, to* which is correct for distinguishing between the nominal and verbal forms respectively. The pertinent statement that such words which follow the comma, e.g. *an, to, be*, etc. "must be utilised" in front of the lemma could even mislead the inexperienced user to believe that (*to*) *give an* must always be used in front of *account*, which is not the case.

For all the African languages the infinitive prefix is given in front of the

verb stem, but the comma convention is used in the Setswana translation equivalents. The treatment of *act*, *to* as reference to the infinitive form of the verb is inconsistent. The Setswana and Afrikaans equivalents follow the convention for English by giving *-dira* for Setswana preceded by a hyphen to indicate its status as a verb stem, followed by a comma and the infinitive prefix *go*, and Afrikaans *doen*, followed by a comma and the infinitive *om te* which is the equivalent of English *to*. The isiXhosa, isiZulu and Sesotho, however, do not follow the convention but give the infinitive forms as they would actually occur in text and speech i.e. with the infinitive prefixes preceding the verb. For Sepedi only the verb stem is given, with no infinitive guidance. To be consistent with the other languages it should have been either *go dira* to reflect actual occurrence or *-dira, go* to be in line with the English convention. Once again the user is ill-served because (s)he might conclude that *to act* is *go dira* in Setswana but *dira* in Sepedi whilst the two forms are exactly the same in Sepedi and Setswana.

Nouns are inconsistently handled as (a) without any part of speech indication, e.g. *accident* (b) with a part of speech label in brackets (*n*), e.g. *acid*, or (c) given with lead words such as *an*, or *the* as for *accused*. No label (*v*) is used for verbs.

CMD also claims that "the format adopted is straightforward and easy to understand; where the need arises for special connotations, a directive word is given between brackets, to indicate the particular shade of meaning associated with the particular leadword". This is useful in terms of homonym and sense distinction for instances such as *bow (in shooting)*, *gargle (mouth wash)*, *gargle (throat wash)*, etc. but less effective for Afrikaans (*language*) or alphabet (*letters of*) where the label does not make much sense, neither does it contribute to better understanding of the meaning of the word represented by the lemma or to delimitation of the range of application.

In CMD circumflexes are only indicated for Setswana. This creates the incorrect impression that, e.g. Sepedi and Setswana have the same word for *begin*, but that the pronunciation differs, i.e. *thoma* versus *thôma*. Circumflexes should be consistently and correctly used especially in cases where different meanings are conveyed, e.g. Sepedi *bola* 'speak' versus *bôla* 'rot'. Likewise, no tonal indication is given. Tonal indication in these dictionaries is at least required to distinguish between homonyms which have the same spelling but differ in tone such as *tlala* 'hunger' and *tlâla* 'become full'.

CMD also claims that using the dictionary does not require specialised knowledge of the languages covered by the dictionary because "all the troublesome grammatical complexities generally associated with the formation of the words have already been dealt with and incorporated into the words". Such "complexities" are indeed problematic for the user for lemma identification in order to look up isiZulu, isiXhosa, Siswati and isiNdebele words, cf. Prinsloo (2011) for a detailed discussion. In MLD the user has to identify the stem of the word before isiZulu and isiXhosa words can be looked up in the sections Xhosa/English/Afrikaans and Zulu/English/Afrikaans where these languages

are given as L1. CMD, however, does not provide lookup of isiZulu and isiXhosa words as L1, so CMD can hardly claim any credit for user-friendliness in terms of "not requiring specialised knowledge" or that "grammatical complexities ... have been dealt with" if it merely provides words from these languages as translation equivalents.

Finally spelling errors occur which is a serious mistake for any dictionary to make since dictionaries are often used to check spelling. So, for example, OJPD misspelled frequently used Sepedi words such as *woman* (**masadi* instead of *mosadi*), *wife* (**mogatsa* instead of *mogatšā*) found within a few randomly selected pages in the alphabetical section and *Monday* and *February* (**Mosupologo* instead of *Mošupologo* and **Dobokwane* instead of *Dibokwane*), respectively in the tables for Days of the week and Months of the year.

No pronunciation guidance is given in these dictionaries.

4.2 Comment on semantics

MLD admits that "shades of meaning and interpretations specific to the culture of one language are often lost in another" and continues that "this problem is obviously compounded when a word is translated through six other languages." (MLD: 6). It is not clear what is exactly meant by "shades of meaning". If it is taken to mean lack of minor differences in e.g. register or range of application it could be tolerable for a dictionary treating many languages simultaneously. If it refers to distinction between different senses of a word it constitutes a major deficiency, if it also refers to lack of distinction between homonyms it is an even bigger shortcoming of the dictionary. It will be illustrated in this section that the deficiencies in the dictionaries under discussion indeed go far beyond minor differences of meaning.

In consulting a dictionary the user expects a reasonable number of items in the *comment on semantics* in the article e.g. at least items giving basic data of the word represented by the lemma such as meaning, examples of use, collocations, different senses, idioms, etc. So, for example, treatment of the lemma *interest* in MID is limited to the single-word translations *inzalo*, *phaello*, *inzala*, *morokotso*, *rente* and *tswalo* in isiZulu, Sesotho, isiXhosa, Setswana, Afrikaans and Sepedi respectively. Sense distinction between *interest* (to be interested in something) and *interest* (profit, growth on investment) is not given. Lack of *sense distinction*, aggravated by inconsistent treatment and misguidance of the user is for example evident from the treatment of the lemma *with* (prep.) in MLD.

MLD

English	Afrikaans	Sepedi	Sesotho	Setswana	IsiXhosa	IsiZulu
with (prep)	met	le	le	ka	nga-	nga-

Table 5: MLD's treatment of *with* (prep)

Here only the sense of "together with/accompanying" is given for Sepedi and Sesotho and not the equally important sense of "with something, i.e. as an instrument" — thus an instance of an incomplete translation equivalent paradigm. The exact opposite applies for Setswana, isiXhosa and isiZulu in Table 5. *Ka* and *nga-* means "with something" but the translation equivalent for "together with" i.e. Setswana *le* and isiXhosa and isiZulu *na* is missing. This is correct for examples such as *O sepela le mosadi* 'He walks with the woman' but fails to add the other translation equivalent *ka* indicating with (an instrument), e.g. *O sepela ka maoto* 'He walks with (his) feet'. The user runs the risk of creating incorrect sentences such as **O sepela le maoto* 'he walks with (his) feet' in Sepedi or **Ukhuluma ngaye* 'he talks to her' in isiZulu. Although the translation equivalent paradigms in Table 5 are incomplete, at least the translations are correct. Even more serious mismatches are instances where homonyms (words with same spelling but different meanings) are confused in the treatment for the different languages. So, for example, is the lemma *might* (n) = power/strength in CMD translated for its nominal meaning in isiZulu and isiXhosa as *amandla* and in Sepedi and Sesotho as *maatla* and *matla* respectively but in Setswana as the modal verb *ka* 'can/may/might' instead of the nominal meaning *maatla*. The inexperienced user can incorrectly conclude that *ka* is the word for power in Setswana and produce **incorrect phrases such as **o tshwanetse go nna le ka* 'you should have power/strength'**.

Much more problematic for the multilingual model are instances of what Prinsloo and Gouws (2006) call grammatical divergence. An underlying detrimental characteristic of multilingual dictionaries is a "one fits all" approach and assumption. MID's treatment of pronouns results in numerous mistakes, inconsistencies and misguidance of the user.

English	IsiZulu	Sesotho	IsiXhosa	Setswana	Afrikaans	Sepedi
Pronouns:	Izabizwana:	Maemedi:	Izimelabizo:	Leemedi:	Voornaamwoorde:	Mainaina:
I	(mina) ngi-	nna	ndi-	nna	ek	nna
you	(wena) u-	wena	wena/u-	wena	jy	wena
he/she/it	(yena) u-/ (yena) u-/ (yona) i-	yena/yena/ yona	yena/yena/ yona	ena/ena/yona	hy/sy/dit	yena/yena/ yona
we	(thina) si-	re	si-	rona	ons	re
you	(nina) ni-	o	wena	wena	julle	le
they	(bona) ba-	ba	bona	bona	hulle	ba
me	mina	nna	mna	nna	ek	nna/ya ka
you	nawe	wena	wena	wena	jy/jou	wena
him/her/it	yena/yona	yena/yena/yona	yena/yena/yona	ena/ena/yona	hom/haar/dit	yena/yena/yona
us	thina	rona	thina/sithi	rona	ons	rena
you	wena	wena	wena/u-	wena	julle	wena
them	bona	bona	bona	bona	hulle	bona

Figure 9: Pronouns in MID: 303

This single table is fraught with different types of errors. For *I*, both the pronoun *mina* and the subject concord *ngi-*, which has a secondary pronominal function, is given for isiZulu but for Sesotho, Setswana and Sepedi only the pronoun *nna* and not the pronominal subject concord *ke-* is given. For isiXhosa only the subject concord *ndi-*, and not the pronoun, is given. The same information is offered for *you*, but now both the pronoun *wena* and the subject concord *u-* are given for isiXhosa. The same inconsistencies prevail for *we*, *you*, and *they*. In the case of *me* the equivalent *ya ka* is given which refers to *mine* (a possessive pronoun: *ya* = of, *ka* = me) for Sepedi but not considered for the other languages.

In the case of *he/she/it* use of the slash "/" convention is unclear. The presentation in Figure 9 gives the impression that *yena* is applicable to all cases for *he/she* and that *yona* for all possible referents of *it*. This is insufficient guidance for the African languages where third persons are subdivided into numerous noun classes resulting in different words for *he/she* and *it*.

Subject concords (Sc.); object concords (Oc.); demonstratives (Dem.); possessive concords (Poss.); emphatic pronouns (Ep.) and quantitative pronouns (Qp.).

Person or noun class	Example	Sc.	Oc.	Dem.	Poss.	Ep.	Qp.
1st Person singular	nna 'I'	ke	n-				
1st Person plural	rena 'we'	re	re				
2nd Person sing.	wena 'you' (singular)	o	go				
2nd Person plural	lena 'you' (plural)	le	le				
Class 1	monna 'man'	o/a	mo	yo	wa	yena	yohle
Class 2	banna 'men'	ba	ba	ba	ba	bona	bohle
Class 3	molato 'trouble, problem'	o	o	wo	wa	wona	wohle
Class 4	melato 'problems'	e	e	ye	ya	yona	yohle
Class 5	lesogana 'young man'	le	le	le	la	lona	lohle
Class 6	masogana 'young men'	a	a	a	a	ona	ohle
Class 7	selo 'object, thing'	se	se	se	sa	sona	sohle
Class 8	dilo 'objects, things'	di	di	tše	tša	tšona	tšohle
Class 9	ntlo 'hut'	e	e	ye	ya	yona	yohle
Class 10	dintlo 'huts'	di	di	tše	tša	tšona	tšohle
Class 14	bogobe 'porridge'	bo	bo	bjo	bja	bjona	bjohle
Class 15	go reka 'to buy'	go	go		ga		
Class 16	fase 'below'			fa			
Class 17	godimo 'above'	go	go		ga	gona	gohle
Class 18	morago 'behind'			mo			

Table 6: Sepedi noun classes, concords and pronouns

He/she/it is pronominalized depending on the noun class as *yena*, *wona*, *lona*, *sona*, or *yona* in Sepedi with similar paradigms for the other African languages.

On the same page (303) *he/she/it is* is translated in Sepedi as *o/o/e*. In the majority of cases *he/she/it is* is translated in Sepedi by the copulative particle *ke*, e.g. *he/she is a teacher*, '*ke morutiši*'. In following the information on this page the user could easily produce incorrect speech and text such as **(monna) o morutiši* 'he (the man) is a teacher'. The shortcomings are thus not limited to inconsistencies but guide the user to incorrect speech and text production.

The risk of the user being misguided by the above-mentioned shortcomings and inconsistencies is very high and refutes one of the basic principles of good lexicography, i.e. that the user should be guided by the dictionary not to make mistakes in especially text and speech production.

Finally the lack of examples of usage in these dictionaries is a significant shortcoming. Such examples help the user to understand the meaning of a word and to use it in text and speech production, and their value is not disputed in lexicography. For example, Kilgarriff et al. (2008) categorically state:

Users appreciate examples. If a dictionary entry includes an example which is a good match for the context in which the user has encountered a word, or for the context in which they want to use it, then the user generally gets what they want quickly and straightforwardly. Thus there is a case for including lots of examples, for lots of different contexts. Kilgarriff et al. (2008: 425)

Conclusion

In this article a critical evaluation of a number of South African multilingual dictionaries has been attempted. It is a well-known fact in lexicographic circles that no single dictionary can be everything for everyone. One should not be unreasonable in judging multilingual dictionaries from a single perspective, e.g. only on the number of treated lemmas while ignoring the many benefits and learning/reading potential of the thematically ordered sections. One also cannot expect detailed treatment of lemmas in each of the languages in a multilingual dictionary. However, as quoted from Haas (1962) the users expect to find the words they are looking for and therefore at least the most common words of the language should be included and correctly treated. As for any other dictionary the multilingual dictionary should also be a product of high lexicographic achievement. One cannot but come to the conclusion that the multilingual dictionaries used in this study are not products of high lexicographic achievement. In a sense it could be argued that multilingual dictionaries indeed try to be everything to everyone. It has been indicated that some shortcomings can be attributed to the model itself, e.g. the impossibility to satisfactorily treat multiple languages within the scope and physical limitations of a single paper dictionary. On the macrostructural level compilers of future multilingual paper dictionaries should ascertain coverage of the basic/common words of the language and make sure that at least the most common words of the languages are covered. On microstructural level compilers should pay much more attention to user guidance, sense distinction, different aspects of

consistency, parity between translation equivalents, conventions used, and to correct all kinds of errors. Consultation with linguistic experts of the African languages would also not be amiss.

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) 8576. The Grantholder acknowledge that the 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

Dictionaries

- (ANNA) Martin, Willy (Main ed.). 2011. *Pharos Groot Woordeboek. Afrikaans en Nederlands (Prisma Groot Woordenboek Afrikaans en Nederlands)*. Cape Town: Pharos.
- (CMD) Jennings, L.E. et al. 1995. *The Concise Multilingual Dictionary. English, Xhosa, Zulu, Northern Sotho, Southern Sotho, Tswana, Afrikaans*. Johannesburg: Ad Donker.
- (EANTA) Uys, Isabel. 2008. *The English, Afrikaans, Northern Sotho, Tswana Aid: Word Lists & Phrases in Four Languages*. Cape Town: Pharos.
- (MED) Rundell, M. (Ed.). 2007. *Macmillan English Dictionary for Advanced Learners*. Second Edition. Oxford: Macmillan.
- (MID) Bennett, J. et al. 2010. *Multilingual Illustrated Dictionary. English, IsiZulu Sesotho, IsiXhosa, Setswana, Afrikaans, Sepedi*. Cape Town: Pharos.
- (MLD) Reynierse, C. (Ed.). 1991. *South African Multi-Language Dictionary and Phrase Book. English, Afrikaans, Northern Sotho, Sesotho, Tswana, Xhosa, Zulu*. Cape Town: Reader's Digest Association South Africa.
- (OJPD) Goodwill, J.S. et al. 1991. *The Oxford Junior Primary Dictionary for Southern Africa with North Sotho, South Sotho, Setswana and Afrikaans Words*. Cape Town: Oxford University Press.

Other sources

- Gouws, R.H., D.J. Prinsloo and M. Dlali. 2014. A Series of Foundation Phase Dictionaries for a Multilingual Environment. *Stellenbosch Papers in Linguistics* 43(2014): 23-43.
- Haas, Mary R. 1962. What Belongs in a Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1962. *Problems in Lexicography*: 45-50. Bloomington: Indiana University.
- Kilgarriff, A., M. Husák, K. McAdam, M. Rundell and P. Rychlý. 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda and Janet DeCesaris (Eds.).

2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 425-432. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

(PEIC) Gauton, Rachéle: The University of Pretoria English Internet Corpus.

Prinsloo, D.J. 2011. A Critical Analysis of the Lemmatisation of Nouns and Verbs in isiZulu. *Lexikos* 21: 169-193.

Prinsloo, D.J. and R.H. Gouws. 2006. Lexicographic Presentation of Grammatical Divergence in Sesotho sa Leboa. *South African Journal of African Languages* 16(4): 184-197.

Probleme in Bezug auf die slowenische lexikographische Terminologie

Simona Štavbar, *Abteilung für Germanistik, Philosophische Fakultät,
Universität Maribor, Slowenien (simonastavbar@gmail.com)*

Zusammenfassung: Der Beitrag behandelt slowenische Terminologie im Bereich der lexikographischen Disziplin und die damit verbundene terminologische Problematik der lexikographischen Disziplin in Slowenien. Für diese ist ein großer Anteil an synonymen und fehlenden lexikographischen Termini charakteristisch. Im Beitrag wird gezeigt, wie und inwieweit man die fehlende slowenische lexikographische Terminologie in Anlehnung an die deutsche lexikographische Terminologie anhand der Wortbildungsverfahren einführen könnte, was zur Entwicklung der lexikographischen Theorie und Erhöhung der Qualität von lexikographischen Produkten führen würde.

Stichwörter: DEUTSCHE LEXIKOGRAPHISCHE TERMINOLOGIE, SLOWENISCHE LEXIKOGRAPHISCHE TERMINOLOGIE, SYNONYME TERMINI, FEHLENDE TERMINI, WORTBILDUNGSVERFAHREN, LEXIKOGRAPHISCHE DISZIPLIN IN SLOWENIEN

Abstract: Problems related to the Slovenian lexicographical terminology. The article discusses Slovenian terminology in the field of lexicography as well as terminological problems related to the lexicographical discipline in Slovenia. The terminology of this discipline displays a great number of synonymous lexicographical terms on the one hand, and a lack of lexicographical terms on the other hand. Based on the German lexicographical terminology, this article looks into how and to which extent the missing lexicographical terms can be formed in Slovene through different types of word-formation processes. The introduction of much needed lexicographical terms into Slovene would contribute to the formation of lexicographical theory and raise the quality of lexicographical products.

Keywords: GERMAN LEXICOGRAPHICAL TERMINOLOGY, SLOVENIAN LEXICOGRAPHICAL TERMINOLOGY, SYNONYMOUS TERMS, LACK OF TERMS, WORD FORMATION, SYNONYMOUS LEXICOGRAPHICAL TERMS, LEXICOGRAPHICAL DISCIPLINE IN SLOVENIA

1. Zum Stand und zur Problematik der slowenischen lexikographischen Terminologie

Im Folgenden wird die slowenische Terminologie im Bereich der lexikographischen Disziplin und die damit verbundene terminologische Problematik der lexikographischen Disziplin in Slowenien besprochen. In Anlehnung an die

deutsche lexikographische Terminologie werden die Möglichkeiten vorgestellt, wie man die fehlende lexikographische Terminologie anhand der Wortbildungsverfahren einführen könnte und es wird darauf hingewiesen, dass man die uneinheitliche lexikographische Terminologie im Slowenischen, die sich in Synonymie niederschlägt¹, vereinheitlichen sollte. Das Einführen von fehlenden lexikographischen Termini im Slowenischen würde nämlich zur Entwicklung der lexikographischen Theorie und zur Erhöhung der Qualität von lexikographischen Produkten führen. Auf die angeführte Problematik bezüglich der Fachterminologie, ihrer lexikographischen und terminographischen Erfassung macht Jesenšek aufmerksam:

Trotz der Vielzahl und Vielfalt der Fachwörterbücher, der steigenden Anzahl terminographischer Produkte und intensiver empirischer und theoriegeleiteter Forschung bleiben nicht wenige Problembereiche in der Auseinandersetzung mit den Fachsprachen und ihrer lexikographischen bzw. terminographischen Erfassung weiterhin ungeklärt. (Jesenšek 2013: 11)

Es werden zuerst die Gründe angeführt, die zur uneinheitlichen Terminologie in der slowenischen lexikographischen Disziplin beitragen.

2. Gründe für die uneinheitliche lexikographische Terminologie im Slowenischen

Für die uneinheitliche lexikographische Terminologie, die sich in der Verwendung synonyme Termini und auch im Fehlen von lexikographischen Termini widerspiegelt, gibt es mehrere Gründe. Ihr Vorkommen ist verbunden mit der Tätigkeit unterschiedlicher terminologischer Schulen, die in der Vergangenheit bei der Begründung der Lexikographie als selbständige Disziplin ihre eigene lexikographische Terminologie formten und in die lexikographische Disziplin einführen. Als sich diese Terminologie etablierte, wurden in die Zielsprache Termini anderer Sprachen eingeführt. Im slowenischen Raum hatte besonders die *Tschechische terminologische Schule*² einen relativ großen Einfluss auf die Art und den Umfang der lexikographischen Terminologie.³ Mit der späteren Etablierung unterschiedlicher Theorien bzw. theoretischer Vorschläge wurden in die Zielsprache neue Termini eingeführt. Dies führte jedoch zur uneinheitlichen Verwendung der lexikographischen Termini (z. B. der deutsche Terminus *Lemma* und die Varianten dieses Terminus *iztočnica/slovarska iztočnica/geslo* im Slowenischen). Das Angeführte ist auch das Resultat der mangelnden Vereinbarung bzw. der unzureichenden Zusammenarbeit unter den Fachleuten dieses Fachgebiets, da im institutionellen Sinne auf dem internationalen Niveau die Zusammenarbeit gut ist, auf dem nationalen Niveau diese jedoch nicht ausreichend ist. Der nächste Grund für die uneinheitliche Terminologie und für das Aufkommen von synonymen Termini ist der Einfluss anderer Disziplinen, darunter z. B. der Informatik. Hier wird die Terminologie dieser Disziplin von der lexikographischen Disziplin übernommen, wobei es wegen mangelnder

lexikographischer Fachkenntnisse zur uneinheitlichen Verwendung der lexikographischen Termini kommt (vgl. z. B. die oft falsche Verwendung der Termini *informacija* und *podatek* im Slowenischen bzw. *Informationen* und *Daten* im Deutschen).⁴ Auf internationalen Konferenzen, Kongressen und Kolloquien wird überwiegend das Englische als Verständigungssprache verwendet, das Englische überwiegt auch in Fachpublikationen, wie z. B. *Euralex*, *International Journal of Lexicography*, *Lexikos* u. a. Für den englisch- und deutschsprachigen Raum stehen mehrere lexikographische Fachpublikationen zur Verfügung, was man für den slowenischen Raum nicht feststellen kann. Hier erscheinen Beiträge zur allgemeinen Lexikographie, Fachlexikographie und zur metasprachlichen Lexikographie überwiegend in Publikationen anderer linguistischer Disziplinen oder werden in bestimmten thematisch konzipierten Nummern der linguistischen Fachpublikationen publiziert.⁵ Im slowenischen Raum gibt es derzeit auch keine Quellen (Fachwörterbücher, Datenbanken u. a.), in denen lexikographische Terminologie gesammelt, entsprechend behandelt und den Benutzern zugänglich wäre. Zudem stehen auch keine Lehrwerke zu genannten Themenbereichen und auch keine Übersetzungen wichtiger lexikographischer Texte zur Verfügung, die die Beschäftigung mit lexikographischer Terminologie anregen würden.⁶ Aus diesem Grund ist die deutsche Metalexikographie im slowenischen Raum wenig bekannt und damit auch die Terminologie, die mit ihrer Strukturiertheit als Basis für das Einführen der fehlenden lexikographischen Terminologie und für ihre Vereinheitlichung hätte dienen können.⁷ Ein weiterer Grund für das Genannte liegt in der letzten Zeit bestimmt auch in den fehlenden oder unzureichenden Deutschkenntnissen (als Folge der Fremdsprachenpolitik im Ausbildungssystem); das zeigt sich an der Berücksichtigung nur der englischen Literatur zum Thema.

Im Bereich der praktischen lexikographischen Tätigkeit in Slowenien werden von mehreren nationalen Instituten und auch von Unternehmen allgemeine und fachsprachliche Wörterbücher und andere lexikographische Produkte vorbereitet (z. B. vom *Inštitut za slovenski jezik Frana Ramovša*,⁸ *Institut „Jožef Stefan“*,⁹ *Trojina*¹⁰ oder *Amebis*¹¹), im Bereich der theoretischen Lexikographie lassen sich jedoch Schwierigkeiten bezüglich der Verwendung der lexikographischen Terminologie beobachten. In Fachpublikationen und in Wörterbuchaußentexten gibt es oft Fälle von synonymen lexikographischen Termini, die im Folgenden an Belegen dargestellt werden.

3. Zur synonymen Verwendung der Termini in der slowenischen lexikographischen Disziplin und die fehlende Terminologie

In slowenischen linguistischen Fachpublikationen und in Wörterbuchaußentexten wird der lexikographische Terminus *iztočnica* [*Lemma*] oft synonym verwendet, u. z. gibt es die Varianten *iztočnica*, *geslo*, *geselska iztočnica* und *slovarska iztočnica*. In der Fachliteratur ist die Form *iztočnica* etabliert, die auch im *SP* (2001)¹² und im *SSKJ2* (2014)¹³ belegt ist sowie in Wörterbuchaußentexten und

in Fachartikeln. Im SSKJ (1994)¹⁴ ist der Terminus *geslo* belegt (vgl. *Uvod* [Einführung]); *iztočnica* ist dort mit der Häufigkeitsangabe *redko* [selten] markiert. Im SP (2001) sind beide Termini belegt. Das Gleiche gilt auch für das SSKJ2 (2014). Der Terminus *iztočnica* hat sich, den Quellen zufolge, erst mit der Zeit etabliert. In Verbindung mit dem Terminus *geslo* ist im SP (2001) auch der Terminus *geselski članek* belegt, der im SSKJ2 (2014) als *slovarski sestavek* belegt ist und an anderen Stellen noch als *geselski sestavek* [Wörterbuchartikel]. In Fachartikeln werden in Verbindung mit *iztočnica* die Termini *varianta iztočnice* und *iztočnična varianta* angeführt. Im SSKJ2 (2014) ist der Terminus *različica* (*iztočnice*) belegt, was das slowenische Äquivalent zum entlehnten Terminus *varianta* (*iztočnice*) ist. Im Deutschen handelt es sich hierbei um Zusammensetzung, u. z. *Lemma-variante*. Als Einwort- bzw. Mehrwortterminus ist *ustreznik* bzw. *tujejezični ustreznik*/*tuji ustreznik* und *ustreznica* bzw. *tujejezična ustreznica* belegt. Das erste Beispiel bezieht sich auf Termini und das zweite auf Lexeme. Im Deutschen verwendet man in beiden Fällen den Terminus *Äquivalent*. Synonym kommen auch *terminološki kvalifikator* und *terminološki označevalnik* [Fachgebietsangabe] vor, wobei dieser Terminus auch deskriptiv als *označevalnik s področja strokovne rabe* belegt ist. An diesem und auch am Beispiel von *varianta* (*iztočnice*)/*različica* (*iztočnice*) lässt sich beobachten, dass die entlehnte und die slowenische Form oft parallel vorkommen. Als Synonyme sind Termini *besednovrstni podatek* und *besednovrstni kvalifikator* belegt, was im Deutschen *Wortartangabe* ist. Das slowenische Äquivalent des deutschen Terminus *Belegbeispielangabe* ist im SSKJ2 (2014) als Ein- bzw. Mehrwortterminus *zglede[i]* und *zglede iz dejanske rabe* angegeben. Das Angeführte deutet auf einen relativ großen Anteil an synonymen Termini und auf den Bedarf nach Vereinheitlichung bzw. Standardisierung der lexikographischen Terminologie im Slowenischen.

Neben der Vereinheitlichung besteht zudem der Bedarf nach den lexikographischen Termini, die es im Slowenischen nicht gibt. Das Einführen der fehlenden lexikographischen Termini würde nämlich eine höhere Qualität von lexikographischen Beschreibungen sichern. Erarbeiten und anbieten müsste man z. B. Benennungen für die Arten der Wörterbücher, u. z. *fachliches Sprachwörterbuch* (**jezikovni strokovni slovar*),¹⁵ *fachliches Sachwörterbuch* (**stvarni strokovni slovar*), *monoinformatives Wörterbuch* (**enoizborni slovar*), *Aussprachewörterbuch* (**izgovorni slovar*), *polyinformatives Wörterbuch* (**večizborni slovar*), *Sprachstadienwörterbuch* (**slovar jezikovnega obdobja/slovar jezikovnih obdobj*) u. a. In Verbindung mit dem Lemma z. B. *Lemmakandidat* (**kandidat za iztočnico*), *Lemmazeichen* (*iztočnični znak?*) und *Lemmazeichengestaltung* (*oblika iztočničnega znaka?*). Zudem gibt es keine Benennungen für die Bestandteile des Wörterbuchartikels und für Strukturen, z. B. *glattalphabetische Makrostruktur* (**tekočeabecedna makrostruktura*), *nischenalphabetische Makrostruktur* (**besedotvorno tekočeabecedna makrostruktura*), *nestalphabetische Makrostruktur* (**prekinjena tekočeabecedna makrostruktura*), *einfache Mikrostruktur* (**enostavna mikrostruktura*), *erweiterte Mikrostruktur* (**razširjena mikrostruktura*), *Verweisstruktur* (**kazalčna struktura*), *typographischer Strukturanzeiger* (**tipografski strukturni kazalnik*), *nicht-*

typographischer Strukturanzeiger (**netipografski strukturni kazalnik*) und viele andere lexikographische Termini.

4. Eigenschaften der Termini

Wie anfangs angegeben wurde, führt zur Erweiterung des lexikographischen Fachwortschatzes das Verfahren der Wortbildung. Um den Fachwortschatz im Bereich der Lexikographie im Slowenischen zu erweitern, werden zudem Lexeme aus anderen Sprachen entlehnt, wobei hier die Gebersprache das Deutsche ist.¹⁶ Im Folgenden wird kontrastiv vorgegangen, u. z. werden die lexikographischen Termini im Slowenischen in Anlehnung an das Deutsche angeführt und angeboten, da das Deutsche die Gebersprache für die slowenischen Termini ist und die slowenische lexikographische Terminologie in Anlehnung an das deutsche terminologische System und unter Berücksichtigung der deutschen lexikographischen Theorie festgelegt werden soll.¹⁷

4.1 Wortbildung

Nach Fleischer und Barz (2012: 18) handelt es sich bei der Wortbildung um „die Produktion von Lexemen auf der Grundlage und mithilfe vorhandenen Sprachmaterials.“ Das Sprachmaterial wird immer morphosyntaktisch verändert (Donalies 2002: 19). Wortbildung umfasst die Verfahren, die man für die Bildung von Lexemen zur Verfügung hat und auch den Bestand an wortgebildeten Lexemen sowie deren kommunikative Fähigkeiten (Fleischer und Barz 2012: 1). Die am häufigsten verwendeten Verfahren der Wortbildung sind — auch im Bereich der Bildung lexikographischer Termini im Deutschen und Slowenischen — die Komposition und die Derivation. Zudem gibt es noch die Konversion und Kurzwortbildung, die in der Fachlexikographie mit Deutsch sehr oft zur Bildung von Sachnamen eingesetzt wird.

Wortbildung hat einen wichtigen Anteil an der Erweiterung des lexikographischen Fachwortschatzes. Im Folgenden wird somit näher auf die Verfahren der Komposition, Derivation, Konversion, Kurzwortbildung und Entlehnung sowie Lehnwortbildung eingegangen.

4.1.1 Komposition

Komposita sind komplexe Lexeme, die sich aus zwei wortfähigen unmittelbaren Konstituenten zusammensetzen. Ihrer Abfolge nach werden die beiden unmittelbaren Konstituenten entweder als Erstglied oder als Zweitglied im Lexem unterschieden (Fleischer und Barz 2012: 84). Die Bildung von zusammengesetzten Benennungen ist im Deutschen sehr produktiv (Donalies 2002: 53). Das gilt auch für den Bereich der lexikographischen Terminologie und für die Terminologie aus anderen, benachbarten Fachgebieten, wo Komposita und

Mehrworttermini besonders zahlreich bzw. dominierend sind. Es gibt folgende Möglichkeiten zur Bildung von Komposita im Deutschen:

a) Substantiv + Substantiv

Fachlexikographie, Typologigraph, Wörterbuchfunktion, Wörterbuchtyp, Rezeptionswörterbuch, Abkürzungswörterbuch, Synonymenwörterbuch, Antonymenwörterbuch, Archaismenwörterbuch, Neologismenwörterbuch, Sachwörterbuch, Sprachwörterbuch, Lemmazeichen, Lemmavariante, Formkommentar, Wörterbuchartikel, Lemmaliste, Datenbank, Synonymenangabe, Phrasemangabe, Lemmakandidat

b) Verbstamm + Substantiv

Verweisstruktur, Zugriffsstruktur, Planungsphase

c) Adjektiv + Substantiv

Primärquelle, Sekundärquelle, Spezialwörterbuch, Kurzwort, Leitelement, Fremdwörterbuch

d) Adverb + Substantiv

Außertext, Vorspann, Nachspann

e) Numerale + Substantiv

Letztbeleg, Allbuch, Erstbeleg

Im Vergleich zum Deutschen lässt sich im Slowenischen eine wesentlich geringere Wortbildungsproduktivität im Bereich der Komposition feststellen. Die aus dem Deutschen entlehnten Termini werden im Slowenischen in der Mehrheit der Fälle als attributive Wortverbindungen realisiert, u. z. gibt es jeweils ein Substantiv als Bezugswort, das links oder rechts attribuiert ist, wobei die Fälle mit Linksattribuierung überwiegen:

Substantiv + Linksattribut

(1)
 tipolo-ški graf-Ø
 typology-ADJ.M.NOM graph-M.NOM.SG
 'typological graph' [*Typologigraph*]
 (tipologija, F.NOM)

(2)
 recepcij-ski slovar-Ø
 reception- dictionary-
 ADJ.M.NOM M.NOM.SG
 'receptive dictionary' [reception dictionary]
 [*Rezeptionswörterbuch*] (recepcija, F.NOM)

(3)
 krajšav-ni slovar-Ø
 abbreviation- dictionary-
 ADJ.M.NOM M.NOM.SG
 'abbreviated dictionary' [abbreviation dictionary]
 [*Abkürzungswörterbuch*]
 (krajšava, F.NOM)

(4)
 jezik-ovni slovar-Ø
 language- dictionary-
 ADJ.M.NOM M.NOM.SG
 'linguistic dictionary' [language dictionary]
 [*Sprachwörterbuch*]
 (jezik, M.NOM)

(5)
 iztočnič-ni znak-Ø
 lemma-ADJ.M.NOM sign-M.NOM.SG
 'lemmatic sign' [lemma sign] [*Lemmazeichen*]
 (iztočnica, F.NOM)

(6)
 oblik-ovni komentar-Ø
 form-ADJ.M.NOM comment-M.NOM.SG
 'form comment' [*Formkommentar*]
 (oblika, F.NOM)

- (7) slovar-ski članek-Ø
dictionary-ADJ.M.NOM article-M.NOM.SG
'dictionary article' [Wörterbuchartikel]
(slovar, M.NOM)
- (8) kazalč-na struktur-a
cross-reference- structure-
ADJ.F.NOM F.NOM.SG
'cross-reference structure' [Verweisstruktur]
(kazalec, M.NOM)
- (9) načrtov-(a)lna faz-a
plan-ADJ.F.NOM phase-F.NOM.SG
'planning phase' [Planungsphase]
(načrtovati, INF)
- (10) primar-ni vir-Ø
primary-M.NOM source-M.NOM.SG
'primary source' [Primärquelle]
(primaren, ADJ; cf. German: primär)
- (11) sekundar-ni vir-Ø
secondary-M.NOM source-M.NOM.SG
'secondary source' [Sekundärquelle]
(sekundaren, ADJ; cf. German: sekundär)
- (12) zun-anje besedil-o
out-ADJ.N.NOM text-N.NOM.SG
'outer text' [Außentext]
(zunaj, ADV*)
- (13) poseb-ni slovar-Ø
'separately'[special]- dictionary-
ADJ.M.NOM M.NOM.SG
'special dictionary' [Spezialwörterbuch]
(posebej, ADV)
*ADV=adverb; 1=shift of meaning
- (14) uvod-ni del-Ø
introduction-ADJ.M.NOM part-M.NOM.SG
'introductory part' [Vorspann]
(uvod, M.NOM)

Substantiv + Rechtsattribut

- (1) funkcij-a slovarj-a
function-F.NOM.SG dictionary-M.GEN.SG
'function of a dictionary' [Wörterbuchfunktion]
- (2) vrst-a slovarj-a
type-F.NOM.SG dictionary-M.GEN.SG
'type of a dictionary' [Wörterbuchtyp]
- (3) slovar-Ø arhaizm-ov
dictionary-M.NOM.SG archaism-GEN.PL
'dictionary of archaisms' [Archaismenwörterbuch]
- (4) slovar-Ø neologizm-ov
dictionary-M.NOM.SG neologism-GEN.PL
'dictionary of neologisms' [Neologismenwörterbuch]
- (5) različic-a iztočnic-e
variant-F.NOM.SG lemma-F.GEN.SG
'variant of a lemma' [Lemmavariante]
- (6) seznam-Ø iztočnic-Ø
list-M.NOM.SG lemma-GEN.PL
'list of lemmata' [Lemmaliste]
- (7) struktur-a dostopanj-a
structure-F.NOM.SG access-N.GEN.SG
'structure of access' [Zugriffsstruktur]
- (8) slovar-Ø tujk-Ø
dictionary-M.NOM.SG foreign word-GEN.PL
'dictionary of foreign words' [Fremdwörterbuch]

In einigen Fällen stehen im Slowenischen beide Möglichkeiten zur Verfügung, oft als paralleles Vorkommen von entlehntem (sog. Internationalismen) und einheimischem Sprachmaterial:¹⁸

- (1) *Synonymenwörterbuch* → slovar sinonimov/sopomenski slovar/sinonimni slovar,
(2) *Antonymenwörterbuch* → slovar antonimov/protipomenski slovar/antonimni slovar

- (1)
- | | | | |
|---|----------------|--|---------------------|
| slovar-Ø | sinonim-ov | sopomen-ski | slovar-Ø |
| dictionary-M.NOM.SG. | synonym-GEN.PL | synonym-ADJ.M.NOM | dictionary-M.NOM.SG |
| ‘dictionary of synonyms’ [<i>Synonymenwörterbuch</i>] | | ‘synonymous dictionary’ [<i>Synonymenwörterbuch</i>]
(sopomenka, F.NOM) | |
-
- | | |
|--|---------------------|
| sinonim-ni | slovar-Ø |
| synonym-ADJ.M.NOM | dictionary-M.NOM.SG |
| ‘synonymous dictionary’ [<i>Synonymenwörterbuch</i>]
(sinonim, M.NOM) | |
- (2)
- | | | | |
|---|----------------|---|---------------------|
| slovar-Ø | antonim-ov | protipomen-ski | slovar-Ø |
| dictionary-M.NOM.SG | antonym-GEN.PL | antonym-ADJ.M.NOM | dictionary-M.NOM.SG |
| ‘dictionary of antonyms’ [<i>Antonimenwörterbuch</i>] | | ‘antonymous dictionary’ [<i>Antonimenwörterbuch</i>]
(protipomenka, F.NOM) | |
-
- | | |
|--|---------------------|
| antonim-ni | slovar-Ø |
| antonym-ADJ.M.NOM | dictionary-M.NOM.SG |
| ‘antonymous dictionary’ [<i>Antonimenwörterbuch</i>]
(antonim, M.NOM) | |

Einige Beispiele deutscher lexikographischer Termini werden im Slowenischen als präpositionale Wortverbindungen realisiert:

(1) *Synonymenangabe* → *podatek o sinonimu*, (2) *Phrasemangabe* → *podatek o frazemu*, (3) *Lemmakandidat* → *kandidat za iztočnico*

- (1)
- | | | |
|--|----|----------------|
| podatek-Ø | o | sinonim-u |
| data-NOM.SG | on | synonym-LOC.SG |
| ‘data on a synonym’ [<i>Synonymenangabe</i>] | | |
- (2)
- | | | |
|--|----|----------------|
| podatek-Ø | o | frazem-u |
| data-NOM.SG | on | phrasem-LOC.SG |
| ‘data on a phrasem’ [<i>Phrasemangabe</i>] | | |
- (3)
- | | | |
|--|-----|--------------|
| kandidat-Ø | za | iztočnic-o |
| candidate-NOM.SG | for | lemma-ACC.SG |
| ‘candidate for a lemma’ [<i>Lemmakandidat</i>] | | |

Aus dem bislang Angeführten ist ersichtlich, dass Zusammensetzungen im Slowenischen zwar vorhanden sind, sie spielen jedoch eine untergeordnete Rolle. Zudem gibt es vereinzelt Fälle von deutschen Zusammensetzungen, die im Slowenischen als nichtattribuierte Einworttermini belegt sind, z. B. (1) *Nest-anordnung* → *gnezdenje*, (2) *Nachspann* → *dodatek*

- (1)
- | |
|---|
| gnezdenj-e |
| nesting-N.NOM.SG |
| ‘nesting’ ¹ [<i>Nestanordnung</i>] |
| 1=substantive; 2=shift of meaning |
- (2)
- | |
|---|
| dodatek-Ø |
| appendix [back matter] ² -M.NOM.SG |
| ‘appendix’ [back matter] [<i>Nachspann</i>] |

Bei längeren Zusammensetzungen in den Fachsprachen lässt sich beobachten, dass auch hier wie in der Allgemeinsprache das Verfahren der Reduktion einsetzt (Poethe 2000: 204). Mittels Kürzung lassen sich ökonomischere Benennungen erzeugen. Lexikographische terminologische Einheiten in Form von Zusammensetzungen werden in einigen Fällen bei der ersten Anführung im

Text in ihrer Grundform angeführt, bei jeder weiteren Erwähnung jedoch in der Kurzform:

Lesartenetiketierung → *Ettiketierung, Kurzettiketierung; Lesartendisambiguierung* → *Disambiguierung; Sprachnorm* → *Norm, Registerhaupteintrag* → *Haupteintrag, Artikelnester* → *Nester, Artikelnischen* → *Nischen, Wörterbuchartikel* → *Artikel, Mikrostrukturanzeiger* → *Strukturanzeiger, Artikelmikrostruktur* → *Mikrostruktur*

Solche Grundformen mit den dazugehörigen Kurzformen lassen sich für das Slowenische, mit einigen Ausnahmen (z. B. (1) *slovarska iztočnica* → *iztočnica* [Lemma], (2) *zglied[i] iz dejanske rabe* → *zglied[i]* [Belegbeispielangabe(n)], nicht belegen. Der Grund dafür liegt in der wenig ausgearbeiteten lexikographischen Terminologie im Slowenischen, die eine geringe Strukturiertheit des lexikographischen terminologischen Systems und der Relationen zwischen den Begriffen aufweist.

- (1)
- | | | | | |
|----------------------------|--|----------------|--|-----------------|
| slovar-ska | | iztočnic-a | | iztočnic-a |
| dictionary-ADJ.F.NOM | | lemma-F.NOM.SG | | lemma-F.NOM.SG |
| 'dictionary lemma' [Lemma] | | | | 'lemma' [Lemma] |
- (2)
- | | | | | |
|---|----|------------------|--------------|-----------------------------------|
| zglied-i | iz | dejan-ske | rab-e | zglied-i |
| example-NOM.PL | of | act-ADJ.F.GEN.SG | use-F.GEN.SG | example-NOM.PL |
| 'examples of actual use' [Belegbeispielangaben] | | | | 'examples' [Belegbeispielangaben] |

Daneben hat das Deutsche grundsätzlich die Möglichkeit, relativ lange und komplexe Einwortbenennungen zu bilden. Diese sind in der lexikographischen Terminologie des Deutschen recht zahlreich vertreten und stellen in Fällen von terminologischen Lücken in anderen Sprachen ein kaum lösbares Übersetzungsproblem dar (Butina-Koller et al. 2010: XXV-XXX), besonders, weil man sie oft nur in Form von festen Wortverbindungen übersetzen kann, was zugleich eine Abweichung vom Formaspekt des Terminus bedeutet. Besonders problematisch ist das in Fällen, wo es im Deutschen zu Bildungen mit hoher Gliederanzahl kommt: *Verweisadressangabe, Belegbeispielangabe, Wörterbuchbenutzungstest, Wörterbuchbenutzungsforschung, Datendistributionsstruktur, Fachwörterbuchartikel, Sprachstadienwörterbuch, Artikelstrukturschema, Lemmazeichengestaltungangabe, Lemmakandidatenliste*

Für solche Zusammensetzungen im Deutschen gilt, dass sie im Slowenischen in der Regel durch attributive oder präpositionale Wortverbindungen wiedergegeben werden, wobei man im Slowenischen besonders eine hohe Anzahl von mehrgliedrigen attributiven Verbindungen beobachten kann:

- (1) *Wörterbuchbenutzungstest* → *test rabe slovarja*, (2) *Wörterbuchbenutzungsforschung* → *raziskovanje rabe slovarja*, (3) *Datendistributionsstruktur* → *strukturna porazdelitve podatkov*, (4) *Fachwörterbuchartikel* → *slovarski članek strokovnega slovarja*, (5) *Sprachstadienwörterbuch* → *slovar jezikovnega obdobja*, (6) *Artikelstruk-*

turschema → *shema strukture slovarskega članka*, (7) *Lemmakandidatenliste* → *seznam kandidatov za iztočnico*

- | | | | | | | | | | | | | | | | | | | | | | | |
|--|---------------|---------------|-------------|-------------|---------------|-------------|----------|----------|----------|--|---------------|------------|--------------|-----------|-------------|-------------|--------------|-------------|-----------|----------|----------|----------|
| <p>(1)</p> <table border="0"> <tr> <td>test-Ø</td> <td>rab-e</td> <td>slovarj-a</td> </tr> <tr> <td>test-</td> <td>use-</td> <td>dictionary-</td> </tr> <tr> <td>M.NOM.SG</td> <td>F.GEN.SG</td> <td>M.GEN.SG</td> </tr> </table> <p>'test of dictionary use' [<i>Wörterbuchbenutzungstest</i>]</p> | test-Ø | rab-e | slovarj-a | test- | use- | dictionary- | M.NOM.SG | F.GEN.SG | M.GEN.SG | <p>(2)</p> <table border="0"> <tr> <td>raziskovanj-e</td> <td>rab-e</td> <td>slovarj-a</td> </tr> <tr> <td>research-</td> <td>use-</td> <td>dictionary-</td> </tr> <tr> <td>N.NOM.SG</td> <td>F.GEN.SG</td> <td>M.GEN.SG</td> </tr> </table> <p>'research of dictionary use' [<i>Wörterbuchbenutzungsforschung</i>]</p> | raziskovanj-e | rab-e | slovarj-a | research- | use- | dictionary- | N.NOM.SG | F.GEN.SG | M.GEN.SG | | | |
| test-Ø | rab-e | slovarj-a | | | | | | | | | | | | | | | | | | | | |
| test- | use- | dictionary- | | | | | | | | | | | | | | | | | | | | |
| M.NOM.SG | F.GEN.SG | M.GEN.SG | | | | | | | | | | | | | | | | | | | | |
| raziskovanj-e | rab-e | slovarj-a | | | | | | | | | | | | | | | | | | | | |
| research- | use- | dictionary- | | | | | | | | | | | | | | | | | | | | |
| N.NOM.SG | F.GEN.SG | M.GEN.SG | | | | | | | | | | | | | | | | | | | | |
| <p>(3)</p> <table border="0"> <tr> <td>struktur-a</td> <td>porazdelitv-e</td> <td>podatk-ov</td> </tr> <tr> <td>structure-</td> <td>distribution-</td> <td>data-</td> </tr> <tr> <td>F.NOM.SG</td> <td>F.GEN.SG</td> <td>GEN.PL</td> </tr> </table> <p>'structure of data distribution' [<i>Datendistributionsstruktur</i>]</p> | struktur-a | porazdelitv-e | podatk-ov | structure- | distribution- | data- | F.NOM.SG | F.GEN.SG | GEN.PL | <p>(4)</p> <table border="0"> <tr> <td>slovar-ski</td> <td>članek-Ø</td> <td>strokovn-ega</td> <td>slovarj-a</td> </tr> <tr> <td>dictionary-</td> <td>article-</td> <td>specialized-</td> <td>dictionary-</td> </tr> <tr> <td>ADJ.M.NOM</td> <td>M.NOM.SG</td> <td>M.GEN.SG</td> <td>M.GEN.SG</td> </tr> </table> <p>'dictionary article of a specialized dictionary' [<i>Fachwörterbuchartikel</i>]</p> | slovar-ski | članek-Ø | strokovn-ega | slovarj-a | dictionary- | article- | specialized- | dictionary- | ADJ.M.NOM | M.NOM.SG | M.GEN.SG | M.GEN.SG |
| struktur-a | porazdelitv-e | podatk-ov | | | | | | | | | | | | | | | | | | | | |
| structure- | distribution- | data- | | | | | | | | | | | | | | | | | | | | |
| F.NOM.SG | F.GEN.SG | GEN.PL | | | | | | | | | | | | | | | | | | | | |
| slovar-ski | članek-Ø | strokovn-ega | slovarj-a | | | | | | | | | | | | | | | | | | | |
| dictionary- | article- | specialized- | dictionary- | | | | | | | | | | | | | | | | | | | |
| ADJ.M.NOM | M.NOM.SG | M.GEN.SG | M.GEN.SG | | | | | | | | | | | | | | | | | | | |
| <p>(5)</p> <table border="0"> <tr> <td>slovar-Ø</td> <td>jezikovn-ega</td> <td>obdobj-a</td> </tr> <tr> <td>dictionary-</td> <td>linguistic-</td> <td>period-</td> </tr> <tr> <td>M.NOM.SG</td> <td>N.GEN.SG</td> <td>N.GEN.SG</td> </tr> </table> <p>'dictionary of a linguistic [language] period' [<i>Sprachstadienwörterbuch</i>]</p> | slovar-Ø | jezikovn-ega | obdobj-a | dictionary- | linguistic- | period- | M.NOM.SG | N.GEN.SG | N.GEN.SG | <p>(6)</p> <table border="0"> <tr> <td>shem-a</td> <td>struktur-e</td> <td>slovarsk-ega</td> <td>člank-a</td> </tr> <tr> <td>scheme-</td> <td>structure-</td> <td>dictionary-</td> <td>article-</td> </tr> <tr> <td>F.NOM.SG</td> <td>F.GEN.SG</td> <td>M.GEN.SG</td> <td>M.GEN.SG</td> </tr> </table> <p>'scheme of a structure of a dictionary article' [<i>Artikelstruktur-schema</i>]</p> | shem-a | struktur-e | slovarsk-ega | člank-a | scheme- | structure- | dictionary- | article- | F.NOM.SG | F.GEN.SG | M.GEN.SG | M.GEN.SG |
| slovar-Ø | jezikovn-ega | obdobj-a | | | | | | | | | | | | | | | | | | | | |
| dictionary- | linguistic- | period- | | | | | | | | | | | | | | | | | | | | |
| M.NOM.SG | N.GEN.SG | N.GEN.SG | | | | | | | | | | | | | | | | | | | | |
| shem-a | struktur-e | slovarsk-ega | člank-a | | | | | | | | | | | | | | | | | | | |
| scheme- | structure- | dictionary- | article- | | | | | | | | | | | | | | | | | | | |
| F.NOM.SG | F.GEN.SG | M.GEN.SG | M.GEN.SG | | | | | | | | | | | | | | | | | | | |
| <p>(7)</p> <table border="0"> <tr> <td>seznam-Ø</td> <td>kandidat-ov</td> <td>za</td> <td>iztočnic-o</td> </tr> <tr> <td>list-</td> <td>candidate-</td> <td>for</td> <td>lemma-</td> </tr> <tr> <td>M.NOM.SG</td> <td>GEN.PL</td> <td></td> <td>F.ACC.SG</td> </tr> </table> <p>'list of candidates for a lemma' [<i>Lemmakandidatenliste</i>]</p> | seznam-Ø | kandidat-ov | za | iztočnic-o | list- | candidate- | for | lemma- | M.NOM.SG | GEN.PL | | F.ACC.SG | | | | | | | | | | |
| seznam-Ø | kandidat-ov | za | iztočnic-o | | | | | | | | | | | | | | | | | | | |
| list- | candidate- | for | lemma- | | | | | | | | | | | | | | | | | | | |
| M.NOM.SG | GEN.PL | | F.ACC.SG | | | | | | | | | | | | | | | | | | | |

In Fällen von Mehrwortbenennungen gibt es im Deutschen zahlreiche Beispiele terminologischer Bildungen, bei denen sich die Struktur des Terminus aus einem oder mehreren attributiven Adjektiven vor einem substantivischen Kompositum oder einer Ableitung zusammensetzt oder die Struktur Genitiv- bzw. Präpositionalattribut-Argumente in Nachstellung enthält. Die deutsche lexikographische Terminologie ist nämlich durch ein erhöhtes Erscheinen von komplexen Lexemen charakterisiert. Solche komplexe Strukturen haben einen differenzierenden und spezifizierenden Charakter. Beispiele solcher Bildungen sind in der lexikographischen Terminologie zahlreich: *textbezogene Wörterbuchfunktion*, *wissensbezogene Wörterbuchfunktion*, *abstrakte hierarchische linke Kernstruktur*, *konkrete hierarchische rechte Kernstruktur*, *basaler Wörterbuchartikel*, *typographischer Strukturanzeiger*, *nicht-typographischer Strukturanzeiger*, *lexikographische Datenerhebung*

Wie in vielen bislang angeführten Beispielen, in denen kein slowenischer Terminus nachgewiesen werden konnte, werden slowenische lexikographische Termini in solchen Fällen durch eine wörtliche Übersetzung geschaffen. Da sich das Wortbildungssystem im Slowenischen von dem Deutschen unterscheidet, wirken viele slowenische Termini nicht natürlich, was besonders an den unteren Mehrwortbildungen in Form von längeren Wortketten ersichtlich ist:

- (1) *textbezogene Wörterbuchfunktion* → *slovarska funkcija v zvezi z besedilom*, (2) *wissensbezogene Wörterbuchfunktion* → *slovarska funkcija v zvezi z znanjem*, (3) *abstrakte hierarchische linke Kernstruktur* → *abstraktna hierarhična leva jedrna struktura*, (4) *konkrete hierarchische rechte Kernstruktur* → *konkretna hierarhična desna struk-*

tura, (5) *basaler Wörterbuchartikel* → *temeljni slovarski članek*, (6) *typographischer Strukturanzeiger* → *tipografski strukturni kazalnik*, (7) *lexikographische Datenerhebung* → *slovaropisno pridobivanje podatkov*

- (1)
 slovarsk-a funkcij-a v zvez-i z besedil-om
 dictionary-F.NOM.SG function-F.NOM.SG in relation-F.LOC to text-N. INS
 ‘dictionary function in relation to a text’ [*textbezogene Wörterbuchfunktion*]
- (2)
 slovarsk-a funkcij-a v zvez-i z znanj-em
 dictionary-F.NOM.SG function-F.NOM.SG in relation-F.LOC to knowledge-N.INS
 ‘dictionary function in relation to knowledge’ [*wissensbezogene Wörterbuchfunktion*]
- (3)
 abstraktn-a hierarhičn-a lev-a jedrn-a struktur-a
 abstract-F.NOM.SG hierarhical-F.NOM.SG left-F.NOM.SG core-F.NOM.SG structure-F.NOM.SG
 ‘abstract hierarchical left core structure’ [*abstrakte hierarchische linke Kernstruktur*]
- (4)
 konkretn-a hierarhičn-a desn-a jedrn-a struktur-a
 concrete- hierarhical- right- core- structure-
 F.NOM.SG F.NOM.SG F.NOM.SG F.NOM.SG F.NOM.SG
 ‘concrete hierarchical right core structure’ [*konkrete hierarchische rechte Kernstruktur*]
- (5)
 temeljn-i slovarsk-i članek-Ø
 basic-M.NOM.SG dictionary-M.NOM.SG article-M.NOM.SG
 ‘basic dictionary article’ [*basaler Wörterbuchartikel*]
- (6)
 tipografsk-i strukturn-i kazalnik-Ø
 typographical-M.NOM.SG structural-M.NOM.SG indicator-M.NOM.SG
 ‘typographical structural indicator’ [*typographischer Strukturanzeiger*]
- (7)
 slovaropisn-o pridobivanj-e podatkov
 lexicographical-N.NOM.SG acquisition-N.NOM.SG data-GEN.PL
 ‘lexicographical acquisition of data’ [*lexikographische Datenerhebung*]

4.1.2 Derivation

Bei der Wortableitung bzw. Derivation handelt es sich um die Verbindung einer Derivationsbasis (Wortstamm, Konfix oder Syntagma) mit einem Derivationsaffix (Präfix, Suffix oder Zirkumfix) (vgl. Fleischer und Barz 2012: 86). Die Suffixderivation betrifft vor allem Substantive und Adjektive, die Präfixderivation und Konversion sind für Verben charakteristisch (ebd.).

Im Folgenden werden die Wortbildungsprodukte der Derivation — die Suffixderivation, die Präfixderivation und die Zirkumfixderivation — an Beispielen der Lexeme aus dem Bereich der Lexikographie angeführt:

a1) Suffixderivation bei Substantiven

Charakteristisch hierfür sind folgende Suffixe: *-e* (*Liste*, *Experte*), *-ei/-erei* (*Datei*), *-el* (*Artikel*), *-er* (*Benutzer*, *Nutzer*), *-heit/-keit/-igkeit* (*Häufigkeit*, *Prüfbarkeit*), *-nis* (*Verzeichnis*), *-(i)s* (*Basis*), *-ung* (*Markierung*, *Adressierung*). Solche Belege werden

im Slowenischen entweder als Simplex, als Ableitung oder als Zusammensetzung realisiert. Simplex im Slowenischen ist: (1) *seznam* [*Liste; Verzeichnis*]; Belege für die Suffixableitung sind: (2) *strokovnjak* [*Experte*], (3) *uporabnik* [*Nutzer, Benutzer*], (4) *sestavek* [*Artikel*], (5) *pogostost* [*Häufigkeit*], (6) *preverljivost* [*Prüfbarkeit*], (7) *osnova* [*Basis*], (8) *označevanje* [*Markierung*], (9) *naslavljanje* [*Adressierung*]; Zusammensetzung: (10) *datoteka* [*Datei*].

(1) seznam-Ø list-NOM.SG 'list' [<i>Liste; Verzeichnis</i>]	(2) strokovn-jak expert-NOM.SG (suffix: -jak) 'expert' [<i>Experte</i>] (strokoven, ADJ)
(3) uporab-nik use-NOM.SG (suffix: -nik) 'user' [<i>Nutzer, Benutzer</i>] (uporaba, S*)	(4) sestav-ek compile-NOM.SG (suffix: -ek) 'compilation' [<i>article</i>] [<i>Artike</i>] (sestaviti, V*)
(5) pogost-ost frequent-NOM.SG (suffix: -ost) 'frequency' [<i>Häufigkeit</i>] (pogost, ADJ)	(6) preverljiv-ost verifiable-NOM.SG (suffix: -ost) 'verification' [<i>Prüfbarkeit</i>] (preverljiv, ADV)
(7) osnov-a basic-NOM.SG (suffix: -a) 'basis' [<i>Basis</i>]	(8) označ-evanje label-NOM.SG (suffix: -evanje) 'label' [<i>Markierung</i>] (označiti, V)
(9) naslavljanje address-NOM.SG (suffix: -anje) 'address' [<i>Adressierung</i>] (nasloviti, V)	(10) datotek-a file-NOM.SG (stem: dat-; interfixal formant: -o-; stem: -(biblio)teka) 'file' [<i>Datei</i>] (podatkovna + (biblio)teka)

*V=verb, S=substantive

a2) Suffixderivation bei Adjektiven

Man findet folgende Suffixe: *-bar* (*prüfbar*), *-er* (*inner, basaler*), *-ich* (*fachlich*) *-isch* (*lexikographisch, typographisch*), *-los* (*erfolglos*). Im Slowenischen werden die obigen Beispiele als Präfix- bzw. Suffixableitung realisiert. Präfixableitung: (1) *neuspešen* [*erfolglos*]; Belege für die Suffixableitung: (2) *preverljiv* [*prüfbar*], (3) *notranji* [*inner*], (4) *temeljni* [*basaler*], (5) *strokoven* [*fachlich*], (6) *leksikografski* [*lexikographisch*], (7) *tipografski* [*typographisch*].

(1) ne-uspeš-en NEG-success- ADJ.M.NOM 'unsuccessful' [<i>erfolglos</i>] (uspeh, M.NOM)	(2) prever-(l)jiv test-ADJ.M.NOM (suffix: -(l)jiv) 'testable' [<i>prüfbar</i>] (preveriti, V)
--	---

(3)
notr-anji
inside-ADJ.M.NOM (suffix: -anji)
'inner' [*inner*]
(noter, ADV)

(4)
temelj-ni
basis-ADJ.M.NOM (suffix: -ni)
'basic' [*bazaler*]

(5)
strok-oven
special field-ADJ.M.NOM (suffix: -oven)
'special field'¹ [*fachlich*]
(stroka, F.NOM.)

(6)
leksikograf-ski
lexicography-ADJ.M.NOM (suffix: -ski)
'lexicographical' [*lexikographisch*]

(7)
tipograf-ski
typography-ADJ.M.NOM (suffix: -ski)
'typographical' [*typographisch*]
(tipografija, F.NOM.)
1=adj. only before a noun: *special field dictionary*

a3) Suffixderivation bei Verben

Für die Suffixderivation bei Verben gibt es nur wenige produktive Suffixe, u. z. *-el(n)/-l(n)* (*vermitteln*), *-er(n)/-r(n)* (*verändern*), *-ier(en)/-isier(en)/-efizier(en)* (*standardisieren*, *lemmatisieren*, *sortieren*, *publizieren*). Die angegebenen Belege werden im Slowenischen als Präfixableitung realisiert: (1) *posredovati* [*vermitteln*], (2) *spremeniti* [*verändern*]; Belege für die Suffixableitung sind: (3) *standardizirati* [*standardisieren*], (4) *lematizirati* [*lemmatisieren*], (5) *urejati* [*sortieren*], (6) *objaviti* [*publizieren*].

(1)
po-sred-ovati
PREP*-middle- (prefix: po-, verbal suffix:
INF -ova-, verbal formant: -ti)
'to mediate' [to communicate]¹ [*vermitteln*]

(2)
s-premen-iti
PREP-change- (prefix: s-, verbal suffix: -i-,
INF verbal formant: -ti)
'to change' [*verändern*]

(3)
standard-
izirati
standard-INF (verbal suffix: -izira-, verbal
formant: -ti)
'to standardize' [*standardisieren*]

(4)
lemat-izirati
lemma-INF (verbal suffix: -izira-, verbal
formant: -ti)
'to lemmatize' [*lemmatisieren*]

(5)
urej-ati
sorted-INF (verbal suffix: -a-, verbal
formant: -ti)
'to sort' [*sortieren*]
*PREP=preposition; 1=shift of meaning

(6)
objav-iti
publication-INF (verbal suffix: -i-, verbal
formant: -ti)
'to publish' [*publizieren*]

b1) Präfixderivation bei Substantiven

Es lassen sich folgende Präfixe beobachten: *haupt-* (*Hauptlemma*), *ur-* (*Urbeleg*, *Urliste*). Im Slowenischen wird der erste Beleg als attributive Wortverbindung realisiert: *glavna iztočnica* [*Hauptlemma*], bei den zwei darauffolgenden Belegen handelt es sich um Fälle von Bildungen mit dem Präfix *-pra* und um Beispiele der Wortbildungsart „*sestavljanje* bzw. *sestava*“¹⁹ (Präfigierung): (1) *praprimer* [*Urbeleg*], (2) *praseznam* [*Urliste*].

- | | |
|---|---|
| <p>(1)
 praprimer-Ø
 original.example-NOM.SG (prefix: pra-)
 ‘original example’ [Urbeleg]</p> | <p>(2)
 praseznam-Ø
 original.list-NOM.SG (prefix: pra-)
 ‘original list’ [Urliste]</p> |
|---|---|

b2) Präfixderivation bei Adjektiven

Die Anzahl der Präfixe ist gering (Fleischer und Barz 2012: 351): *un-* (*unabhängig, unmittelbar*). Im Slowenischen handelt es sich um Fälle von „*sestavljanje* bzw. *sestava*“: (1) *neodvisen* [*unabhängig*], (2) *neposreden* [*unmittelbar*].

- | | |
|--|---|
| <p>(1)
 ne-odvis-en
 NEG-depend-ADJ.M.NOM (prefix: ne-)
 ‘independent’ [<i>unabhängig</i>]
 I=shift of meaning</p> | <p>(2)
 ne-posred-en
 NEG-in.the.middle-ADJ.M.NOM (prefix: ne-)
 ‘not in the middle’ [immediate, direct]¹ [<i>unmittelbar</i>]</p> |
|--|---|

b3) Präfixderivation bei Verben

Nach Fleischer und Barz (2012: 380) gehört die Präfixderivation zu den in der verbalen Wortbildung am häufigsten genutzten Wortbildungsarten. Verbale Präfixderivate sind untrennbar. Hier gibt es folgende Präfixe: *be-* (*benutzen, bearbeiten*), *ver-* (*verweisen, umverteilen*), *zer-* (*zergliedern*), *über-* (*überprüfen*), *um-* (*umschreiben*), *unter-* (*unterstützen*). Im Slowenischen handelt es sich bei den folgenden Fällen um Präfixderivation: (1) *uporabiti* [*benutzen*], (2) *obdelati* [*bearbeiten*], (3) *usmerjati* [*verweisen*], (4) *prerazporediti* [*umverteilen*], (5) *razčleniti* [*zergliedern*], (6) *preveriti* [*überprüfen*], (7) *opisati* [*umschreiben*], (8) *podpreti* [*unterstützen*].

- | | |
|--|---|
| <p>(1)
 u-porab-iti
 PREP-use-INF (prefix: u-,
 verbal suffix: -i-, verbal
 formant: -ti)
 ‘to use’ [<i>benutzen</i>]</p> | <p>(2)
 ob-del-ati
 PREP-work- (prefix: ob-, verbal suffix: -a-,
 INF verbal formant: -ti)
 ‘to rework’ [revise] [<i>bearbeiten</i>]</p> |
| <p>(3)
 u-smerj-ati
 PREP-refer-INF (prefix: u-, verbal suffix: -a-,
 verbal formant-ti)
 ‘to refer [to]’ [<i>verweisen</i>]</p> | <p>(4)
 pre-razpored-iti
 ADV-distribution- (prefix: pre-, verbal suffix:
 INF -i-, verbal formant: -ti)
 ‘to redistribute’ [<i>umverteilen</i>]</p> |
| <p>(5)
 raz-člen-iti
 ADV-segment- (prefix: raz-, verbal suffix:
 INF -i-, verbal formant: -ti)
 ‘to segment’ [<i>zergliedern</i>]</p> | <p>(6)
 pre-ver-iti
 ADV-verify- (prefix: pre-, verbal suffix: -i-,
 INF verbal formant: -ti)
 ‘to verify’ [<i>überprüfen</i>]</p> |
| <p>(7)
 o-pis-ati
 X*-describe- (prefix: o-, verbal suffix: -a-,
 INF verbal formant: -ti)
 ‘to describe’ [<i>umschreiben</i>]</p> | <p>(8)
 pod-pr-eti
 ADV- (prefix: pod-, verbal suffix: -e-,
 ‘support’- verbal formant: -ti); ADV <i>pod</i>
 INF meaning <i>spodaj</i>, i.e. <i>under</i>; -<i>preti</i>
 meaning <i>push, press</i>
 ‘to support’ [<i>unterstützen</i>]</p> |

*X=no specification possible;

c1) Zirkumfixderivation bei Adjektiven²⁰

Hier ist die Derivation nicht auf die Kombinationen mit *ge-* beschränkt, z. B. *inter-* (*interfachlich*). Das Beispiel hier ist im Slowenischen ein Beleg für die Wortbildungsart „*sestavljanje* bzw. *sestava*“: (1) *medstrokov*en [*interfachlich*].

- (1)
med-strok-oven
PREP-discipline-ADJ.M.NOM (prefix: med-)
'interdisciplinary' [*interfachlich*]

c2) Zirkumfixderivation bei Verben

Die Derivation mit Zirkumfixen *be-/ver- ... -ig(en)* und *ver- ... -ier(en)* ist schwach ausgebaut. Hierbei entstehen untrennbare Verben (Fleischer und Barz 2012: 433), z. B. *beabsichtigen*. Im Slowenischen handelt es sich um Präfixderivation: (1) *nameravati* [*beabsichtigen*].

- (1)
na-mer-avati
PREP-'intend'-INF (prefix: na-, verbal suffix: -ava-,
verbal formant: -ti; *nameriti*
meaning *to direct* [ro])
'to intend' [*beabsichtigen*]

4.1.3 Konversion

Bei der Konversion handelt es sich um den Übergang von Lexemen einer Wortklasse in die andere (vgl. Fleischer und Barz 2012: 87). Man unterscheidet zwei Arten von Konversion, u. z. die morphologische und die syntaktische Konversion. Im Falle der morphologischen Konversion „stehen zwei phonologisch gleiche Stämme in einer Motivationsbeziehung zueinander“ (ebd.: 88). Beispiele der morphologischen Konversion im Bereich der lexikographischen Lexik sind z. B. *verweis(en)* → *Verweis*, *beleg(en)* → *Beleg* u. a.²¹ In Fällen mit parallelem Vorkommen von Verb und abstraktem Substantiv ist aus der synchronen Perspektive jedoch schwierig zu bestimmen, in welche Richtung die Konversion verläuft (vgl. ebd.).

Bildungen mit *Ab-*, *An-*, *Zu-* sind Konversionen von trennbar zusammengesetzten Verben: *zugreifen* → *Zugriff*, *angeben* → *Angabe*, *abstehen* → *Abstand*.

Um syntaktische Konversion handelt es sich, wenn das Konvertat ein Flexionselement seiner Basis behält, z. B. das grammatische Morphem *-en* in Fällen des substantivierten Infinitivs (vgl. ebd.): *nachschlagen* → *Nachschlagen*, *adressieren* → *Adressieren*, *wissen* → *Wissen*, *nutzen* → *Nutzen*, *lesen* → *Lesen* u. a. Aus dem Angeführten ist ersichtlich, dass es sich um gemeinsprachliche Lexeme handelt, die auch in der Lexikographie verwendet werden.

Im Vergleich zum Deutschen gibt es hier jedoch durch die Sprachstruktur bedingte Unterschiede. Typisch für das Slowenische ist, dass Verben gewöhnlich als Basis für Substantivableitungen dienen, u. z. mit oder ohne Wortbil-

dungsaffixe. In dieser Hinsicht gibt es Fälle der Suffixableitung mit Nullsuffix: (1) *dostopati* → *dostop*, (2) *nasloviti* → *naslov*, (3) *odmakniti* → *odmik*, oder Fälle mit Suffixableitung: (4) *rabiti* → *raba*, (5) *poiskati* → *iskanje*.

- (1)
- | | | |
|----------------------------------|----------------------|-----------------------------|
| do-stop-ati | | dostop-Ø |
| PREP-access-INF | (prefix: do-, verbal | access- |
| | suffix: -a-, verbal | NOM.SG |
| | formant: -ti) | |
| ‘to access’ [<i>zugreifen</i>] | | ‘access’ [<i>Zugriff</i>] |
- (2)
- | | | |
|-------------------------------------|---|------------------------------|
| naslov-iti | | naslov-Ø |
| address-INF | (verbal suffix: -i-, verbal formant: -ti) | address-NOM.SG |
| ‘to address’ [<i>adressieren</i>] | | ‘address’ [<i>Adresse</i>] |
- (3)
- | | | |
|----------------------------------|---|---------------------------------|
| od-mak-niti | | odmik-Ø |
| PREP-move-INF | (prefix: od-, verbal suffix: -ni-, verbal formant: -ti) | removal-NOM.SG |
| ‘to remove’ [<i>entfernen</i>] | | ‘removal’ [<i>Entfernung</i>] |
- (4)
- | | | |
|-------------------------------|---|-------------------------------|
| rab-iti | | rab-a |
| use-INF | (verbal suffix: -i-, verbal formant: -ti) | use-NOM.SG (suffix: -a) |
| ‘to use’ [<i>verwenden</i>] | | ‘usage’ [<i>Verwendung</i>] |
- (5)
- | | | |
|--------------------------------------|--|--|
| po-isk-ati | | isk-anje |
| PREP-consult-INF | (prefix: po-, verbal suffix: -a-, verbal | consult-NOM.SG |
| | formant: -ti) | (suffix: -anje) |
| ‘to consult’ [<i>nachschlagen</i>] | | ‘consultation’ [<i>Nachschlagen</i>] |

4.1.4 Kurzwortbildung

Neben den behandelten Wortbildungsarten Komposition und Derivation werden lexikographische Fachausdrücke auch mittels Kürzung gebildet. Kurzwörter sind Produkte unterschiedlicher Kürzungsvorgänge, bei denen die längere Vollform reduziert wird, sie haben Wortcharakter und werden als Benennungen verwendet (Fleischer und Barz 1995: 218; 2012: 91). Dem Vorteil der Benutzung von Kurzwörtern, der vor allem in der Sprachökonomie liegt, kann man als Nachteil insbesondere einen Mangel an Verständlichkeit und Präzision gegenüber der längeren Form hervorheben. Kurzwörter sind nämlich häufig mehrdeutig, wenn sie sich auf mehrere Fachgebiete und auf mehrere Sprachen beziehen (Arntz, Picht und Mayer 2009: 120). Somit erweisen sie sich als problematisch, da man mit dem gleichen Kurzwort etwas Unterschiedliches benennt.²² Kurzwörter haben manchmal einen differenzierenden Charakter, da ähnliche Benennungen mittels Kurzwörter leichter unterschieden werden. In Verbindung mit der Wortbildung im Bereich der Lexikographie lässt sich die Tendenz feststellen, dass Kurzwörter neben den sprachökonomischen auch andere Funktionen haben. Interessanterweise fungieren sie im Deutschen auch

als Titel, also als Sachnamen für lexikographische Nachschlagewerke, z. B. DWDS (*Digitales Wörterbuch der deutschen Sprache*), WLWF (*Wörterbuch zur Lexikographie und Wörterbuchforschung*), OWID (*Online-Wortschatz-Informationssystem Deutsch*) u. a. Dies gilt nicht nur für das Deutsche, sondern auch für das Slowenische, z. B. SSKJ (*Slovar slovenskega knjižnega jezika*) [*Wörterbuch der slowenischen Standardsprache*], SMeS (*Slovenski medicinski elektronski slovar*) [*Slowenisches elektronisches Wörterbuch der Medizin*] oder *Evroterm* (*Terminološka zbirka izrazov Evropske unije*) [*Terminologische Sammlung von Ausdrücken der Europäischen Union*], GETS (*Geodetski tezaver in slovar*) [*Thesaurus und Wörterbuch der Geodäsie*], wobei der Anteil dieser im Slowenischen im Vergleich zum Deutschen deutlich geringer ist.²³

Hier sei besonders auf die folgenden Kurzformtypen verwiesen:

a) Abkürzungen, die entstehen, wenn eine Benennung nicht in ihrer vollen Form ausgeschrieben wird. Dieser Typ von Kurzformen ist in der lexikographischen Terminologie, besonders in Wiegands Fachtexten häufig vorhanden und wird dort bei grafischen Darstellungen hauptsächlich aus praktischen Gründen eingesetzt, vgl. z. B. *Lematis*. = *Lenmatisierung*, *Homographentr.* = *Homographentrennung*, *exh.* = *exhaustiv*, *initial.* = *initialalphabetisch*, *final.* = *finalalphabetisch* (1998: 205-206). Solche Abkürzungen werden in der Fachliteratur auch Gebrauchsabkürzungen bezeichnet (vgl. Fleischer und Barz 1995: 218). Für das Slowenische sind sie nicht charakteristisch.

b) Initialwörter (Akronyme), bei denen es sich um die aus dem Anfangsbuchstaben mehrerer Wörter entstandene neue Ausdrucksform handelt (vgl. Bußmann 1990: 42). Wenn eine solche Form wie ein Wort ausgesprochen wird, liegt eine (1) „Sprechkürzung“²⁴ vor und wenn nicht, dann handelt es sich um Fälle von (2) „Buchstabierkürzung“ (Arntz, Picht und Mayer 2009: 121). In der lexikographischen und computerlinguistischen Terminologie findet man Beispiele von beidem:

1. Beispiele für „Sprechkürzung“ im Deutschen:

KWIC → *key word in context*, OWID → *Online-Wortschatz-Informationssystem Deutsch*, OBELEX → *Online-Bibliografie zur elektronischen Lexikographie*, COSMAS → *Corpus Search, Management and Analysis System*, OWL → *Web Ontology Language*

1a Beispiel für „Sprechkürzung“ im Slowenischen:

GETS → *Geodetski tezaver in slovar* [*Thesaurus und Wörterbuch der Geodäsie*], SMeS → *Slovenski medicinski slovar* [*Slowenisches elektronisches Wörterbuch der Medizin*]

2. Beispiele für „Buchstabierkürzung“ im Deutschen:

DWDS → *Digitales Wörterbuch der deutschen Sprache*, WLWF → *Wörterbuch zur Lexikographie und Wörterbuchforschung*, DTD → *Document Type Definition*, XML

→ *eXtensible Markup Language*, WDG → *Wörterbuch der deutschen Gegenwartssprache*, MLR → *multifunktionale lexikalische Ressource*

2a Beispiele für „Buchstabierkürzung im Slowenischen:

SSKJ → *Slovar slovenskega knjižnega jezika* [Wörterbuch der slowenischen Standardsprache], SSSJ → *Sprotni slovar slovenskega jezika* [Growing Dictionary of the Slovenian Language],²⁵ SNB → *Slovar novejšega besedja* [Wörterbuch des neueren Wortschatzes]

c) Silbenkurzwörter, die aus dem Anfang des Wortes und weiteren Silben einer Ausgangsbenennung bestehen, eignen sich besonders für die Bildung eines einprägsamen und problemlos aussprechbaren Wortes (Arntz, Picht und Mayer 2009: 121). Diese Art der Wortkürzung ist im Deutschen oft vorhanden (vgl. z. B. *elxiko* → elektronisches, lexikalisch-lexikographisches korpusbasiertes Informationssystem oder *dereko* → das deutsche Referenzkorpus). In diesem und in einigen oben angeführten Beispielen zeigt sich bei der Kurzwortbildung die Tendenz, Namen für Wörterbücher und andere Informationssysteme zu bilden. Somit ist die Kurzwortbildung hier auf Sachnamenbildung ausgerichtet.

4.2 Entlehnung und Lehnwortbildung

Nach Fleischer und Barz (2012: 102) entsteht der Fremdwortschatz im Deutschen „durch Entlehnung „fertiger“ Wörter und durch Wortbildung mit exogenen Elementen innerhalb des Deutschen“. Unter Entlehnung versteht man die unveränderte bzw. weitgehend unveränderte Übernahme eines Lexems aus einer anderen Sprache — auf sprachliches Material wird also außerhalb der gleichen Sprache gegriffen (vgl. Donalies 2002: 16).

Bei der Lehnwortbildung greift man auf die ‘modernen’ Gebersprachen (z. B. das Englische) oder auf die ‘klassischen’ Sprachen zurück (das Griechische oder das Lateinische).²⁶ Die griechischen und lateinischen Wortstämme als exogene Wortbildungseinheiten gelangen ins Deutsche oft auch über andere Sprachen (z. B. über das Englische) (vgl. Fleischer und Barz 2012: 104).

Fleischer und Barz (2012: 102) unterscheiden bei der Lehnwortbildung zwei Gruppen von motivierten Wortbildungen. Zur ersten Gruppe gehören die Bildungen mit exogenen Einheiten und zur zweiten Gruppe gehören Kombinationen aus exogenen und indigenen Einheiten. Im Bereich der lexikographischen Terminologie gehören zur ersten Gruppe Lexeme wie *computerextern*, *Kontentvalidierung*, *Hyperlink*, *Reliabilität* u. a. Für die zweite Gruppe lässt sich bezüglich der Lehnwortbildung beobachten, dass sich ein exogenes Erstglied oft mit einem indigenen Zweitglied (oft sind das Substantive) verbindet. Dies lässt sich an folgenden Beispielen der lexikographischen Termini beobachten: *Mediostruktur*, *Mikrostruktur*, *Mikrostrukturanzeiger*, *Makrostruktur*. Diese Teilung auf zwei Gruppen gilt auch für das Slowenische, wobei sich die deutschen und die slowenischen Belege bezüglich der exogenen und indigenen Einheiten oft

nicht überlappen. Zur ersten Gruppe gehören z. B. (1) *mediostruktura* [*Mediostruktur*], (2) *mikrostruktura* [*Mikrostruktur*], (3) *makrostruktura* [*Makrostruktur*] und zur zweiten Gruppe z. B. (4) *hiperpovezava* [*Hyperlink*].

- | | |
|---|--|
| <p>(1)
 mediostruktur-a
 middle.structure-F.NOM.SG
 ‘middle structure’ [<i>Mediostruktur</i>]</p> | <p>(2)
 mikrostruktur-a
 very small.structure-F.NOM.SG
 ‘very small structure’ [<i>Mikrostruktur</i>]</p> |
| <p>(3)
 makrostruktur-a
 large.structure-F.NOM.SG
 ‘large structure’ [<i>Makrostruktur</i>]</p> | <p>(4)
 hiperpovezav-a
 beyond the usual size.link-F.NOM.SG
 ‘link beyond the usual size’ [<i>Hyperlink</i>]</p> |

Auch in der Fremdwortbildung unterscheidet man die Wortbildungsarten der Komposition, Derivation, Konversion und Kurzwortbildung. Hier sei nur kurz auf die Möglichkeiten der Komposition und Derivation im Bereich der lexikographischen Terminologie eingegangen. Bei der Komposition unterscheidet man zwei Gruppen, u. z. 1. Komposita aus zwei wortfähigen Konstituenten, von denen jeweils eine oder zwei exogen sind und 2. Komposita aus zwei Konfixen oder aus einem Konfix und einer wortfähigen Konstituente (Fleischer und Barz 2012: 111).²⁷

1. Komposita mit einer bzw. zwei exogenen Konstituenten im Deutschen:
Portallemmaliste, Computerunterstützung, Konsultationspraxis

Im Slowenischen werden die angegebenen Termini unterschiedlich realisiert. Zum einen ist die Gebersprache für die slowenische Terminologie das Deutsche (*praksa dostopanja* [*Konsultationspraxis*], *seznam iztočnic portala* [*Portallemmaliste*]), zum anderen das Englische (*računalniška podpora* [*computer support*]). Auch bezüglich der Wortbildungsverfahren lassen sich Unterschiede nachweisen, u. z. werden diese Termini im Slowenischen als attributive Wortverbindungen realisiert — mit dem exogenen Sprachmaterial: (1) *seznam iztočnic portala*, (2) *praksa dostopanja* oder ohne: (3) *računalniška podpora*.

- | | |
|--|---|
| <p>(1)
 seznam-Ø iztočnic-Ø portal-a
 list- lemma- portal-
 M.NOM.SG GEN.PL GEN.SG
 ‘list of portal lemmata’ [<i>Portallemmaliste</i>]</p> | <p>(2)
 praks-a dostopanj-a
 practice-F.NOM.SG access-GEN.SG
 ‘practice of access’ [<i>Konsultationspraxis</i>]</p> |
| <p>(3)
 računalnišk-a podpor-a
 computational-F.NOM.SG support-F.NOM.SG
 ‘computational support’ [<i>Computerunterstützung</i>]</p> | |

2. Konfixkomposita aus einem bzw. aus zwei Konfixen und einer wortfähigen Konstituente im Deutschen:
Mikroarchitektur, Metalexikographie, Metatext, Kontext, Mediostruktur, Mikrostruktur, Makrostruktur

Im Slowenischen bestehen die Zusammensetzungen entweder aus einer oder aus mehreren exogenen unmittelbaren Konstituenten *mikro-/meta-/medio-/makro-*: (1) *mikrozgradba* [*Mikrostruktur*], (2) *metabesedilo* [*Metatext*], (3) *mediostruktura* [*Mediostruktur*], (4) *mikrostruktura* [*Mikrostruktur*], (5) *makrostruktura* [*Makrostruktur*]. Zudem gibt es Zusammensetzungen mit der indigenen unmittelbaren Konstituente *sam-* in (6) *samodejen* [*automatisch*] und *sestavljanke* mit der indigenen Konstituente *so-*: (7) *sobesedilo* [*Kontext*].

- | | |
|--|---|
| <p>(1)
mikrozgradb-a
very small.structure-F.NOM.SG
'very small structure' [<i>Mikrostruktur</i>]</p> | <p>(2)
metabesedil-o
at a higher level.text-N.NOM.SG
'text at a higher level' [<i>Metatext</i>]</p> |
| <p>(3)
mediostruktura-a
middle.structure-F.NOM.SG
'middle structure' [<i>Mediostruktur</i>]</p> | <p>(4)
mikrostruktura-a
very small.structure-F.NOM.SG
'very small structure' [<i>Mikrostruktur</i>]</p> |
| <p>(5)
makrostruktura-a
large.structure-F.NOM.SG
'large structure' [<i>Makrostruktur</i>]</p> | <p>(6)
samodej-en
self-acting-ADJ.M.NOM
'self-acting' [<i>automatisch</i>]</p> |
| <p>(7)
sobesedil-o
surrounding.text-N.NOM.SG
'context' [<i>Kontext</i>]</p> | |

Bezüglich der Derivation gibt es die folgenden Möglichkeiten der Kombination von exogenen und indigenen Wortbildungseinheiten, wobei hier die Grenze zwischen der Fremdwortbildung und indigener Wortbildung fließend ist (vgl. Fleischer und Barz 2012: 112):

1. exogene Basis und indigenes Suffix:

-ung (*Digitalisierung*), *-heit* (*Differenziertheit*), *-schaft* (*Autorschaft*), *-al* (*digital*), *-e* (*solide*), *-el(l)* (*flexibel*, *maschinell*), *-ion* (*Annotation*, *Kollokation*) *-isch* (*typographisch*, *symmetrisch*), *-ieren* (*definieren*, *strukturieren*, *analysieren*).

Im Slowenischen gibt es die exogene oder die indigene Basis mit indigenem bzw. exogenem Suffix. Exogene Basis und exogenes bzw. indigenes Suffix weisen folgende Termini auf: (1) *diferenciranost* [*Differenziertheit*], (2) *avtorstvo* [*Autorschaft*], (3) *digitalen* [*digital*], (4) *fleksibilen* [*flexibel*], (5) *anotacija* [*Annotation*], (6) *kolokacija* [*Kollokation*], (7) *tipografski* [*typographisch*], (8) *simetričen* [*symmetrisch*], (9) *definirati* [*definieren*], (10) *strukturirati* [*strukturieren*], (11) *analizirati* [*analysieren*] (12) *digitalizacija* [*Digitalisierung*]; indigene Basis und indigenes Suffix weisen folgende Belege auf: (13) *zanesljiv* [*solide*], (14) *strojen* [*maschinell*].

- | | |
|--|---|
| <p>(1)
diferenciran-ost
differentiated-F.NOM.SG
'differentiation' [<i>Differenziertheit</i>]</p> | <p>(2)
avtor-stvo
author-N.NOM.SG
'authorship' [<i>Autorschaft</i>]</p> |
|--|---|

- | | |
|---|---|
| (3)
digital-en
digitize-ADJ.M.NOM
'digital' [<i>digital</i>] | (4)
fleksibil-en
flexibility-ADJ.M.NOM
'flexible' [<i>flexibel</i>] |
| (5)
anot-acija
annotate-NOM.SG
'annotation' [<i>Annotation</i>] | (6)
kolok-acija
collocate-NOM.SG
'collocation' [<i>Kollokation</i>] |
| (7)
tipograf-ski
typography-ADJ.M.NOM
'typographic' [<i>typographisch</i>] | (8)
simetr-ičen
symmetry-ADJ.M.NOM
'symmetric' [<i>symmetrisch</i>] |
| (9)
defin-irati
definition-INF
'to define' [<i>definieren</i>] | (10)
struktur-irati
structure-INF
'to structure' [<i>strukturieren</i>] |
| (11)
analiz-irati
analysis-INF
'to analyse' [<i>analysieren</i>] | (12)
digitaliz-acija
digitize-NOM.SG
'digitalization' [<i>Digitalisierung</i>] |
| (13)
zanes-ljiv
rely.on-ADJ.M.NOM
'reliable' [<i>solide</i>] | (14)
stroj-en
machine-ADJ.M.NOM
'machine-assisted' [<i>maschinell</i>] |

2. indigene Basis und exogenes Suffix: *-ieren* (*alphabetisieren*)

Im Slowenischen ist sowohl die Basis als auch das Suffix exogen: *alfabetirati*. Dieser Terminus wird jedoch im Slowenischen meist als Wortverbindung (1) *abecedna razporeditev* [*alphabetische Anordnung*] realisiert.

- (1)
abeced-na razporeditev-Ø
alphabet-ADJ.F.NOM arrangement-F.NOM.SG
'alphabetical arrangement' [*alphabetische Anordnung*]

3. exogene Basis und indigenes Präfix: *un-* (*unstrukturiert, unmarkiert*)

Bei den folgenden Beispielen handelt es sich im Slowenischen um *sestavljanje* bzw. *sestava* mit dem indigenen Präfix *ne-* und der exogenen bzw. indigenen Basis: (1) *nestrukturiran* [*unstrukturiert*], (2) *neoznačen* [*unmarkiert*].

- | | |
|---|---|
| (1)
ne-struktur-iran
NEG-structure-ADJ.M.NOM (prefix: ne-)
'unstructured' [<i>unstrukturiert</i>] | (2)
ne-označ-en
NEG-mark-ADJ.M.NOM (prefix: ne-)
'unmarked' [<i>unmarkiert</i>] |
|---|---|

4. indigene Basis und exogenes Präfix:
in- (indirekt), *sub-* (Subkommentar)

Im Slowenischen geht es um Beispiele von *sestavljene* oz. *sestava*, u. z. mit der indigenen Basis und dem indigenen Präfix: (1) *neposreden* [indirekt], (2) *podkomentar* [Subkommentar].

(1)	(2)
ne-posred-en	pod-komentar-Ø
NEG-direct-ADJ.M.NOM (prefix: ne-)	PREP-comment-M.NOM.SG (prefix: pod-)
'indirect' [indirekt]	'sub-comment' [Subkommentar]

Die Mehrheit der neuen Termini bringt heute die technologische Entwicklung mit sich, also ein neuer Begriff führt zum neuen Terminus. Es passiert zudem oft, dass man mit der Übernahme einer technischen Neuentwicklung auch die Benennung aus der betreffenden Sprache mit übernimmt. Besonders in den letzten Jahren könnte man eine Menge der englischen Fachwörter auf einigen sich schnell entwickelnden Fachgebieten beobachten, die ins Deutsche übernommen wurden. Entlehnung aus modernen Sprachen erweist sich als die produktivste Art der Entlehnung (vgl. Vintar 2008: 51). Die schnelle Entwicklung von Informationswissenschaften sorgt dafür, dass Termini aus dem englischen Sprachraum übernommen werden, was sich konkret am Anteil der englischen Termini im Fachbereich der Computerlexikographie widerspiegelt (*Link, Marker, Parser, Semantic Web, Tag, Tagging*²⁸ und *usage notes, Extensible Markup Language, guide word, Hyperlink* u. a.). Im Slowenischen werden zu vielen englischen Entsprechungen eigene Termini gebildet, z. B. *link* → *povezava*, *marker* → *označevalnik*, *parser* → *razčlenjevalnik*, *semantic web* → *semantični splet*, *tag* → *značka*, *tagging* → *označevanje*, *usage notes* → *oznake*, *Extensible Markup Language* → *razširljivi označevalni jezik*, *hyperlink* → *nadpovezava*. Cabré (1999: 90) führt an, dass das Entleihen von Termini oft unnötig ist, da die Nehmersprachen selbst über das Potenzial verfügen, eigene Termini zu bilden.

Im Gegensatz zur Lehnwortbildung überträgt man bei der Lehnübersetzung die einzelnen Wortelemente in die Zielsprache ohne die innere Struktur der Benennung zu verändern (z. B. *Computer Aided Translation* → *maschinengestützte Übersetzung* [računalniško podprto prevajanje], *conceptual graph* → *konzeptueller Graph* [konceptualni graf], *artificial intelligence* → *künstliche Intelligenz* [umetna inteligenca], *extensible markup language* → *erweiterbare Auszeichnungssprache* [razširljivi označevalni jezik]. Solche Lehnübersetzungen bedient man sich im Slowenischen besonders dann, wenn in Anlehnung an das deutsche lexikographische System komplexe Termini im Slowenischen gebildet werden sollen, z. B. *lexikographische Datenauswertung* → (1) *slovaropisna analiza podatkov*, *Datendistributionsstruktur* → (2) *struktura porazdelitve podatkov*, *hierarchische vollständig erweiterte Mikrostruktur* → (3) *hierarhična popolnoma razširjena mikrostruktura*, *konkrete hierarchische linke Kernstruktur* → (4) *konkretna hierarhična leva jedrna struktura* u. a.

- (1)
slovaropisn-a analiz-a podatkov-ov
lexicographical- analysis- data-GEN.PL
F.NOM.SG F.NOM.SG
'lexicographical analysis of data' [*lexikographische Datenauswertung*]
- (2)
struktur-a porazdelitv-e podatkov-ov
structure-F.NOM.SG distribution-GEN.SG data-GEN.PL
'structure of data distribution' [*Datendistributionsstruktur*]
- (3)
hierarhičn-a popolnoma razširjen-a mikrostruktur-a
hierarchical- fully expanded- micro-structure-
F.NOM.SG F.NOM.SG F.NOM.SG F.NOM.SG
'hierarchical fully expanded microstructure' [*hierarchische vollständig erweiterte Mikrostruktur*]
- (4)
konkretn-a hierarhičn-a lev-a jedrn-a struktur-a

concrete- hierarchical- left- core- structure-
F.NOM.SG F.NOM.SG F.NOM.SG F.NOM.SG F.NOM.SG
'concrete hierarchical left core structure' [*konkrete hierarchische linke Kernstruktur*]

In einigen Fällen wird im Deutschen nur ein Teil der Benennung unverändert übernommen — gewöhnlich ist das die erste Konstituente — und der andere Teil wird übersetzt (*Log-File-Analyse*, *Log-File-Protokoll*, *Bottom-up-Lexikographie*, *Offline-Wörterbuch*, *Think-Aloud-Protokoll*, *Markup-Sprache*). Das Gleiche lässt sich auch für die aus dem Deutschen entlehnte lexikographische Terminologie im Slowenischen beobachten: *glattalphabetische Makrostruktur* → *tekočebecedna makrostruktura*, *nicht-typographischer Strukturanzeiger* → *netipografski strukturni kazalnik*, *computergestützte Lexikographie* → *računalniško podprta leksikografija* u. a. Zu erwähnen ist hier auch der Fall der sogenannten internen Entlehnung, d. h. der Übergang bzw. die Übernahme einer Benennung in eine andere Fachsprache zur Bezeichnung eines verwandten Begriffs. Dies führt zur Mehrdeutigkeit bzw. Polysemie (vgl. z. B. den deutschen Terminus *Stichwort* und den slowenischen Terminus *geslo*).

5. Zusammenfassende Bemerkungen

Im Hinblick auf die lexikographische Terminologie lassen sich folgende allgemeine Charakteristika beobachten. Durch die Verfahren der Komposition und Derivation wird die lexikographische Terminologie durch Wortschatzerweiterung ausgebaut, Spezifizierung führt zur Erhöhung der Genauigkeit in der lexikographischen Terminologie. Durch das Vorgehen der Kürzung erfolgt die Steigerung der Wirtschaftlichkeit. Konversion resultiert durch den Verzicht des Agens in der Erhöhung der Anonymität. Auf der Basis der Wortbildungsverfahren lassen sich konkret folgende Charakteristika anhand des kontrastiven

Vergleichs der deutschen und der slowenischen lexikographischen Terminologie feststellen: Deutsche Komposita werden im Slowenischen als attributive bzw. als präpositionale Wortverbindungen realisiert. Zusammensetzungen sind im Slowenischen zwar vorhanden, sie haben jedoch eine untergeordnete Rolle. Daneben hat das Deutsche grundsätzlich die Möglichkeit, relativ lange und komplexe Einwortbenennungen zu bilden, wobei diese ins Slowenische meist übersetzt werden, was zugleich eine Abweichung vom Formaspekt des Terminus bedeutet. Bezüglich der Derivation lässt sich beobachten, dass deutsche Derivate im Slowenischen unterschiedlich realisiert werden. Im Vergleich zum Deutschen gibt es bezüglich der Konversion durch die Sprachstruktur bedingte Unterschiede. Typisch für das Slowenische ist, dass Verben gewöhnlich als Basis für Substantivableitungen dienen, u. z. mit oder ohne Wortbildungsaffixe. Kurzwortbildung ist in beiden Sprachen realisiert. Bei der Lehnwortbildung lässt sich zudem beobachten, dass sich die deutschen und die slowenischen Belege bezüglich der exogenen und indigenen Einheiten oft nicht überlappen. Bei der Lehnwortbildung ist neben dem Deutschen auch das Englische (besonders in der Computerlexikographie) die Gebersprache für die slowenische Terminologie. Im Slowenischen werden zu vielen englischen Entsprechungen eigene Termini gebildet. Mithilfe der Lehnübersetzung werden, wie oben angegeben, in Anlehnung an das deutsche lexikographische System komplexe Termini im Slowenischen gebildet. Aus dem Angeführten ist ersichtlich, dass man nur für die deutschen Komposita und Kurzwortbildungen Muster für das Überführen der deutschen lexikographischen Terminologie ins Slowenische aufstellen kann, für die anderen Verfahren gilt dies nicht, was auch die oben besprochenen Belege zeigen.

Die behandelten lexikographischen Termini im Deutschen, die Termini bzw. Lexeme, die auf dem Fachgebiet der Lexikographie verwendet werden und die entsprechenden Äquivalente im Slowenischen sind in der folgenden Tabelle angeführt:²⁹

Lexikographische Termini im Deutschen und die Termini/Lexeme, die auf dem Fachgebiet der Lexikographie verwendet werden (alphabetisch angeordnet)	Slowenische Äquivalente zu den lexikographischen Termini im Deutschen und zu den, auf dem Fachgebiet der Lexikographie verwendeten Termini/Lexemen
Abkürzungswörterbuch	krajšavni slovar
abstrakte hierarchische linke Kernstruktur	abstraktna hierarhična leva jedrna struktura
adressieren	nasloviti
Adresse	naslov
Adressierung	naslavljanje
alphabetische Anordnung	abecedna razporeditev
analysieren	analizirati
Annotation	anotacija

Antonymenwörterbuch	slovar antonimov
Äquivalentangabe	podatek o ustreznici
Archaismenwörterbuch	slovar arhaizmov
Artikelstrukturschema	shema strukture slovarskega članka
Außentext	zunanje besedilo
Aussprachewörterbuch	izgovorni slovar
automatisch	samodejen
Autorschaft	avtorstvo
basaler Wörterbuchartikel	temeljni slovarski članek
Basis	osnova
bazaler	temeljni
beabsichtigen	nameravati
bearbeiten	obdelati
Belegbeispielangabe	zgled iz dejanske rabe
benutzen	uporabiti
Benutzer	uporabnik
computergestützte Lexikographie	računalniško podprta leksikografija
Computerunterstützung	računalniška podpora
Datei	datoteka
Datenbank	podatkovna baza
Datendistributionsstruktur	strukturna porazdelitve podatkov
definieren	definirati
Differenziertheit	diferenciranost
digital	digitalen
Digitalisierung	digitalizacija
einfache Mikrostruktur	enostavna mikrostruktura
entfernen	odmakniti
Entfernung	odmik
erfolglos	neuspešen
erweiterbare Auszeichnungssprache	razširljivi označevalni jezik
erweiterte Mikrostruktur	razširjena mikrostruktura
Experte	strokovnjak
Fachgebietsangabe	podatek o stroki
Fachlexikographie	strokovna leksikografija

fachlich	strokovni
fachliches Sachwörterbuch	stvarni strokovni slovar
fachliches Sprachwörterbuch	jezikovni strokovni slovar
Fachwörterbuchartikel	slovarski članek strokovnega slovarja
flexibel	fleksibilen
Formkommentar	oblikovni komentar
Fremdwörterbuch	slovar tujk
glattalphabetische Makrostruktur	tekočeebecedna makrostruktura
Häufigkeit	pogostost
Hauptlemma	glavna iztočnica
hierarchisch vollständig erweiterte Mikrostruktur	hierarhično popolnoma razširjena mikrostruktura
Hyperlink	hiperpovezava
indirekt	neposreden
inner	notranji
interfachlich	medstrokovni
Kollokation	kolokacija
konkrete hierarchische linke Kernstruktur	konkretna hierarhična leva jedrna struktura
konkrete hierarchische rechte Kernstruktur	konkretna hierarhična desna jedrna struktura
Konsultationspraxis	praksa dostopanja
Kontext	sobesedilo
künstliche Intelligenz	umetna inteligenca
Leitelement	vodilni element
Lemma	iztočnica
Lemmakandidat	kandidat za iztočnico
Lemmakandidatenliste	seznam kandidatov za iztočnico
Lemmaliste	seznam iztočnic
lemmatisieren	lematizirati
Lemmavariante	različica iztočnice
Lemmzeichen	iztočnični znak
Lemmzeichengestaltangabe	oblika iztočničnega znaka
lexikographisch	leksikografski
lexikographische Datenauswertung	slovaropisna analiza podatkov
lexikographische Datenerhebung	slovaropisno pridobivanje podatkov
Liste	seznam

Makrostruktur	makrostruktura
Markierung	označevanje
maschinell	strojen
maschinengestützte Übersetzung	računalniško podprto prevajanje
Mediostruktur	mediostruktura
Metatext	metabesedilo
Mikrostruktur	mikrostruktura
monoinformatives Wörterbuch	enoizborni slovar
nachschlagen	poiskati
Nachschlagen	iskanje
Nachspann	dodatek
Neologismenwörterbuch	slovar neologizmov
nestalphabetische Makrostruktur	prekinjena tekočabecedna makrostruktura
Nestanordnung	gnezdenje
nicht-typographischer Strukturanzeiger	netipografski strukturni kazalnik
nischenalphabetische Makrostruktur	besedotvorno tekočabecedna makrostruktura
Nutzer	uporabnik
Phrasenangabe	podatek o frazemu
Planungsphase	načrtovalna faza
polyinformatives Wörterbuch	večizborni slovar
Portallemmaliste	seznam iztočnic portala
Primärquelle	primarni vir
prüfbar	preverljiv
Prüfbarkeit	preverljivost
publizieren	objaviti
Rezeptionswörterbuch	repcijski slovar
Sachwörterbuch	stvarni slovar
Sekundärquelle	sekundarni vir
solide	zanesljiv
sortieren	urejati
Spezialwörterbuch	posebni slovar
Sprachstadienwörterbuch	slovar jezikovnega obdobja
Sprachwörterbuch	jezikovni slovar
standardisieren	standardizirati

Stichwort	geslo
strukturieren	strukturirati
Subkommentar	podkomentar
symmetrisch	simetričen
Synonymenangabe	podatek o sinonimu
Synonymenwörterbuch	slovar sinonimov
textbezogene Wörterbuchfunktion	slovarska funkcija v zvezi z besedilom
typographisch	tipografski
typographischer Strukturanzeiger	tipografski strukturni kazalnik
Typologiegraph	tipološki graf
überprüfen	preveriti
umschreiben	opisati
umverteilen	prerazporediti
unabhängig	neodvisen
unmarkiert	neoznačen
unmittelbar	neposreden
unstrukturiert	nestrukturiran
unterstützen	podpreti
Urbeleg	prapimer
Urliste	praseznam
verändern	spremeniti
vermitteln	posredovati
verweisen	usmerjati
Verweisstruktur	kazalčna struktura
verwenden	rabiti
Verwendung	raba
Verzeichnis	seznam
Vorspann	uvodni del
wissensbezogene Wörterbuchfunktion	slovarska funkcija v zvezi z znanjem
Wörterbuchartikel	slovarski članek
Wörterbuchbenutzungsforschung	raziskovanje rabe slovarja
Wörterbuchbenutzungstest	test rabe slovarja
Wörterbuchfunktion	funkcija slovarja
Wörterbuchtyp	vrsta slovarja

zergliedern	razčleniti
zugreifen	dostopati
Zugriff	dostop
Zugriffsstruktur	struktura dostopanja

Tabelle 1: Lexikographische Termini und die auf dem Fachgebiet der Lexikographie verwendeten Termini/Lexeme im Deutschen mit den entsprechenden Äquivalenten im Slowenischen

Anmerkungen

- Poethe führt an, dass der Grad der Normierung in den Geisteswissenschaften wesentlich niedriger ist als dies in den Naturwissenschaften der Fall ist (Poethe 2000: 202).
- Terminologische Schulen* unterschieden sich nach unterschiedlichen Ansätzen in der Praxis und unterschiedlichen Forschungsinteressen (vgl. Laurén und Picht 2006: 167).
- Interessanterweise sind einige lexikographische Termini, darunter z. B. *Lemmazeichen* und *Lemmazeichengestaltangabe* im Slowenischen nicht belegt. Ins Deutsche wurden sie von Wiegand (1983) eingeführt.
- Einige Bemerkungen zum Gebrauch von *Daten* und *Informationen* bei Wiegand (1998: 160-171).
- Darunter vor allem die Fachpublikationen *Slavistična revija* und *Jezik in slovstvo* und die Monographie *Terminologija in sodobna terminografija* (2009). Seit 2013 erscheint die Fachpublikation *Slovenščina 2.0*.
- Lehrbücher für den universitären Bereich, wie z. B. *Lexikographie und Wörterbuchbenutzung* (Engelberg und Lemnitzer 2009) und *A Handbook of Lexicography: The Theory and Practice of Dictionary-making* (Svensén 2009) gibt es im Slowenischen nicht.
- Dies gelang mit dem *Wörterbuch zur Lexikographie und Wörterbuchforschung* (2010), womit ein wichtiger Beitrag im Sinne der Sammlung und Bearbeitung von lexikographischen Termini im Deutschen und Äquivalenten in neun Sprachen geleistet wurde und das Resultat internationaler Zusammenarbeit ist.
- Mehr über die Tätigkeit des Instituts unter: <http://isjfr.zrc-sazu.si/#v> [letzter Zugriff: 1. 8. 2016].
- Mehr über die Tätigkeit des Instituts unter: <http://www.ijs.si/ijsw/E8> [letzter Zugriff: 1. 8. 2016].
- Mehr zur Tätigkeit von *Trojina* unter: <http://www.trojina.si/> [letzter Zugriff: 1. 8. 2016].
- Mehr über die Tätigkeit des Unternehmens unter: <http://www.amebis.si/> [letzter Zugriff: 1. 8. 2016].
- SP (2001) = *Slovenski pravopis* [Slowenisches Rechtschreibwörterbuch]
- SSKJ2 (2014) = *Slovar slovenskega knjižnega jezika 2* [Wörterbuch der slowenischen Standardsprache 2]
- SSKJ (1994) = *Slovar slovenskega knjižnega jezika* [Wörterbuch der slowenischen Standardsprache]
- Bei allen mit dem Asteriskus und Fragezeichen gekennzeichneten Belegen handelt es sich um terminologische Vorschläge.

16. Viele der angeführten Ein- und Mehrwortbenennungen in der deutschen lexikographischen Terminologie gehen auf Wiegand zurück, der besonders für den Bereich der theoretischen Lexikographie viele beitrug.
17. Hier wird von der deutschen theoretischen Basis ausgegangen, deswegen ist die deutsche lexikographische Terminologie der Ausgangspunkt für die Bildung der fehlenden lexikographischen Terminologie im Slowenischen. Ohne eine theoretische Grundlage ist keine systematische Terminologie möglich.
18. Bei entlehnten Termini kann der Fall auftreten, dass Fachwörter, die den ursprünglichen lateinischen oder griechischen Terminus ersetzen, als Doubletten bestehen bleiben.
19. Der Terminus *sestaoljanje* geht auf Toporišič zurück (2000: 159-160) und *sestava* auf Vidovič Muha (2011: 333).
20. Für die Zirkumfixderivation mit Substantiven konnten keine Belege gefunden werden.
21. Eichinger (2000: 73) zählt zu der morphologischen Konversion auch deverbale Konversionen, die mit einem Wechsel des Stammvokals verbunden sind, z. B. *binden* → *Band*.
22. Das Kurzwort *OPAL* ist 1. *Objektbezogene Preis-, Aufwands- und Leistungsermittlung* oder 2. *Online publizierte Arbeiten zur Linguistik* am IDS Mannheim. Das Kurzwort wird also auf zwei unverwandten Fachgebieten verwendet, wo es Unterschiedliches benennt.
23. Manchmal handelt es sich bei den Namen von Wörterbüchern, Datenbanken und anderen lexikographischen Produkten um Eponyme, vgl. z. B. *Duden*, *Wahrig* im Deutschen und *Fran* im Slowenischen.
24. Diese Formen sind leicht aussprechbar und wirken interessant. Bei solchen Formen handelt es sich oft um Produktnamen.
25. Dieses Wörterbuch enthält neue, noch nicht lexikalisierte Wörter.
26. Lehnwortbildung auf der Basis der gräkolateinischen Wortelemente erweist sich als vorteilhaft, da es vor allem in den Naturwissenschaften bereits eine gängige Praxis ist, sprachliches Material aus den klassischen Sprachen zu entlehnen (vgl. z. B. die botanischen Termini). Hier sei zudem bemerkt, dass die Begriffe 'klassisch' und 'modern' für die Bezeichnung von gräkolateinischen und englischen Elementen eher ungeschickt und falsch ist (Eins 2008: 4). Eins verweist nämlich darauf, dass sich die gräkolateinischen Elemente auch anderer Wortbildungsmuster als nur 'klassischen' bedienen können und die englischen Elemente auch zur Verbreitung der gräkolateinischen Konstituenten beigetragen haben (ebd.).
27. Vgl. auch oben unter Wortbildung.
28. Alle Beispiele stammen aus dem Glossar des Nachschlagewerks *Computerlexikographie* (Kunze und Lemnitzer 2007).
29. Als Quelle für die lexikographischen Termini im Deutschen dienten das *Wörterbuch zur Lexikographie und Wörterbuchforschung* (2010), *Lexikographie und Wörterbuchbenutzung* (Engelberg und Lemnitzer 2009), *Computerlexikographie* (Kunze und Lemnitzer 2007) und die Fachpublikation *Lexicographica*. Als Quelle für die slowenischen Äquivalente zu den deutschen lexikographischen Termini dienten die Beiträge der Fachpublikationen *Slavistična revija*, *Jezik in slovo* und die Außentexte der slowenischen Wörterbücher (zugänglich über das Portal *Fran*: <http://www.fran.si/> [letzter Zugriff: 1. 8. 2016]). Bei einigen Äquivalenten im Slowenischen handelt es sich um Neubildungen.

Literatur

- Arntz, Reiner, Heribert Picht und Felix Mayer.** 2009. *Einführung in die Terminologearbeit*. 6. Aufl. Hildesheim/Zürich/New York: Georg Olms.
- Bußmann, Hadumod.** 1990. *Lexikon der Sprachwissenschaft*. Stuttgart: Alfred Kröner.
- Butina-Koller, Ekaterina et al.** 2010. Vorwort der Mitarbeiter zur terminologischen Äquivalenz. Wiegand, Herbert Ernst et al. (Hrsg.). 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung* (Band 1 A–C): XXV–XXX. Berlin/New York: Walter de Gruyter.
- Cabré, Maria Teresa.** 1999. *Terminology. Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins.
- Donalies, Elke.** 2002. *Die Wortbildung des Deutschen. Ein Überblick*. Tübingen: Gunter Narr.
- Eichinger, Ludwig M.** 2000. *Deutsche Wortbildung. Eine Einführung*. Tübingen: Narr.
- Eins, Wieland.** 2008. *Muster und Konstituenten der Lehnwortbildung: Das Konfix-Konzept und seine Grenzen*. Hildesheim: Georg Olms.
- Engelberg, Stefan und Lothar Lemnitzer.** 2009. *Lexikographie und Wörterbuchbenutzung*. 4. Aufl. Tübingen: Stauffenburg.
- Fleischer, Wolfgang und Irmhild Barz.** 1995. *Wortbildung der deutschen Gegenwartssprache*. 2. Aufl. Tübingen: Max Niemeyer.
- Fleischer, Wolfgang und Irmhild Barz.** 2012. *Wortbildung der deutschen Gegenwartssprache*. 4. Aufl. Berlin/Boston: Walter de Gruyter.
- Fran. Slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU.* <http://www.fran.si/> [letzter Zugriff: 1. 8. 2016].
- Jesenšek, Vida.** 2013. Einführung. Fachsprachenlexikographie und Terminographie: eine kritisch-analytische Betrachtung der Theorie und Praxis. Jesenšek, Vida (Hrsg.). 2013. *Specialised Lexicography*: 11–20. Berlin/Boston: Walter de Gruyter.
- Jezik in slovstvo.* <http://www.jezikinslovstvo.com> [letzter Zugriff: 1. 8. 2016].
- Kunze, Claudia und Lothar Lemnitzer.** 2007. *Computerlexikographie — Eine Einführung*. Tübingen: Gunter Narr.
- Laurén, Christer und Picht, Heribert.** 2006. Approaches to Terminological Theories: A Comparative Study of the State-of-the-Art. Picht, Heribert (Hrsg.). 2006. *Modern Approaches to Terminological Theories and Applications*: 163–184. Bern/Berlin u. a.: Peter Lang.
- Ledinek, Nina et al. (Hrsg.).** 2009. *Terminologija in sodobna terminografija*. Ljubljana: Založba ZRC, ZRC SAZU.
- Poethe, Hannelore.** 2000. Fachsprachliche Aspekte der Wortbildung. Die Leistung der Wortbildung für Fachsprache und Fachtext. Barz, Irmhild, Marianne Schröder und Ulla Fix (Hrsg.). 2000. *Praxis- und Integrationsfelder der Wortbildungsforschung*: 199–218. Heidelberg: Winter.
- Slavistična revija.* <http://www.srl.si/> [letzter Zugriff: 1. 8. 2016].
- Slovar slovenskega knjižnega jezika.* 1994. Ljubljana: SAZU, ZRC SAZU in Inštitut za slovenski jezik Frana Ramovša.
- Slovar slovenskega knjižnega jezika 2.* 2014. SAZU, ZRC SAZU in Inštitut za slovenski jezik Frana Ramovša.
- Slovenski pravopis.* 2001. Ljubljana: SAZU, ZRC SAZU in Inštitut za slovenski jezik Frana Ramovša.

- Svensén, Bo.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge u. a.: Cambridge University Press.
- Toporišič, Jože.** 2000. *Slovenska slovnica*. 4. Aufl. Maribor: Obzorja.
- Vidovič Muha, Ada.** 2011. *Slovensko skladiščno besedotvorje*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Vintar, Špela.** 2008. *Terminološka veda in računalniško podprta terminografija*. Ljubljana: Filozofska fakulteta.
- Wiegand, Herbert Ernst.** 1983. Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. Wiegand, Herbert Ernst (Hrsg.). 1983. *Studien zur neuhochdeutschen Lexikographie III*: 401-474. Germanistische Linguistik 1-4/82. Hildesheim u. a.: Georg Olms Verlag.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Mit 159 Illustrationen im Text. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst et al. (Hrsg.).** 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung*. (Band 1 A-C). Berlin/New York: Walter de Gruyter.

Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-UVa*

Sven Tarp, *International Centre for Lexicography, Universidad de Valladolid, Spain; Department of Afrikaans and Dutch, University of Stellenbosch, South Africa and Centre for Lexicography, University of Aarhus, Aarhus, Denmark (st@asb.dk)*

and

Pedro A. Fuertes-Olivera, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa and International Centre for Lexicography, Universidad de Valladolid, Valladolid, Spain (pedro@emp.uva.es)*

Abstract: This paper initially discusses some of the consequences which the technological development has for lexicography, especially in terms of the different types of empirical basis which can be used in dictionary projects. The most important advantages and disadvantages of using the Internet as a corpus are then listed and compared to the usefulness of "traditional" corpora. As an example, the paper shows how the Internet is used as the main empirical source in order to select lemmata and meaning items in the Online Dictionaries of Spanish Valladolid-UVa. The methods and tools employed in the project are discussed together with the requirements to the lexicographers' competences, knowledge and skills. Finally, the paper provides some general conclusions as well as recommendations and hypotheses for future lexicographical work and research.

Keywords: INTERNET LEXICOGRAPHY, ONLINE LEXICOGRAPHY, LEXICOGRAPHICAL METHODOLOGY, EMPIRICAL BASIS, LEMMA SELECTION, MEANING SELECTION, LEXICOGRAPHICAL DATABASES, SPANISH DICTIONARIES, MONOLINGUAL DICTIONARIES, GENERAL DICTIONARIES

Opsomming: Voordele en nadele van die gebruik van die internet as 'n korpus: Die geval van die Aanlynwoordeboeke van Spaans Valladolid-UVa.

Hierdie artikel bespreek aanvanklik sommige van die gevolge wat tegnologiese ontwikkeling vir die leksikografie inhou, veral in terme van die verskillende soorte empiriese basisse wat vir woordeboekprojekte gebruik kan word. Die belangrikste voordele en nadele van die gebruik van die internet as 'n korpus word dan gelyk en vergelyk met die nuttigheid van "tradisionele" korpora. As voorbeeld toon die artikel hoe die internet as die belangrikste empiriese bron gebruik word om lemmata en betekenisitems vir die Aanlyn Woordeboeke van Spaans Valladolid-UVa uit te soek.

* This article was presented as a paper at the Twenty-first Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Xitsonga and Sesotho sa Leboa National Lexicography Units, Tzaneen, South Africa, 4–6 July 2016.

Die metodes en werktuie wat in die projek gebruik word, word bespreek, sowel as die vereistes wat aan die leksikograwe se bevoegdhede, kennis en vaardighede gestel word. Ten slotte verskaf die artikel 'n paar algemene gevolgtrekkings, asook aanbevelings en hipoteses rakende leksikografiese werk en navorsing.

Sleutelwoorde: INTERNETLEKSIKOGRAFIE, AANLYN LEKSIKOGRAFIE, LEKSIKOGRAFIESE METODOLOGIE, EMPIRIESE BASIS, LEMMASELEKSIE, BETEKENISSELEKSIE, LEKSIKOGRAFIESE DATABASISSE, SPAANSE WOORDEBOEKE, EENTALIGE WOORDEBOEKE, ALGEMENE WOORDEBOEKE

1. Introduction

Looking at the overall dictionary compilation process as described by Fuertes-Olivera and Tarp (2014: 85), there are three instances where lexicographers may need access to empirical data in order to do a good job. The first instance is when they are looking for information about the foreseen users' lexicographical needs with a view of preparing a dictionary concept which can assist these users in solving their needs. The second one is when they are selecting and preparing the lexicographical data to be included in the dictionary. And the third one is when they evaluate the usefulness of the dictionary in terms of user satisfaction. In fact, there is also a fourth situation where external empirical data may be required, i.e. when analysing the market in order to determine the sales possibilities of the product, but this is more related to the business side of the project than to lexicographical aspects in the narrow sense of the word. Anyway, in each of these situations there is a set of methods that may appear to be more or less appropriate, i.e. more or less reliable and fast in terms of both productivity and quality of the final product.

In the following, we will look at the empirical bases and the corresponding methods that can be applied when selecting lemmata and meaning items (senses) in a lexicographical online project. We will then discuss some of the most important advantages and disadvantages when using the Internet directly as a corpus, and compare them to the usefulness of "traditional" text corpora. As an example we will take a project currently carried out at the International Centre of Lexicography at the University of Valladolid, namely the *Diccionarios en Línea de Español "Universidad de Valladolid"*, in the following referred to as the *Online Dictionaries of Spanish Valladolid-UVa*. The project, which was originally initiated as a collaboration between the Valladolid-based Centre and its sister Centre for Lexicography at Aarhus University, is based on the lexicographical function theory and inspired by a similar Danish project; cf. Fuertes-Olivera and Bergenholtz (2015). Finally, we will present the hitherto experience and introduce the need to count on intuition as an intangible but highly relevant and unavoidable method in dictionary making.

2. Relationship between lexicography and technology

In a historical perspective, a both intimate and complex relationship can be

observed between lexicography and technology. This implies, among other things, that technological development may lead not only to new tools with which lexicographers can perform their art and craft, but also to new empirical bases from which they can retrieve their data as well as the need for, and possibility of, developing new methods that can be applied in this respect. The reflection is especially relevant in historical periods as the present one where new disruptive technologies are being introduced in lexicography with consequences that can still not be completely grasped:

Today we are in the middle of a new transition of the material and technological basis of lexicography with the introduction of new production tools and methods as well as new platforms and media for presenting the lexicographic product and the extensive use of corpora for the collection of material. The development and technological innovation are going faster than ever before. (...) We know the point of departure but we still only have a vague idea of where we will eventually arrive. (Gouws and Tarp 2016)

Generally, there is a variety of sources from which lexicographers can obtain their data. Bergenholtz and Tarp (1995: 90-96) discuss, among the most important, introspection, multispection, external experts, existing dictionaries, handbooks, textbooks, example cards, and text corpora. With the exception of the three former, these empirical sources have only been possible thanks to the technological development at its various stages: the invention of paper, pens, bookbinding, printing machines, computers, and databases. Since then, with the introduction and development of the web technology, another empirical data source has been put at the disposal of lexicography, namely the Internet.

It is interesting to note that Bergenholtz and Tarp (1995) discuss the consultation of external experts as a form of multispection and, implicitly, the use of one's own knowledge as a form of introspection. It may be so, but it seems nonetheless that there is a difference between the use of introspection in terms of language skills and competence, as it is normally understood within linguistics, and the use of up-to-date expert knowledge stored in someone's memory. As Tarp (2008: 131-136) has argued, in the preparation of various types of dictionary it is important to distinguish between language skills and learned knowledge of a given language, for instance as it is provided by linguistic theory. In this respect, it seems reasonable also to distinguish between the use of language competence and the use of expert knowledge in the dictionary compilation process. Hence, although there is certain terminological confusion in the existing lexicographical literature, introspection — rather than an empirical basis in itself — should be considered a method to obtain specific types of empirical data. It is a method to "look" into oneself in order to retrieve material for different purposes. The "internal" empirical bases on which lexicographers can draw by means of this method are language competences, skills, and knowledge, to which can be added personal experience in general.

The various empirical sources are seldom used alone. In a book review, Kilgarriff (2012) provides an example on how two different types of empirical basis are combined:

I noticed a lexicographical bloomer. On pp 211-213 we have an analysis of the English phrasal verb *call back*. It is given six meanings of which the sixth is given the example "I cannot call his face back." As an English native speaker, I go *eeeeeeugh*. This is blazingly wrong. (We might say "I cannot recall his face.") A little research revealed that this 'example sentence' exists in a number of dictionaries and translation tools: a dictionary error that has been copied and recopied from dictionary to dictionary. (Kilgarriff 2012: 28)

When Kilgarriff says he "goes eeeeeugh" as a native speaker of English, this suggests that his mother-tongue competence warns him about a possible problem which he subsequently confirms and explains through the consultation of other empirical bases, in this case existing dictionaries and translation tools. This is obviously the right method to apply in such cases because it implies that "the lexicographer's primary source of evidence for how a word behaves switches from subjective to objective; from introspection to looking at contexts" (Kilgarriff 1997: 111). It should be noted that Kilgarriff here speaks about the *primary* source of evidence, not the only one, although he errs when defining introspection and "looking at contexts" as sources, inasmuch as they are both methods to access the real sources of evidence.

3. Corpus versus Internet: Preliminary discussion

The first electronic text corpora were introduced in the 1960's, and since then they have never stopped growing. The two first decades after their appearance were characterized by a fierce battle of ideas between the researchers who defended the relevance of corpora for both linguistics and lexicography, and those who opposed this idea with various arguments, generally in favour of introspection as a much more appropriate method to get empirical material. One of the defenders of introspection was Lees (1962) who declared straightforwardly:

You are a native speaker of English; in ten minutes you can produce more illustrations of any point in English grammar than you will find in many millions of words of random text. (Lees 1962: 110)

Little by little the discussion faded out. Half a century after their introduction, there is no longer any doubt that electronic text corpora can be of great value not only to linguistic research but also to lexicographers when performing a series of tasks in connection with the compilation of dictionaries. This has been argued by various scholars engaged in practical lexicography, among them Bergenholtz (1996), Atkins and Rundell (2008), and Hanks (2012). The proof of the pudding is the existence of many high-quality dictionaries which have been compiled based upon this type of empirical basis (see e.g. Sinclair 1997), although the eagerness has sometimes gone too far, none at least in connection with the selection of terms and definitions in specialized dictionaries; cf. Tarp (2016), and Xue and Tarp (2016).

However, a negative consequence of this generally positive development is that introspection as a method to make use of one's own competences and knowledge is occasionally underestimated or even ignored. Although "lexicographers should never rely solely on the introspective approach" (Bergenholtz and Tarp 1995: 92), especially in cases of doubt, it is frequently forgotten that introspection always lays as a filter at the bottom of the lexicographer's choices inasmuch as no seriously working dictionary maker would introduce linguistic or any other type of data with which he or she disagrees — says "eueeeugh" in Kilgarriff's expressive but very accurate terminology — without first negotiating their correctness with other empirical sources.

Today, corpora composed of texts containing hundreds of millions of words are available to the compilers of dictionaries. In this respect, Big Data is already a reality, but the understandable excitement created by this development should never be allowed to overshadow the fact that no corpus, however big, can stand up to the enormous collection of texts and words which can be accessed through the Internet. The development of methods allowing for the use of this almost unlimited empirical basis constitutes undoubtedly a challenge more and more relevant to lexicography.

According to Fuertes-Olivera (2012: 51), a *lexicographical corpus*, i.e. a corpus that can be used to assist dictionary making, can be defined as "any collection of texts where lexicographers can find inspiration for completing the dictionary structures they need when making a real dictionary". As already mentioned, the Internet is made up by a collection of texts. Thus, if a lexicographer can find inspiration in this big collection of texts, the Internet can also be considered a type of lexicographical corpus according to the above definition. This is also the point of view of Kilgarriff and Grefenstette (2003: 334) who write that "the answer to the question 'Is the web a corpus?' is yes."

In this respect, there are two different ways of using the Internet in relation to a lexicographical project, namely 1) constructing a corpus of texts found on the Internet, and 2) using the Internet directly as a corpus, in both cases by means of search engines and other tools. Each of these two types of lexicographical corpus has its advantages and disadvantages. Below are listed some of the advantages when the Internet is used directly as a corpus, in comparison to the use of "traditional" corpora that are made up of collections of texts, whether or not these texts are taken from the Internet or elsewhere:

- The lexicographers have access to many more texts than the ones included in any corpus of selected texts.
- The texts are always up-to-date.
- Time and money are saved when it is not necessary to compose a separate corpus (which is a requirement in relation to specific types of dictionaries, in particular specialized ones).
- The search process can easily be limited to specific geographic areas, a fact that is especially important for a multinational language as Spanish.

- The use of the Internet may lead to the identification and selection of more meaning units than those that can be found in a separate corpus.

As to the disadvantages when using the Internet directly as a corpus, the following seem to be the most important:

- The quality and origin of the texts cannot be controlled.
- The authors of some of the texts may not be real persons.
- The authors may have a low proficiency level in the language in question.
- The texts may not have been revised and corrected.
- It is difficult to calculate the frequency of the linguistic phenomena appearing in the texts.

Some of the above disadvantages may not be relevant to concrete dictionary projects. Gudmann (2014: 32), for instance, argues that "information about frequency (...) is not particularly relevant to a general monolingual reception dictionary". In other cases, the disadvantages can be neutralized, or at least considerably reduced, by a well-trained lexicographer who plays an *active role* based on his or her language competence, skills, knowledge and experience. We will return to this question later on. At this point, our preliminary conclusion is that in spite of the undeniable disadvantages, it is perfectly possible, and even beneficial, to use the Internet as the main empirical source, without resorting to the "traditional" text corpora, when the objective is the production of dictionaries of still higher quality.

4. Selecting lemmata

We sincerely doubt that the traditional corpus composed of a collection of texts is the most appropriate empirical source for lemma selection, especially if this selection has to be done from scratch. To the best of our knowledge, the big general dictionaries that use corpora for this purpose are mostly dictionaries that had their basic lemma stock selected before the introduction of corpora which are now "only" used to provide additional lemmata, among other data. A different method and empirical basis are therefore required when the challenge is a quick and reliable selection of lemmata to a completely new lexicographical project of the magnitude of the Online Dictionaries of Spanish Valladolid-UVa which are planned to handle more than one hundred thousand lemmata and many more senses. The primary empirical basis chosen for this project was therefore the Internet accompanied by a method which will be described in this section.

The basic idea is that the Internet already contains a considerable number of smaller or bigger word lists for free access and use. The challenge is therefore to find these lists and make use of them. This is done by means of an Internet crawler that has been specially designed for this purpose by the Dan-

ish company Ordbogen.com which, due to its business model, is the world's currently most successful provider of online dictionaries in subscription.

Once a number of useful word lists have been found by the Internet crawler, these lists are copied and pasted into a so-called *lemma loader* (see Figure 1), another tool developed by Ordbogen.com and conceived by Professor Emeritus Henning Bergenholtz from the Centre for Lexicography in Aarhus. The lemma loader assigns automatically a lemma to a card in the database and has the advantage that it does not reduplicate the lemmata, but rejects them if they are already stored in the database.

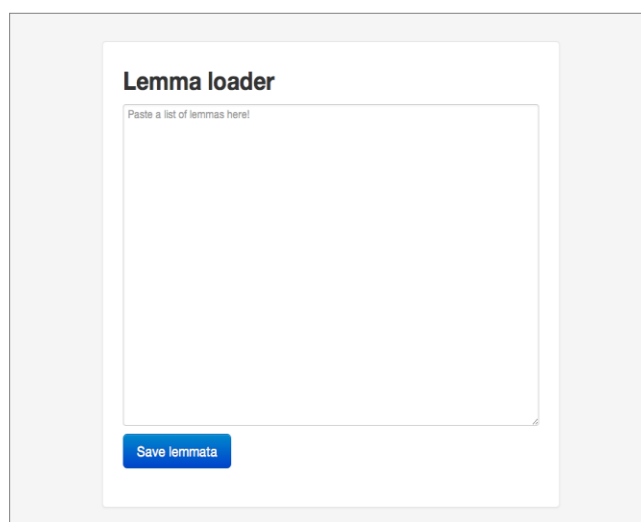


Figure 1: Screenshot of the lemma loader with the field where the copied wordlists are pasted

The experience shows that this method to select lemmata by means of an Internet crawler and a lemma loader is a very efficient, fast and totally reliable method in the case of one-word lemmata. This is reflected by the fact that only one month after the lemma selection to the Online Dictionaries of Spanish Valladolid-UVa started (July 2013), the database contained already 58.000 cards with one-word lemmata.

The next step in the lemma selection process is a manual revision which takes place when the formal grammatical data are attached to the lemma card. The revision is performed by the editor-in-chief (Pedro A. Fuertes-Olivera). Due to the characteristics of the dictionaries and the almost unlimited storage capacity of the database, the project does not work with lemma inclusion criteria but only with exclusion criteria. With this approach, only lemmata that clearly represent spelling mistakes or cannot be documented in the empirical basis, i.e. the Internet, are excluded, even when the latter can be found in some

old word lists but not on the Internet as such. Until now there have only been few cases where a lemma had to be excluded, a fact which also points to the efficiency of the method.

In order to guarantee a systematic treatment of the language, after the initial selection a number of thematic lists containing colours, numbers, cities of a certain size, rivers of more than 1.000 km, etc. were elaborated and the corresponding words introduced in the database as lemmata. This work lasted from September to November 2013 and resulted in the inclusion of additional 10.000 lemmata.

Apart from the mentioned empirical sources, other sources are also used in order to provide a flow of new lemmata. For instance, when the lexicographers are working on the Internet in order to identify meaning items to the selected lemmata (see Section 5), they simultaneously detect a considerable number of synonyms, antonyms and word combinations which are continuously introduced into the database as new lemmata.

Finally, there is the question of idioms and other fixed expressions which are generally selected as lemmata in their own right in the Online Dictionaries of Spanish Valladolid-UVa. Here there are four sources: 1) When the lexicographers are detecting meaning items they occasionally come up with such fixed expressions which are sent the editor-in-chief who then evaluates and analyses them by googling on the Internet. 2) Existing dictionaries are also used as sources in this respect, and 3) the same is the CREA Corpus composed and published by the Royal Spanish Academy for free use. 4) Finally, a number of fixed expressions are also found in other sources, e.g. books and articles read by the lexicographers in connection with other tasks.

As can be seen, it is only the three last sources of fixed expressions where the Internet is not the empirical basis for the selection of lemmata for the Online Dictionaries of Spanish Valladolid-UVa. Generally, the overall process can be characterized as a very fast, efficient and low-cost process which, until now, has resulted in about 20% more lemmata than those contained in the hitherto biggest Spanish dictionaries.

5. Selecting meaning items

The method developed to select meaning items in the Online Dictionaries of Spanish Valladolid-UVa is strongly inspired by a similar method used in the Danish Internet Dictionaries (see Bergenholtz and Agerbo 2014), but it also has some particularities of its own. Roughly speaking, the meaning selection method encompasses the following 15 steps or actions:

1. A lemma contained in the database is chosen in the lexicographer's user interface (see Figure 2)
2. The button "Google" to the left in the lexicographer's interface is activated.

3. A "traditional" Google-search result appears (See Figure 3).
4. The first (3-20) pages are skipped because they only contain lexicographically irrelevant data.
5. The minitexts appearing on each page are read in order to get a general idea of what it is all about.
6. Using the method "copy and paste" the relevant parts of the minitexts are copied into a Word document.
7. Simultaneously, collocations, examples, synonyms, antonyms and word formations are selected in order to be introduced into the respective fields in card representing the sense in question in the lexicographer's interface (see Figure 4 and 5). Idioms and fixed expressions are sent to the editor-in-chief for further evaluation.
8. A number of Google pages are reviewed until no more new data appear and everything is repeated. The number of pages depends on the characteristics of each lemma as well as the lexicographer's intuition based upon experience.
9. Once a satisfactory amount of empirical data has been selected, these data are grouped according to meaning.
10. Based on the groups of data the first definitions are written according to the lexicographical instructions prepared by the editor-in-chief.
11. Now the lexicographer decides whether he or she is satisfied, or if it is necessary to repeat the process, or part of the process, in order to obtain a satisfactory amount of empirical evidence.
12. When the lexicographer has finished meaning selection and written the definitions of the senses addressed to a lemma, a message is sent to the editor-in-chief.
13. The editor-in-chief revises the definitions and compares them with the ones appearing in four Spanish dictionaries (see Section 6). If something is missing, this may lead to a new search process as it is a basic principle in the project that no definition is copied from other dictionaries.
14. If the definitions are related to specialised terms appearing in general language, external experts may be consulted in order to control their correctness.
15. When the editor-in-chief is satisfied — and other relevant data such as grammar, synonyms, antonyms, word formations, collocations and example sentences have been included — the lemma in question is indicated for online publication.

The screenshot shows a web-based interface for editing or viewing the entry for the word "cachupina". On the left, there is a sidebar with various tools and options, including "Clase de palabra", "Sinónimo", "Colocación", "Polisemia", "Encontrar", "Google", "Diccionario", "Nuevo", "Copiar", "Borrar", and "Guardar". The main area contains a search bar with "cachupina" entered, and a dropdown menu set to "Neutro". Below this, there are fields for "Primera ref." and "Ref. gr.", both currently empty. A "Nota gramatical" field is also present. The core of the interface is a grid of grammatical categories, each with a corresponding form: "nombre" (with "una cachupina", "la cachupina", "unas cachupinas", "las cachupinas" and "Gram: 1"), "nombre propio", "verbo", "adjetivo", "pronombre", "artículo", "expresión", "conjunción", "adverbio", "interjección", "numeral", "abreviatura", "preposición", and "dicho". Each category has a small form with a label and a value field.

Figure 2: The lexicographer's interface for grammar to the word "cachupina"

The screenshot shows the first page of Google search results for the word "cachupina". The search bar at the top contains "cachupina" and the Google logo. Below the search bar, there are navigation options: "Alle", "Billeder", "Maps", "Videoeer", "Mere", and "Søgeværktøjer". The search results are listed below, starting with "Side 10 af ca. 8.350 resultater (0,31 sekunder)". The first result is "CVC. El legado de Sefarad. Memoria de Sefarad en América. La niña ..." from cvc.cervantes.es. The second result is "Las gafas del enemigo -Hispano Post" from www.hispanopost.com. The third result is "en el cerro de la culebra, el ciclope es ecocida - EnlaceVeracruz 212" from archivo.vazquezchagoya.com. The fourth result is "Los Andes-Pelluhue Chile Slideshow & Video | TripAdvisor" from tripwov.tripadvisor.com. The fifth result is "Foto de COROZAL (SUCRE), Colombia" from www.fotopaises.com.

Figure 3: Result of Google search for the word "cachupina"

The screenshot shows a web interface for editing the word "cachupina". On the left is a vertical menu with options: Clase de palabra, Sinónimo, Colocación, Polisemia, Encontrar, Google, Diccionario, Nuevo, Copiar, Borrar, and Guardar. Below the menu are the labels "pedrofinito" and "evafinito". The main area has a header with "cachupina" and dropdown menus for "Neutro", "Hom", and "Pol 1". The definition is "denominación que los nativos americanos daban a las españolas peninsulares que habían emigrado y se habían asentado en el continente americano". Below this are sections for "Sinónimo" and "Antónimo", each with a text input field and a "Neutro" dropdown. The "Nota léxica" section contains the text "Se usa en Hispanoamérica.". The "Nota sinonímica" section is empty. The "Nota de uso" section contains "Se usa de forma despectiva.". Below that is a "Segunda referencia" section with a "Borrar" button and an "Enlace" section containing the URL "http://es.wikipedia.org/wiki/Gachup%C3%91" and the text "wikipedia".

Figure 4: The lexicographer's interface for the introduction of definition, synonyms and antonyms to the word "cachupina"

This screenshot shows the same interface as Figure 4 but with different content. The "Colocación" section contains three text boxes: "García (segundo) de Escalante y su mujer, Catalina González, 'la Cachupina'", "las cachupinas llegaron a la zona tiempo después que sus maridos", and "resultar atractivas las cachupinas a los nativos". The "Ejemplo" section has a text box with "La encargada de la hacienda era una cahupina viuda que había heredado todo de su". Below this are sections for "Expresión idiomática" and "Explicación idiomática", each with a "Neutro" dropdown and a text input field. The "Refrán" section has a "Neutro" dropdown and a text input field. The "Formación de palabras" section has three text boxes containing "cachopín", "cachupín", and "cachupino". The "Gramática" section has a text input field.

Figure 5: The lexicographer's interface for the introduction of collocations, example sentences, fixed expressions and word formations to the word "cachupina"

Currently (October 2016), the database of the Online Dictionaries of Spanish Valladolid-UVa contains about 55,000 finished cards (each of them representing one sense) that are ready for publication. The experience until now shows that for 70 percent of the lemmata it is sufficient to work with the minitexts that appear as a result of the Google search. For the remaining 30 percent it is therefore necessary to activate one or more of the links in order to find additional data in the unfolded documents. In this last case, the required data can be found 90 percent of the time. Only in 10 percent of the cases, representing 3 percent of the totality of lemmata, is it necessary to perform a new search with a variant of the lemma in question. This means that for 97 percent of all lemmata one Google search is sufficient to obtain the empirical material required to select meaning items and write definitions of the desired standard.

6. Comparison with similar Spanish dictionaries

As mentioned above, after writing the definitions of the different senses, these senses are compared with those found in four Spanish dictionaries, namely:

- María Moliner: *Diccionario de Uso del Español* (DUE)
- Aquilino Sánchez Pérez: *Gran Diccionario de Uso del Español Actual* (GDUEA)
- Manuel Seco: *Diccionario del Español Actual* (DEA)
- Real Academia Española: *Diccionario de la Lengua Española* (DLE)

These four dictionaries are among the biggest and most prestigious general dictionaries of Spanish. In this respect, the comparison also serves as a sort of quality control and indication of what could be improved (see action 13 in Section 5). The comparison has so far been favourable to the Online Dictionaries of Spanish Valladolid-UVa, as it shows that each lemma treated with the described method has an average of 30-40 percent more senses than the ones found in the four other dictionaries.

Table 1 shows the number of senses which the five mentioned dictionaries provide to eight different lemmata. It is in no way representative but just an indication of how the method described in the previous Section in some cases can generate a bigger number of senses.

Word	DUE	GDUEA	DEA	DLE	Valladolid
ababol	1	1	1	2	3
cabila	1	1	2	2	3
cable	4	4	4	6	11
cabestro	4	4	2	4	6
eclipsar	2	2	2	2	3

eclipsarse	2	2	2	3	4
halagar	4	2	2	4	3
machaca	5	5	5	6	15

Table 1: Comparison between five Spanish dictionaries in terms of number of senses addressed to eight selected lemmata

The tendency reflected in Table 1 is corroborated by Gudmann (2015) who has studied five Spanish online dictionaries, among them the one published by the Royal Spanish Academy, and identified a surprisingly big number of meaning lacunae. Another illustration of this phenomenon is the treatment of *cachupín* and *cachupina* which are two words used in some parts of Latin America and presented as one and the same lemma in the four other dictionaries mentioned above, each of them with only one sense. Figure 7 shows how they are handled by the Real Spanish Academy in the online version of its *Diccionario de la Lengua Española*:

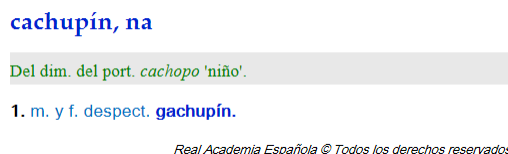


Figure 6: The lemma "cachupín, na" in the *Diccionario de la Lengua Española*

This way of presenting the two words can rightly be considered sexist, as if the feminine word *cachupina* was just subordinated to the masculine word *cachupín*. In the Online Dictionaries of Spanish Valladolid-UVa, the two words are therefore treated separately and listed as two different lemmata. This also suggests that a separate Google search for meaning items has been performed for each of the two words. The surprising result of this method is that the masculine word now appears with two senses while the feminine word includes no less than eight senses, i.e. a total of ten senses, as can be seen in the beta screenshots of the reception dictionary extracted from the database and presented in Figure 7 and 8.

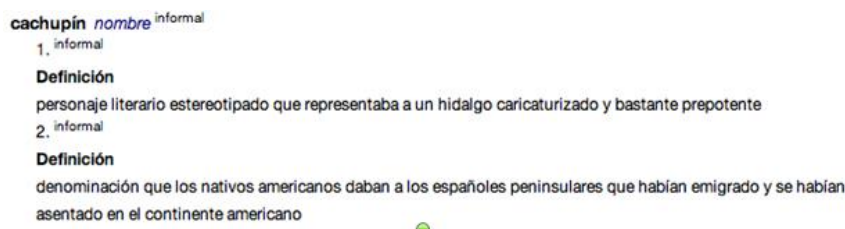



Figure 7: The lemma "cachupín" in the Online Dictionaries of Spanish Valladolid-UVa

cachupina *nombre* informal 

- Definición**
denominación que los nativos americanos daban a las españolas peninsulares que habían emigrado y se habían asentado en el continente americano
- Definición**
baile tradicional de Chile, de origen incierto, pertenece a los denominados "bailes de la tierra"
- Definición**
arbusto perenne de la familia de las Malváceas, destaca por su grandes y vistosas flores de variados colores, dependiendo de la especie, sus hojas son de un verde brillante, es originario de Asia central aunque se ha adaptado perfectamente a otras zonas del mundo
- Definición**
flor de tamaño medio y vistosos colores dependiendo de la especie que da el arbusto perenne de la familia de las Malváceas, de hojas verde brillante y originario de Asia central aunque se ha adaptado perfectamente a otras zonas del mundo
- Definición**
denominación cariñosa que se da a las perras
- Definición**
forma infantil o muy cursi de llamarle al órgano sexual femenino
- Definición**
prenda médica utilizada en centros de salud mental que consiste en una camisa abrochada a la espalda con mangas largas y cerradas que se anundan forzando que el paciente cruce los brazos sobre su pecho, se utiliza para inmovilizar a los pacientes e impedir que se hagan daño a sí mismos o a los demás
- Definición**
cinturón o peto normalmente realizado en cuero que tiene una especie de compartimentos para poder guardar los cartuchos cuando se va de caza

Figure 8: The lemma "cachupina" in the Online Dictionaries of Spanish Valladolid-UVA

The above differences should not be absolutized as the number of lemmata and senses collected in the database of the Online Dictionaries of Spanish Valladolid-UVA cannot be compared directly with the four other dictionaries of Spanish. There are two reasons for this. The first one is that the latter four are all printed dictionaries which suffer from the well-known space restraints. And the second one is that the "paper philosophy" continues to influence current Spanish lexicography even when it goes online. In this respect, a serious problem is that the selection criteria have still not been adapted to the new technology. Rundell (2015) shows in a condensed way how these criteria have been turned upside down by recent developments:

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'? (Rundell 2015: 312)

This change in selection criteria suggests that lexicographers, who only a few decades ago should justify the inclusion of any new lemma or sense because it frequently meant the exclusion of other data, are now challenged with the need to justify the non-inclusion of new lemmata and senses. In this respect, we do not know how many unpublished senses the other Spanish dictionaries may have in their databases. And neither do we know how many additional meaning items could be identified in their corpora if it was required. The comparison between the five dictionaries presented in Section 6 should therefore be taken as an indication of the new possibilities which the use of the Internet put at the disposal of lexicography, and not as an evidence-based fact that accounts for all aspects. At the end of the day what should be compared are not the dictionaries as such but the methods and empirical bases used to provide their lexicographical data.

With this in mind, our **current hypothesis**, which should be subjected to further research, is that *a corpus composed of selected texts is more appropriate to identify the most typical and frequent words, senses and behaviours of words, whereas the use of the Internet directly as a corpus is a more appropriate method when it is a question of detecting the less typical and frequent words, expressions, senses and behaviours of words.*

7. Quality and productivity

From the previous discussion it follows that the use of the Internet directly as a corpus represents a promising method to exploit two relevant results of recent technological development, namely 1) that the Internet today comprises an almost "unlimited" number of texts and words, and 2) that a modern digital dictionary is sustained by a database with almost "unlimited" storage capacity. The choice of method is therefore both an interesting topic for academic discussions and a question with big practical and economic consequences. Today the challenge for publishing houses and lexicographical teams in general is not only to compile high-quality dictionaries but also to guarantee high productivity in the compilation process. An increasing number of lexicographers are becoming aware that their discipline is submerged in a crisis which in a certain manner could be described as a struggle between life and death. This crisis is determined by two opposed tendencies in current lexicography: On the one hand, many publishers of high-quality dictionaries are closing down their dictionary departments due to lack of income and a sustainable business model. On the other hand, an increasing number of free-access dictionaries of dubious quality are placed on the Internet by generally well-intended but insufficiently trained people.

The results of this development are many and mostly negative. Although

dictionaries, due to the online media, have more users than ever before, many of these users experience problems when they try to make use of the information retrieved from such dubious dictionaries, a fact which many language and translation teachers will recognize. The unavoidable result is that an increasing number of users, who are conscious of their needs, turn their back to lexicography and look in other information tools for assistance. Some big Spanish enterprises that are willing to pay for the service are, for instance, critical of the current standard of Spanish online dictionaries, leaving a market open for projects like the Valladolid-UVa. However, the problem is further exacerbated by the fact that many modern users of the Internet, especially young people, expect the service to be free as discussed by Gouws and Tarp (2016).

Publishing houses all over the world are struggling to find a solution to these challenges which are very complex. One of the necessary counter-measures to the present crisis is undoubtedly to *increase productivity* in the dictionary compilation process with a view to reducing costs and finding a sustainable business model, i.e. increase productivity without compromising quality. Productivity has many faces and can only be increased through the *integration of user-friendly technology, efficient methods and well-trained and motivated lexicographers*. In modern dictionary projects as the Valladolid-UVa, the lexicographer's interface is the central working tool by means of which the lexicographers introduce their data into the database. In this respect, Tarp (2015) writes:

The lexicographer's interface is basically a *means of production*. It should therefore be designed with a view to guaranteeing both high productivity and the highest possible quality of the resulting product, i.e. the data stored in the database. This requires above all that it contains all the fields needed to introduce lexicographical data of the foreseen types into the database. But it is also important that the interface is as user-friendly as possible in order to facilitate the lexicographer's job, reduce the number of mistakes, economise on the resources employed, and shorten the total production time. (Tarp 2015: 234)

Figures 2, 4 and 5 in Section 6 reproduce three screenshots of the lexicographer's interface related to the lemma *cachupina*. The first one represents the mother card in which the grammatical data common to all senses of the lemma are introduced. The second and third screenshots show two pages of the card representing the first sense of the lemma. The pages are structured according to the different tasks that have to be performed and in such a way that the need to swift between one and another by means of the functional buttons to the left is reduced to a minimum. The main idea is that the lexicographers should feel comfortable when they work with this interface. Its user-friendly design together with the methods described in the previous sections is the precondition for the high productivity that characterizes the compilation of the Online Dictionaries of Spanish Valladolid-UVa.

The identification of meaning items and the introduction of definitions and other lexicographical data into the database started in March 2014. The experience shows that a lexicographer can finish an average of 4 to 6 senses per

hour with the described method and production tool. The experience also shows that productivity decreases after four or five hours as the job demands a high degree of concentration, for which reason the four lexicographers doing this part of the job only work four hours daily on the project. But if we make an abstraction from this fact, it would mean that a full-time lexicographer in an 8-hour work day would be able to finish about 40 senses and in a 40-hour week about 200 senses, which adds up to a total of about 9.000 senses per lexicographer in a 45-week work year.

The overall outcome is that the four half-time lexicographers attached to the project have finished a total of about 40.000 sensed (cards) from March 2014 to July 2016. This could be compared with the small army of lexicographers who are working on the dictionary of the Royal Spanish Academy and which we believe to be around 20 people. If this team worked with the described tool and method they would be able to produce about 180.000 senses per year. If they worked three years it would run into more than half a million senses and after only five and a half year the number would reach one million senses, i.e. the most comprehensive Spanish dictionary ever produced. In this respect, the gauntlet is down! Basically, it is a question of fully adapting to and exploiting the new technologies and techniques put at the disposal of lexicography. The Online Dictionaries of Spanish Valladolid-UVa provides one example of how this can be done although we are not claiming that it is the only road that leads to Rome.

8. The lexicographer's competences and active role

Tools, methods and empirical bases do not make a dictionary on their own. However advanced the technology, the most important thing in dictionary making is still the human factor in the form of skilled, knowledgeable and motivated lexicographers. Dictionaries, as Gudmann (2014: 31) rightly states, "are still made by real human beings through a creative process without a correct answer carved in stone."

So what precisely is required from the lexicographers participating in the project? Here it is once more necessary to make a distinction between *knowledge* and *skills*. This means, on the one hand, that a modern online project as the Online Dictionaries of Spanish Valladolid-UVa cannot prosper without a manager (lexicographer-in-chief) who has a profound knowledge of lexicographical theory and methodology as well as the ability to design a dictionary concept, write instructions, select and train a team of collaborators, and supervise the daily work.

On the other hand, it also means that the project needs a team of skilled lexicographers who are highly productive and able to generate lexicographical data of the required quality. These practical lexicographers should above all have linguistic competences in Spanish, i.e. they should be native Spanish speakers. In addition, they should also have "a sound knowledge of the world

generally and about at least one specific field" (Bergenholtz 2013: 5). The specific knowledge could be about linguistics but it could also be about other disciplines relevant to the project. In this respect, Bergenholtz (2013) reports that the team of lexicographers working on the Danish Internet Dictionaries is composed of people from language studies, mathematics, chemistry, molecular biology, physics, legal science, economics and chemistry. When the Spanish lexicographers were tested before being employed in the Valladolid-UVa project they should, apart from language competence and knowledge, also prove other relevant skills such as their ability to use computers, navigate on the Internet, find relevant data according to the instructions, and transform these data into easily understandable Spanish definitions.

However, this is only the starting point. In some of the actions listed in Section 5, it becomes clear that the selection of meaning items is not an exact science with "a correct answer carved in stone". On the contrary, a successful result depends to a large extent on the lexicographer's active role and decisions, which are not only based on his or her language competences and knowledge, but also on experience. This is, at least, the case for the following actions:

- Action 4: How many pages should be skipped?
- Action 6: Which parts of the minitexts are relevant?
- Action 8: How many pages should be reviewed?
- Action 9: When is the amount of empirical data satisfactory?
- Action 11: When is the lexicographer satisfied with the process?

It goes without saying that the decisions taken in these cases will affect the quality of the final product and that the decision time itself will have consequences for productivity. The challenge is therefore to reduce decision time and raise the quality of the decisions. In order to understand what happens, or should happen, it seems beneficial to refer to the "five stages of skill acquisition" proposed by Dreyfus and Dreyfus (1986) who operate with the following types of performer according to skills: novice, advanced beginner, competent performer, proficient performer and expert. Flyvbjerg (2001) has summarized the characteristics of these five stages in the learning process:

- (1) *Novices* act on the basis of context-independent elements and rules.
- (2) *Advanced beginners* also use situational elements, which they have learned to identify and interpret on the basis of their own experience from similar situations.
- (3) *Competent performers* are characterized by the involved choice of goals and plans as a basis for their actions. Goals and plans are used to structure and store masses of both context-dependent and context-independent information.
- (4) *Proficient performers* identify problems, goals, and plans intuitively from their own experientially based perspective. Intuitive choice is checked by analytical evaluation prior to action.
- (5) Finally, *experts'* behaviour is intuitive, holistic, and synchronic, understood in the way that a given situation releases a picture of problem, goal, plan,

decision, and action in one instant and with no division into phases. This is the level of true human expertise. Experts are characterized by a flowing, effortless performance, unhindered by analytical deliberations. (Flyvbjerg 2001: 20-21)

Flyvbjerg (2001: 21) adds that the above model contains a "qualitative jump" from the three first stages to stage 4 and 5, and that "the jump implies an abandonment of rule-based thinking as the most important basis for action, and its replacement by context and intuition". In another publication, Dreyfus and Dreyfus (1992) emphasize this point:

It seems that beginners make judgments using strict rules and features, but that with talent and a great deal of involved experience the beginner develops into an expert who sees intuitively what to do without applying rules and making judgments at all. The intellectualist tradition has given an accurate description of the beginner and of the expert facing an unfamiliar situation, but normally an expert does not deliberate. He does not reason. He does not even act deliberately. He simply spontaneously does what has normally worked and, naturally, it normally works. (Dreyfus and Dreyfus 1992: 117)

If this model is transferred to lexicography, it is easier to understand what makes a good lexicographer and what should be taken into account when selecting a team of collaborators for a dictionary project.

In the concrete case of the Valladolid-UVa project, when the job advert was posted, more than 90 candidates sent in their applications. Of these, 12 applicants were pre-selected based on their documented knowledge and competences. They were then (February 2014) offered a 30 hour course taught by Pedro A. Fuertes-Olivera and Helene R. Gudmann, the latter a skilled lexicographer with experience from the Danish Internet Dictionaries. The course included introduction to lexicography and the Valladolid-UVa project as well as instructions on how to collect data on the Internet, write definitions, and prepare the remaining data categories. The 12 candidates then took a test where they should fill in a number of cards in the database based on the instructions. The four best performers were selected for the job and started working for a 3-month trial period which is a requirement of the Spanish legislation.

When they took the test, the four lexicographers who eventually got the job, could most precisely be characterized as *novices* according to the Dreyfus model. In the 3-month trial period they were expected to develop relatively quickly into *advanced beginners* and then into *competent performers*, a transformation that would be reflected in growing productivity and quality of their lexicographical work. If this were not the case, the contract would be cancelled. However, after three months they will more often than not still be very much dependent on lexicographical instructions (rules) and chosen goals and plans as a basis for their actions and decisions, although some of them may start using their intuition based on previous experience. If this happens, they will jump to stage 4, that of the *proficient performer*, which is the minimum level that

should be expected from all lexicographers participating in projects of the magnitude and importance of the Online Dictionaries of Spanish Valladolid-UVa.

Stage 5 in the Dreyfus model is the *expert* or *virtuoso* level which is only reached by a section of practicing lexicographers. It depends above all on experience but cannot be reached without talent which should be spotted in the trial period. At this stage, the lexicographer simply acts spontaneously without deliberating, and therefore "even the best lexicographers, when pressed, can never explain what they are doing, or why" (Wierzbicka 1985: 5). Lexicography has now become *pure art and craft*. However, this does certainly not imply that "lexicography has no theoretical foundation", as Wierzbicka also claims, or that the lexicographical compilation process is not based on rules. It rather signifies that these rules have been completely internalised and integrated with experience-based intuition into a flowing, effortless and holistic performance where the lexicographer, as any other person who performs at this level, cannot explain what he or she is doing. The lexicographical theory and instructions (rules) are still there, at the bottom of everything, as an important instrumentarium that is needed in order to transmit knowledge and skills to future lexicographers.

Today, the lexicographers participating in the project at the University of Valladolid can be characterized as either proficient performers or experts as defined in the Dreyfus model. This implies that they are now able to take quick, qualified and intuitive decisions to act in situations like the ones discussed above (Actions 4, 6, 8, 9 and 11) as well as in all other situations that may be related to the compilation process. In this respect, *experience-based human intuition is an important production factor without which success in a lexicographical project would be impossible*.

All this can be summarized as follows: Firstly, the technology and methods needed to work with the Internet directly as a corpus require skilled, knowledgeable and talented human beings who are motivated to make an extraordinary performance; and secondly, these characteristics should be spotted by the project manager as early as possible, an ability which also demands experience and talent. This is at least the experience from the on-going work on the Online Dictionaries of Spanish Valladolid-UVa.

9. Conclusions

The corpora were introduced in the 1960's whereas the Internet as a generalized phenomenon did not see the light of the day until three decades later, in the 1990's. It is surprising that lexicography so far has made more use of an old technology than of a more recent one. The question is whether it is time to explore the lexicographical possibilities of the Internet. The experience of the Dictionaries Valladolid-Uva clearly indicates that the time is more than ripe. It shows that skilled and well-trained lexicographers working with the right tools

and methods are perfectly able to handle the undeniable disadvantages when using the Internet directly as a corpus instead of the traditional text corpora. This, of course, does not imply that these corpora are no longer of relevance to lexicography as they still have an important role to play when performing a number of tasks. It means above all that lexicography in order to confront its current crisis needs to go online not only to present its products but also to make them with the necessary quality and productivity.

In this paper we have discussed a set of online dictionaries of Spanish which is one of the world's biggest languages with more than four hundred million native speakers. It is evident that the number of Spanish texts placed on the Internet is enormous. However, as the experience of the Danish Internet Dictionaries shows, it is perfectly possible to use the same technology and methodology when working with a smaller language with only five million first-language speakers. In this respect, the possible problem is not as much the number of speakers as the penetration and generalized use of the Internet within a given speech community. This suggests that there may be some African languages with relatively few speakers where the collection of Internet-based texts is still not big enough to compile dictionaries as described in this contribution but they will be the exception to the rule. In most cases, the amount of Internet-based texts is already sufficient, or will be it in the nearby future. The Internet is here to stay, at least for a period of years, and it could easily be a big mistake not to start exploring its lexicographical possibilities already today.

Acknowledgements

Thanks are due to the Spanish Ministry of Economy and Competitiveness for funding the project "La Teoría Funcional de la Lexicografía: Diseño y Construcción de Diccionarios de Internet" (Ref. FFI2014-52462-P).

References

Dictionaries

- Bergenholtz, H. (Ed.). 2016. *De Danske Netordbøger*. Odense: Ordbogen.com.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.) in collaboration with M.Á. Sastre Ruano, E. Álvarez Ramos, M. Fonseca Hernández, M.J. López Carrero, Á. Prieto Salvador and O. Saldaña. *Diccionarios en Línea de Español "Universidad de Valladolid"*. Hamburg: Lemma.com. (Under construction.)
- Moliner, M. 2007. *Diccionario de Uso del Español*. Third edition. Madrid: Gredos.
- Real Academia Española. 2014. *Diccionario de la Lengua Española*. 23rd edition. Madrid: Espasa.
- Sánchez Pérez, A. (Ed.). 2001. *Gran Diccionario de Uso del Español Actual*. Madrid: Sociedad General Española de Librería.
- Seco, M., O. Andrés and G. Ramos. 2011. *Diccionario del Español Actual*. Madrid: Aguilar.

Other Literature

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bergenholtz, H. 1996. Korpusbaseret Leksikografi. *LexicoNordica* 3: 1-15.
- Bergenholtz, H. 2013. The Role of Linguists in Planning and Making Dictionaries in Modern Information Society. Kwary, D., N. Wulan and L. Musyahda. (Eds.). 2013. *Lexicography and Dictionaries in the Information Age. Selected papers from the 8th ASIALEX International Conference*: 1-10. Surabaya: Airlangga University Press.
- Bergenholtz, H. and H. Agerbo. 2014. Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries. *Hermes* 52: 125-139.
- Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Dreyfus, H. and S. Dreyfus. 1986. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Dreyfus, H. and S. Dreyfus. 1992. What is Moral Maturity? Towards a Phenomenology of Ethical Expertise. Ogilvy, J. (Ed.). 1992. *Revisioning Philosophy*: 111-131. Albany, NY: State University of New York Press.
- Flyvbjerg, B. 2001. *Making Social Science Matter. Why Social Inquiry Fails and How It Can Succeed Again*. Cambridge: Cambridge University Press.
- Francis, W.N. 1979. Problems of Assembling and Computerizing Large Corpora. Bergenholtz, H. and B. Schaefer (Eds.). 1979. *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*: 110-123. Königstein/Ts.: Scriptor.
- Fuertes-Olivera, P.A. 2012. Lexicography and the Internet as a (Re-)source. *Lexicographica* 28: 49-70.
- Fuertes-Olivera, P.A. and H. Bergenholtz. 2015. Los Diccionarios en Línea de Español "Universidad de Valladolid". *Estudios de Lexicografía* 4: 71-98.
- Fuertes-Olivera, P.A. and S. Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter.
- Gouws, R.H. and S. Tarp. 2016. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 29(4). (In print.)
- Gudmann, H.R. 2014. *Betydningshuller i Spanske Ordbøger. En Undersøgelse af Betydningenheder i Spanske Monolingvale Almene Receptionsordbøger*. M.A. Thesis. Aarhus: Aarhus University, Department of Business Communication.
- Gudmann, H.R. 2015. Lagunas de Significado en los Diccionarios Españoles. *Estudios de Lexicografía* 4: 161-184.
- Hanks, P. 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.
- Kilgarriff, A. 1997. I Don't Believe in Word Senses. *Computers and the Humanities* 2(31): 91-113.
- Kilgarriff, A. 2012. [Review of] Pedro A. Fuertes-Olivera and Henning Bergenholtz (Eds.). 2012. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. *Kernerman Dictionary News*, July 2012: 26-29.
- Kilgarriff, A. and G. Grefenstette. 2003. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29(3): 333-347.
- Lees, R. 1962. Oral contribution. Quoted by Francis, W.N., 1979: 110.

- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Sinclair, J.M.** 1997. Introduction. *Collins Cobuild English Language Dictionary*: xv-xxi. London: Harper-Collins.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, S.** 2015. Structures in the Communication between Lexicographer and Programmer: Database and Interface. *Lexicographica* 31: 217-246.
- Tarp, S.** 2016. Excesos en el Uso de Corpus en la Lexicografía: «Pesca» de Términos y Definiciones. *Revista de Lexicografía* 21. (In print.)
- Wierzbicka, A.** 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.
- Xue, M. and S. Tarp.** 2016. Corpus-based, Corpus-driven or Corpus-assisted lexicography? The Limited Usefulness of Corpora in Defining Specialised Terms. *Lexicographical Studies* 4: 1-11.

Illustrative Examples in a Bilingual Decoding Dictionary: An (Un)necessary Component?*

Alenka Vrbinc (*alenka.vrbinc@ef.uni-lj.si*), Faculty of Economics,
University of Ljubljana, Ljubljana, Slovenia

and

Marjeta Vrbinc (*marjeta.vrbinc@ff.uni-lj.si*), Faculty of Arts,
Department of English, University of Ljubljana, Ljubljana, Slovenia

Abstract: The article discusses the principles underlying the inclusion of illustrative examples in a decoding English–Slovene dictionary. The inclusion of examples in decoding bilingual dictionaries can be justified by taking into account the semantic and grammatical differences between the source and the target languages. Among the differences between the dictionary equivalent, which represents the most frequent translation of the lemma in a particular sense, and the translation of the lemma in the illustrative example, the following should be highlighted: the differences in the part of speech; context-dependent translation of the lemma in the example; the one-word equivalent of the example; zero equivalence; and idiomatic translation of the example. All these differences are addressed and discussed in detail, together with the sample entries taken from a bilingual English–Slovene dictionary. The aim is to develop criteria for the selection of illustrative examples whose purpose is to supplement dictionary equivalent(s) in the grammatical, lexical, semantic and contextual senses. Apart from that, arguments for translating examples in the target language are put forward. The most important finding is that examples included in a bilingual decoding dictionary should be chosen carefully and should be translated into the target language, since lexical as well as grammatical changes in the translation of examples demand not only a high level of knowledge of both languages, but also translation abilities.

Keywords: ILLUSTRATIVE EXAMPLES, BILINGUAL DECODING DICTIONARY, SEMANTIC DIFFERENCES BETWEEN SOURCE LANGUAGE (SL) AND TARGET LANGUAGE (TL), GRAMMATICAL DIFFERENCES BETWEEN SL AND TL, TRANSLATION OF EXAMPLES, TRANSPOSITION, CONTEXT-DEPENDENT TRANSLATION, ONE-WORD EQUIVALENT, ZERO EQUIVALENT, IDIOMATIC TRANSLATION

Opsomming: Toeliggende voorbeelde in 'n tweetalige dekodeerende woordeboek: 'n (On)nodige Komponent? Die beginsels onderliggend aan die insluiting van toeliggende voorbeelde in 'n dekodeerende Engels–Sloveense woordeboek word in hierdie artikel bespreek. Die insluiting van voorbeelde in dekodeerende tweetalige woordeboeke kan geregtigdig

* This article is an extended and revised version of the paper *The Utility of Examples of Use in an English–Slovene Dictionary*, presented by Alenka Vrbinc at the AsiaLex 2016 conference, which took place in Manila, Philippines, 1–3 June 2016.

word deur die semantiese en grammatiese verskille tussen die bron- en doeltale in ag te neem. Onder die verskille tussen die woordeboekekwivalent, wat die mees frekwente vertaling van die lemma in 'n sekere sin verteenwoordig, en die vertaling van die lemma in die toeligtende voorbeeld, moet die volgende uitgelig word: die verskille in woordsoort; konteks-afhanklike vertaling van die lemma in die voorbeeld; die eenwoordekwivalent van die voorbeeld; zero-ekwivalensie; en idiomatiese vertaling van die voorbeeld. Al hierdie verskille word aangespreek en in besonderhede bespreek saam met die voorbeeldinskrywings geneem uit 'n tweetalige Engels-Sloveense woordeboek. Die oogmerk is die ontwikkeling van kriteria vir die seleksie van toeligtende voorbeelde wat die woordeboekekwivalent(e) in 'n grammatiese, leksikale, semantiese en kontekstuele sin moet aanvul. Afgesien hiervan word argumente ook gevoer vir die vertaling van voorbeelde in die doeltaal. Die belangrikste gevolgtrekking is dat voorbeelde wat in 'n tweetalige dekodeerende woordeboek ingesluit word, versigtig gekies moet word en vertaal moet word in die doeltaal, aangesien leksikale sowel as grammatiese veranderinge in die vertaling van voorbeelde nie net 'n hoë vlak van kennis van beide tale vereis nie, maar ook vertalingsvermoëns.

Slutelwoorde: TOELIGTENDE VOORBEELDE, TWEETALIGE DEKODERENDE WOORDEBOEK, SEMANTIESE VERSKILLE TUSSEN BRONTAAL (BT) EN DOELTAAL (DT), GRAMMATIESE VERSKILLE TUSSEN BT EN DT, VERTALING VAN VOORBEELDE, TRANSPOSISIE, KONTEKS-AFHANKLIKE VERTALING, EENWOORDEKWIVALENT, ZERO-EKWIVALENT, IDIOMATIESE VERTALING

1. Introduction

Although being one of the entry components, illustrative examples are an optional rather than an obligatory element of the microstructure. They have been used in the compilation of dictionaries for a long time: they were introduced into English lexicography as early as 1755 by Samuel Johnson in his *Dictionary of the English Language*, but they had been used even before that, for example by Arab lexicographers since the 8th century or by Greek and Latin lexicographers since the 16th century (Al-Kasimi 1977: 89). Illustrative examples can be found in monolingual dictionaries, where they follow the definition(s) as well as in bilingual dictionaries, where they appear after the translation equivalent(s). In the metalexigraphic literature dealing with this topic, the following questions regarding illustrative examples are often addressed: what distinguishes an example from a sub-entry; what kind of combinations (free, partially fixed, etc.) can be used as examples; which sources of examples should be used; how the choice of a source is related to the type of dictionary and its target users (Adamska-Salaciak 2006: 493). When including illustrative examples, a range of factors should be taken into consideration: the extent of the use of examples; what role the examples will play in exemplification; where the examples come from (are they taken from a corpus?, are they invented by lexicographers, or are they (at least partially) modified by lexicographers, despite having been extracted from a corpus?); how consistently they will be included in a dictionary; how to strike a balance between authentic corpus material and the goal of lexicographic efficiency (Jackson 2002: 181; Adamska-

Salaciak 2013: 226; Krvina 2014: 3). Doubtlessly, the criteria for the inclusion of illustrative examples should be precisely defined before the compilation of a dictionary begins, and all compilers working on the project should stick to these guidelines as strictly as possible throughout work on the dictionary.

The number of examples illustrating the use of a lemma or a certain sense of a lemma cannot be determined in advance, but it can be claimed that the number of examples depends on the semantic and grammatical complexity of the lexical item in question, since more complex items generally require a greater number of examples (Toope 1996: 167). In a bilingual dictionary covering the vocabulary of two languages from very different cultures, more examples will be needed (Jacobsen et al. 1991: 2788), and the same holds true for bilingual dictionaries intended for encoding, in which more examples are needed than in those intended for decoding (Kromann et al. 1991: 2772). How many examples will indeed be included in a given dictionary is a matter that should be settled by the lexicographers, and their judgement should be based on the findings of various linguistic studies, existing dictionaries and corpora as well as (at least to a certain extent) evidence derived from their own linguistic intuition (Toope 1996: 170).

Bilingual dictionaries, which are discussed in this article, have a double function, since they can be used to serve either the decoding or the encoding function, and this is a point that should be taken into consideration when deciding what kind of illustrative examples to include in a dictionary. In practice, we do not have a set of four bilingual dictionaries to cover the vocabulary of two languages, while serving for either decoding or encoding by native speakers of either language. Consequently, it can be claimed that bilingual dictionaries usually combine the features typical of both encoding and decoding dictionaries, which means that they can be referred to as combined encoding and decoding dictionaries (Jacobsen et al. 1991: 2786). This is why bilingual dictionaries often include information the users do not need to comprehend a text in a foreign language or lack information the users would need to produce a text in a foreign language (Landau 2001: 307).

We should first briefly mention some difficulties encountered when trying to find an appropriate dictionary equivalent for a lemma or a sense of a lemma. Three levels of equivalence are distinguished: semantic (the equivalent should be semantically as close to the lemma as possible), translational (an equivalent offered in a dictionary is insertable in context; the equivalent must be the same part of speech as the lemma) and functional (the equivalent can be appropriate for a specific situational context) (Adamska-Salaciak 2006: 496). Lexicographers should try to find an equivalent in the target language (TL) that fits as many different contexts as possible, which means it is as context-free as possible (Atkins and Rundell 2008: 503). In many cases, it is not possible to find a universally acceptable equivalent, because in many source language-target language pairs, we are faced with the problem of a systematic mismatch of semantic distribution between the two languages (ibid.: 504). When a universally acceptable equivalent does not exist in the TL, lexicographers can resort to

the following alternatives: the near-equivalent and the gloss. No translation of a source language (SL) lemma is possible in the case of zero equivalence; in such a case, examples of use should be provided, together with their translation into the TL (Atkins and Rundell 2008: 505-506).

In bilingual dictionaries, illustrative examples are included in the form of a phrase or sentence in the SL that illustrates the use of a single word, compound or idiom, along with its translation (Van Scherrenburg 1990: 10; Al-Kasimi 1977: 96). A contrastive aspect is doubtlessly very much in the foreground, since the same content can be realized syntactically in different ways (Svensén 2009: 150, 155). Therefore, a bilingual dictionary should include illustrative examples that help the users to comprehend or translate a text from a foreign language into their mother tongue (dictionaries for decoding), or to produce sentences or a text in a foreign language, or to translate a text from their mother tongue into a foreign language (dictionaries for encoding). It should be clear that such examples are being used to exemplify the usage of particular senses; they should convey a great deal of information about the actual usage in context (e.g., grammatical correctness, typical collocations, variety of usage and connotation); they should complement or replace a universally acceptable equivalent and offer an appropriate solution if a universally acceptable equivalent cannot be used (for example, when the example is translated contextually) (Atkins and Rundell 2008: 507; Jackson 2002: 130; Landau 2001: 207-208; Béjoint 2000: 135).

The next issue that should be addressed is whether or not illustrative examples included in a bilingual dictionary should be translated. Metalexigraphers have conflicting opinions regarding this matter. Al-Kasimi (1977: 96) and Zöfgen (1991: 2898) advocate that the examples should be translated, whereas Jacobsen et al. (1991: 2786) and Adamska-Sałaciak (2006: 493-494) believe that there is no need to translate examples if they are chosen so as to pose no problems for anyone with a basic knowledge of the TL. Adamska-Sałaciak (2013: 228; 2006: 494), however, admits that, in certain cases, the examples must be translated into the L2, and she also enumerates some of these cases: when they introduce important exceptions (e.g., the example shows that, in certain circumstances, a different translation equivalent is needed or that the equivalent provided is omitted in translation); when they might be too difficult for the average user to interpret on their own; when the dictionary is targeted at beginners. At the same time, she proposes (2013: 228) that in electronic dictionaries, translation can be made an optional feature, to be switched on and off when needed.

The article discusses in detail the principles underlying the inclusion of illustrative examples in an English–Slovene decoding dictionary. The aim is to develop criteria for the selection of illustrative examples whose purpose is to supplement dictionary equivalent(s) in the grammatical, lexical, semantic and contextual senses. Apart from that, arguments are advanced for translating illustrative examples in the TL (i.e., Slovene in our case).

2. Illustrative examples in some existing English–Slovene dictionaries

In a bilingual dictionary for decoding, the user searches for the equivalent of a particular sense of an unknown word or word combination that s/he has encountered in a text. The question should be posed in what cases it is sensible to include illustrative examples in the entry of a decoding dictionary. It can be said with a high degree of certainty that illustrative examples are not needed if the lemma can be translated by means of the listed equivalents. In *Veliki angleško-slovenski slovar Oxford-DZS (The Oxford-DZS Comprehensive English–Slovenian Dictionary)*, a relatively high number of illustrative examples are included, but a more careful investigation of these examples reveals that many of these could be excluded without detriment to users or to the usability of the dictionary, since they are translated by the dictionary equivalent(s) provided. Figure 1 shows the entries for *seismic* (adjective), *seizure* (noun) and *seize* (verb), and, as can be seen, only the illustrative examples shaded grey by the authors of this contribution should, in fact, be included, since these cannot be translated by the direct insertion of the dictionary equivalents.

<p>seismic prid. 1. seizmičen, potresen: seismic activity seizmična aktivnost seismic waves potresni valovi 2. velikanski, gromozanski, ogromen: a shock of seismic proportions šok velikanskih razsežnosti seismic changes in economics tektonski premiki v ekonomiji</p> <p>seizure sam. 1. zavzetje; prevzem: the seizure of the Assembly building zavzetje skupščine Napoleon's seizure of power Napoleonov prevzem oblasti 2. zaseg, zaplemba, zaplenitev: a drug seizure zaseg mamil ... 4. napad: an epileptic seizure napad epilepsije Dad'll have a seizure if he sees you wearing that očija bo kap, če te vidi tako oblečenega</p> <p>seize gl. 1. z(a)grabiti (za), pograbititi (za): he seized my arm zgrabil me je za roko she seized hold of the door handle zgrabila je kljuko na vratih people seized the opportunity ljudje so zgrabili priložnost he was seized by fear zgrabil ga je strah panic seized her zajela jo je panika he seized the letter from her hands pismo ji je iztrgal iz rok 2. ...</p>

Figure 1: Illustrative examples in *Veliki angleško-slovenski slovar Oxford-DZS*

Among the unshaded examples, two should be highlighted, i.e., *seismic changes in economics* and *panic seized her*, since the translation of the lemma in these examples differs from the equivalent offered in the dictionary, but there is no sound reason for introducing a lexically new equivalent which is synonymous with equivalents provided in the dictionary. The denotation of *tektonski* is comparable to that of *velikanski*, *gromozanski*, and *ogromen*, and all three equivalents given in the dictionary are insertable in the illustrative example. The same holds true for the translation of *panic seized her*, where the equivalent provided (i.e., *zgrabiti*) can be used instead of the verb *zajeti*. Similarly, the example *he seized the letter from her hands* can be translated by the verbs given as equivalents, i.e., *z(a)grabiti*, *pograbititi*.

The illustrative examples included in this dictionary are translated without exceptions, but they are not always well chosen and many could be excluded, since the same information is simply repeated. In *Collins COBUILD Angleško–slovenski slovar Bridge* (*Collins COBUILD English–Slovene Dictionary Bridge*) and *Angleško–slovenski slovar Password* (*English Dictionary for Speakers of Slovenian Password*), which are a cross between a monolingual dictionary and a bilingual dictionary and are intended to bridge the gap between the monolingual and the bilingual dictionary, illustrative examples are not translated. In many cases, however, there is such a huge discrepancy between the dictionary equivalent and the translation of the lemma in the example that the decision not to translate examples should be considered a serious shortcoming, since the user cannot possibly know how to translate the lemma in the example. To illustrate this shortcoming, Figure 2 shows some entries from *Angleško–slovenski slovar Password*.

<p>mention noun (often with of) a (usually brief) remark (about): <i>No mention was made of this matter.</i> □ omemba</p> <p>officer noun ... 2 a person who carries out a public duty: <i>a police-officer.</i> □ uradnik</p> <p>sound verb ... 2 to signal (something) by making a sound: <i>Sound the alarm!</i> □ (po)zvoniti</p>

Figure 2: Entries from *Angleško–slovenski slovar Password*

The noun *mention* in *No mention was made of this matter* is not translated by a noun, and it is highly unlikely that an average user would use the method of transposition and translate it by means of the verb *omeniti*, i.e., *Ta zadeva ni bila niti omenjena*. As far as sense 2 of the noun *officer* is concerned, the dictionary equivalent *uradnik* is in accordance with the definition, but the example *police-officer* should be translated into Slovene by a one-word equivalent *policist* (masculine), *policistka* (feminine). It should again be pointed out that an average user would not be able to provide the only acceptable translation if this example were not translated. A similar observation can be made in connection with sense 2 of the verb *sound*, where the equivalent *(po)zvoniti* matches the definition, but the example *Sound the alarm!* is an idiom requiring a very specific translation, which can be either idiomatic, i.e., *biti plat zvoana*, or non-idiomatic, i.e., *opozoriti/opozarjati na nevarnost*.

It is evident from the examples discussed in this section that the illustrative examples included in the existing English–Slovene dictionaries should be chosen far more carefully and should certainly be translated into Slovene, since this is the only way to show the user how the lemma in question can be translated in context correctly and appropriately. In the following section, we will define criteria that should be applied when selecting illustrative examples for a bilingual decoding dictionary.

3. Criteria for the selection of illustrative examples for an English–Slovene decoding dictionary

A bilingual decoding dictionary can include examples illustrating the more general nature of a particular structure, but examples showing contrastive differences between two languages are of even greater importance. Examples are used to illustrate the differences between the listed dictionary equivalent(s) and the translation of the lemma in context or, in other words, the integration of equivalents into a context. They also demonstrate that a lemma should be translated into the TL by a word combination or a structure rather than by a single lexical item (Toope 1996: 96). Among the differences between the dictionary equivalent, which represents the most frequent translation of the lemma in a particular sense, and the translation of the lemma in the illustrative example, the following should be highlighted: differences in the part of speech; context-dependent translation of the lemma in the illustrative example; the one-word equivalent of the example; zero equivalence; and idiomatic translation of the illustrative example. All these differences will be addressed in more detail in this section, together with the sample entries taken from a bilingual English–Slovene dictionary that is in its final stages of completion.

- (1) Transposition involving a change in the grammar from SL to TL: the dictionary equivalent corresponds in the part of speech to the lemma, but the lemma used in a specific context is translated by means of another part of speech:
 - (a) a noun is translated by an adjective:
discretion sam.¹ obzirnost, diskretnost: *be the soul of discretion* biti zelo diskreten
 - (b) a noun used as a premodifier in the illustrative example is translated by an adjective:
nudist sam. nudist, nudistka: *nudist beach* nudistična plaža
narrow gauge sam. ozki tir: *narrow-gauge railway* ozkotirna železnica
 - (c) a noun used in a sentence example is translated by an adverb:
nonsense sam. nesmisel, neumnost, nonsens: *It's nonsense to say ...* Nesmiselno/Neumno je reči ...
 - (d) a noun is translated by a verb:
normal sam. normala: *Things returned to normal.* Stvari so se normalizirale.
 - (e) a noun which is used in a delexical structure is translated by a verb:
nosebleed sam. krvavitev/krvavenje iz nosa: *have a nosebleed* krvaveti iz nosa
 - (f) a noun forming a collocation with a verb is translated by an adjective:
notoriety sam. zloglasnost, razvpitost, notoričnost: *gain/win/achieve notoriety* postati razvpit

- (g) an adjective is translated by a prepositional phrase:

narcotic prid.² narkotičen: *narcotic addiction* zasvojenost z mamili

- (h) an adjective is translated by an adverb:

typical prid. značilen, tipičen: *be typical of sb/sth* biti značilno/tipično za koga/kaj

- (i) a prepositional phrase is translated by an adverb:

night sam. 1. noč: *at night* ponoči • *by night* ponoči • *in/during the night* ponoči 2. večer: *(late) at night* (pozno) zvečer

- (2) Context-dependent translation of the lemma in the illustrative example: provided there is a gap between the equivalent given in the dictionary and the context-dependent translation of a lemma, illustrative examples are to be included and translated appropriately, thus indicating that the dictionary equivalent is not immediately insertable in any context. At the same time, the user finds possible translations of the lemma in different situations and contexts. Since this is a complex issue, illustrative examples whose translation is context-dependent will be subdivided into three groups:

- (a) The translation of the lemma in the illustrative example is based on a listed dictionary equivalent, but the context requires translation with a considerably more specific expression:

average gl.³ biti v povprečju, znašati v povprečju: *I average about five cups of coffee a day*. Na dan spijem povprečno pet skodelic kave. • *Pay rises this year averaged 2%*. Letos so se plače dvignile povprečno za 2 %. • *This car averages 40 miles to the gallon*. Ta avto porabi povprečno 7 litrov bencina na 100 kilometrov.

As is evident from these examples illustrating the use of the verb *average*, the verb is often translated by a combination of a specific (context-dependent) verb (e.g., *spiti* ('drink up'), *dvigniti* ('rise, increase'), *porabiti* ('use up')) and the adverb *povprečno* ('on average').

non-existent prid. neobstoječ: *His Italian is non-existent*. Ne zna italijansko. • *Traffic was almost non-existent*. Prometa skoraj ni bilo.

In the entry for the adjective *non-existent*, illustrative examples are to be included, since the dictionary equivalent *neobstoječ* is not immediately insertable in many examples. In translating the above illustrative examples, the combination *be + non-existent* is translated into Slovene by a verb in the negative. For instance, in *His Italian is non-existent*, *be + non-existent* in the English sentence is translated into Slovene by a much more specific verb, i.e., the verb *znati* ('know'). In the second illustrative example (i.e., *Traffic was almost non-existent*), *be + non-existent* is translated into Slovene by the negative form of the verb *be* (i.e., *ne biti*).

The examples included in the entry for the noun *notch* are also worth mentioning, since the Slovene dictionary equivalent *stopnja* cannot be used in the translation of the illustrative examples:

notch sam. stopnja: *be several notches above the others* biti precej boljši od drugih • *Her spirits lifted a few notches.* Njeno razpoloženje se je bistveno izboljšalo • *raise one's voice by a notch* malo privzdigniti glas

The translation of *notch* in the above examples depends on the premodifier expressing a level of intensity: *several* and *a few* express a higher level, whereas the indefinite article *a* in the last example implies a lower level. Accordingly, the examples are translated by an appropriate adjective/adverb in combination with a verb which is semantically equivalent to the verb in the English example.

- (b) The lemma in the illustrative example is translated by using the method of expansion, i.e., by adding an entire context:

teeming prid. mrgoleč, vrveč: *teeming station* postaja, na kateri mrgoli ljudi

As is clear from this example, the Slovene equivalents correspond semantically as well as morphologically to the English lemma. Nevertheless, in spite of an almost ideal situation, neither equivalent can be employed in the translation of the example, since a translation by means of an adjective is completely inappropriate and unacceptable in Slovene. Consequently, the English participle *teeming* cannot simply be replaced by one of the Slovene participles (*mrgoleč*, *vrveč*); instead, the verb from which the participle *mrgoleč* is derived should be used. Since this verb is transitive in this context, an object (i.e., *ljudje* 'people') should be introduced in the translation of the illustrative example; this object, however, is context-dependent. Structurally, the English adjective *teeming* is translated into Slovene by a relative clause (i.e., *na kateri mrgoli ljudi*).

thick prid. poln: *in the thickest part of the crowd* v delu množice, kjer se je trlo ljudi

The above example shows that the superlative form of the adjective *thick* used as a premodifier cannot be translated into Slovene by the equivalent *poln*. The Slovene translation is context-dependent, which is reflected in the translation of the English adjective *thickest* into Slovene by an adverbial clause of place containing the verb *treti se* ('be thick on the ground') and the noun *ljudje* ('people'), an element which is absent in the English original.

tile sam. ploščica: *carpet tiles* tapison v obliki kvadratov

The Slovene equivalent of the English noun *tile* is *ploščica*, but *ploščica* cannot be used in the translation of *carpet tiles*. In this illustrative example, *tile* is translated by a prepositional phrase (*v obliki kvadratov*), the supplement *v obliki* ('in the form of') is a necessary element, while *kvadrat* ('square') can

be semantically linked with *ploščica* as regards its shape.

- (c) One and the same illustrative example can be translated differently depending on context:

tear gl. trgati, strgati, raztrgati: *tear a hole in sth narediti luknjo v kaj, strgati kaj*

In English, the verb *tear* ('strgati') collocates with the noun *hole* ('luknja'). However, in translating actual examples, it can be seen that the translation of the collocation varies. If the collocation *tear a hole* is used with a noun denoting an item of clothing (e.g., *I tore a hole in my jeans/sweater*. Strgala sem si kavbojke/jopico.), it is translated by the verb *strgati*; in another context, the same collocation can be translated as *narediti luknjo* (e.g., *The blast tore a hole in the wall*. Eksplozija je naredila luknjo v steni.).

- (3) Illustrative examples are translated into Slovene by a one-word equivalent:

nude prid. gòl, nag: *nude photograph akt*

theatre sam. 1. gledališče, teater: *a man of the theatre gledališčnik ... 2. brit.⁴ operacijska dvorana: theatre sister instrumentarka ...*

term sam. (strokovni) izraz, têrmin: *a term of abuse zmerljivka, psovka*

As can be seen, the listed dictionary equivalents and the translations of the examples differ. In all but one case, a one-word equivalent contains a semantic component of the dictionary equivalent (*akt* → an artistic depiction of a nude body; *gledališčnik* → a person who is actively involved in theatre arts; *zmerljivka/psovka* → a rude, disrespectful, offensive word). The only exception is the example *theatre sister*, which is translated as *instrumentarka*, a noun denoting a profession and containing no semantic component of the equivalent *operacijska dvorana*.

- (4) Illustrative examples in the case of zero equivalence (in the dictionary entry marked with the symbol ∅):

have gl. ... ∅: *He had his head cut off*. Obglavili so ga. • *He's had his wallet taken*. Ukradli so mu denarnico. • *We have had our request refused*. Zavrnili so nam prošnjo. ...

The causative use of the verb *have* meaning 'to cause something to be done for you by somebody else' does not have an equivalent in Slovene; consequently, only illustrative examples which are translated correctly and appropriately tell the user how *have* in this sense can be expressed in Slovene. The translation into Slovene contains the translation of the full lexical verbs in the English illustrative example (*cut off* = *obglaviti*; *take* = *ukrasti*; *refuse* = *zavrniti*); these verbs are (usually) in the third person plural, which here suggests generic or impersonal use.

piece sam. ... ∅: *a fine piece of work* čudovito delo • *an interesting piece of research* zanimiva raziskava • *a piece of advice* nasvet ...

In the sense 'a single item', the noun *piece* is used in a partitive structure consisting of a partitive (i.e., *piece*) linked by the preposition *of* to another noun (*work, research and advice*). The noun *piece* is not translated into Slovene, since its only function in the above examples is to make an English uncountable noun countable.

space sam. ... Ø: *in a short space of time v kratkem času* • *in/within the space of three years v treh letih* ...

In one of the senses, the noun *space* can be defined as 'a period of time'. Owing to the contrastive differences between English and Slovene, *space* in this sense cannot be translated into Slovene by an appropriate equivalent, thus requiring the inclusion of illustrative examples with proper translations. In the above examples, *space* is left out rather than translated into Slovene.

used to gl. (za izražanje pogostega ali ponavljajočega se dejanja v preteklosti) Ø: *Did he use to work here? Ali je delal tukaj?* • *"Do you play tennis?" "No, but I used to."* "Ali igraš tenis?" "Ne, ampak sem ga." • *He didn't use to smoke. Ni kadil.* • *He used to live in Paris. Živel je v Parizu.* • *You used to go to the cinema a lot, didn't you? Veliko si hodil v kino, ali ne?* ...

Dictionary equivalents in Slovene do not exist for certain senses of English auxiliary verbs. In such cases, possible translations into Slovene can only be provided by means of illustrative examples showing these lemmata in context. This also applies to the modal verb *used to*, whose lack of translation into Slovene should be marked using a certain symbol, such as Ø implying zero equivalence. In the above examples, Slovene translations all contain the imperfective form of the full lexical verb in the past tense. However, to make the entry more informative and user-friendly, it is advisable to add a short description of its use, one that can be included in brackets preceding the symbol, i.e., *za izražanje pogostega ali ponavljajočega se dejanja v preteklosti* ('used to express a frequent or repeated action in the past').

- (5) Some examples clearly indicate the more or less idiomatic nature of the word combination in question, but since the word combination (still) does not have the status of a phraseological unit in English, it cannot be included and treated in the idioms section. Therefore, we can find such word combinations mostly included as illustrative examples, and it can often be observed that their translations into Slovene are also idiomatic:

tired prid. sit: *sick and tired of sth do vrh glave sit česa/koga*

tenuous prid. šibek, negotov: *He had but a tenuous hold on life. Njegovo življenje je viselo na nitki.*

4. Discussion

Illustrative examples in a bilingual dictionary should be selected on the basis of

contrastive differences between the SL and TL. A lemma is always used in context, which means that, apart from semantic features, collocational, connotative and grammatical limitations should also be taken into consideration.

Although a bilingual dictionary cannot explore all the semantic possibilities of a certain lemma in context (Toope 1996: 98), it should provide equivalents that can be used by the user in speaking and writing as well as in translation; in other words, the first objective of the lexicographer compiling a bilingual dictionary is to find a TL word that fits as many of the contexts as possible and is as context-free as possible (Atkins and Rundell 2008: 503). Although a bilingual dictionary cannot include examples demonstrating all possible context-specific equivalents, it should contain a wide range of examples: examples translated by equivalents that are considered common, but are not listed among equivalents in the dictionary; examples translated by context-specific equivalents; examples translated by adding context-specific items; and examples translated by other translation methods (Toope 1996: 101). Doubtlessly, no dictionary can offer solutions to the translation of a lemma or one of its senses in all contexts, but it should include the information that triggers the user's own intuition and reminds him/her that in specific contexts, other equivalents than the most common ones can be used (Toope 1996: 101).

Transposition is one of the most common reasons for the inclusion of illustrative examples in a bilingual (decoding) dictionary. Transposition is the only translation procedure concerned with grammar and, as Newmark (1988: 88) rightly points out, most translators make transpositions intuitively. As far as an average dictionary user is concerned, the question can be raised whether s/he can also make a transposition intuitively when s/he comes across an unknown word in context. This may depend a great deal on the level of knowledge of both languages, as well as on an individual's feeling for language. Since lexicographers cannot rely on the ability of each individual user to make an appropriate transposition, examples translated from an SL into a TL and involving transposition are absolutely necessary in a bilingual dictionary.

From a dictionary user's point of view, an even more difficult situation arises if the translation of the lemma in context requires the addition of some elements and changes in the grammatical structure, since individual words are translated in context rather than in isolation. The addition of lexical elements and changes in grammatical structure are usually closely connected and depend on a careful study of the wider context in an SL. Apart from that, certain syntactic, collocational, culture-specific and pragmatic differences always exist between two languages, differences of which the user should be made aware to the greatest extent possible. Illustrative examples represent the only part of the dictionary entry that enables the lexicographer to draw the user's attention to these differences. It should, however, be stressed that illustrative examples should be translated, since lexical and grammatical changes in the translation of illustrative examples demand not only a high level of knowledge of both languages, but also translation abilities that cannot be expected from the user of

a bilingual dictionary, who is usually not a linguist with translation experience.

In an English–Slovene decoding dictionary, we can find quite a large number of examples translated into Slovene by one-word equivalents. In the SL, these word combinations are often collocations, which are normally included in dictionaries as examples. They have not yet acquired the status of a compound; thus, they are not included as independent lemmata. This can be explained by the relatively low frequency, as demonstrated by various corpora as well as by the failure to perceive them as multi-word lexical items with their own meaning, which means that they have not yet undergone the process of lexicalization. When we check, for example, the frequency of some of the examples mentioned in the previous section under 3), we discover that the example *nude photograph* has 5 hits in the BNC and 14 hits in ukWaC; similarly, *man of the theatre* is also not in common use (BNC: 7 hits, ukWaC: 30 hits), and the same applies to *theatre sister* (BNC: 3 hits, ukWaC: 56 hits) and, to a lesser extent, to *term of abuse* (BNC: 29 hits, ukWaC: 145 hits).

Zero equivalence is another important criterion for the inclusion of illustrative examples in a bilingual decoding dictionary. If there is no equivalent in the TL, no dictionary equivalents can be offered; consequently, the meaning of a lemma can only be conveyed by means of illustrative examples. Since untranslated examples are of no help to the dictionary user, who is not familiar with the meaning of the lemma in question, the translation of examples is an absolute must.

A dictionary equivalent that is of a more general type does not prove useful when dealing with pseudo-phraseological illustrative examples that require idiomatic translation. This is a strong argument for the inclusion of such examples in a bilingual dictionary, as it is the only way to make the user aware of potential pitfalls and to offer him/her a translation acceptable from the point of view of the TL.

To sum up, the findings of our research investigating the inclusion of illustrative examples in an English–Slovene decoding dictionary clearly show that the dictionary equivalents provided can be semantically and functionally appropriate, without satisfying the criterion of translation equivalence. This is the main reason that illustrative examples should be included to provide the user with all possible translations that are considered acceptable in the TL (cf. Adamska-Sałaciak 2006: 499).

5. Conclusion

Our study confirms the importance of including illustrative examples in a bilingual decoding dictionary. Although we limit ourselves to an English–Slovene dictionary, the criteria for the selection of illustrative examples can be applied to the compilation of any bilingual dictionary intended for decoding. Examples should be selected with care to clearly show contrastive differences between the two languages. Examples that can be translated by the immediate insertion of a dictionary equivalent can be regarded as superfluous, since the

user can translate them without difficulty. Examples that cannot be translated by the given dictionary equivalents without employing changes prove to be very useful if included in a bilingual decoding dictionary, since it is not to be expected that an average dictionary user would be able to comprehend them fully, let alone to translate them suitably and correctly. This leads to the conclusion that it is absolutely necessary and important to translate illustrative examples into the TL. If the examples demonstrating contrastive differences between the two languages are not translated, the user is not made aware of the complex differences between the two languages; consequently, it can be claimed that the dictionary does not fulfil one of its main tasks: to clearly show how two different languages function in everyday use.

To conclude, it is important to adopt the approach of functional lexicography and to seek a compromise solution that would meet the user's needs while simultaneously accommodating a precise lexicographic description.

Endnotes

1. sam. = noun
2. prid. = adjective
3. gl. = verb
4. brit. = British English

References

Dictionaries

Angleško-slovenski slovar Password. 1999. Ljubljana: DZS.

Krek, S. (Ed.). 2005. *Veliki angleško-slovenski slovar Oxford-DZS*. Ljubljana: DZS d.d.

Štern, P. (Ed.). 2000. *Collins COBUILD Angleško-slovenski slovar BRIDGE*. Ljubljana: DZS.

Other Literature

Adamska-Salaciak, A. 2006. Translation of Dictionary Examples — Notoriously Unreliable? Corino, E., C. Marengo and C. Onesti (Eds.). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, Italia, 6-9 settembre 2006*. Volume 1: 493-501. Alexandria: Edizioni dell'Orso.

Adamska-Salaciak, A. 2013. Issues in Compiling Bilingual Dictionaries. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 213-231. London/New Delhi/New York/Sydney: Bloomsbury Academic.

Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Béjoint, H. 2000. *Modern Lexicography: An Introduction*. Oxford/New York: Oxford University Press.

Hausmann, F.J. et al. (Eds.). 1991. *Dictionaries. An International Encyclopedia of Lexicography*. Volume 3. Berlin: De Gruyter.

- Jackson, H.** 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Jacobsen, J.R., J. Manley and V.H. Pedersen.** 1991. Examples in the Bilingual Dictionary. Hausmann, F.J. et al. (Eds.). 1991: 2782-2789.
- Kromann, H.-P., T. Riiber and P. Rosbach.** 1991. Grammatical Constructions in the Bilingual Dictionary. Hausmann, F.J. et al. (Eds.). 1991: 2770-2775.
- Krvina, D.** 2014. Problematika zglede v novem slovarju slovenskega jezika (med čim večjo avtentičnostjo in leksikografsko učinkovitostjo). *Posvet o novem slovarju slovenskega jezika: Novi slovar za 21. stoletje*. Ministrstvo za kulturo 12 February 2014. Available at: http://www.mk.gov.si/fileadmin/mk.gov.si/pageuploads/Ministrstvo/slovenski_jezik/E_zbornik/16-_Domen_Krvina_-_Problematika_zglede_v_NSJ.pdf [10 December 2015].
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York/Cambridge: Cambridge University Press.
- Newmark, P.** 1988. *Approaches to Translation*. New York: Prentice Hall.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Toope, M.** 1996. *Examples in the Bilingual Dictionary*. Unpublished M.A. Thesis. Ottawa: University of Ottawa. [Online]. Available at: <http://www.ruor.uottawa.ca/bitstream/10393/9942/1/MM15769.PDF> [28 January 2016].
- Van Scherrenburg, D.** 1990. *The Arrangement of Information in the General Bilingual Dictionary*. Unpublished M.A. Thesis. Ottawa: University of Ottawa. [Online]. Available at: <http://www.ruor.uottawa.ca/bitstream/10393/5625/1/MM62335.PDF> [10 January 2016].
- Zöfgen, E.** 1991. Bilingual Learner's Dictionaries. Hausmann, F.J. et al. (Eds.). 1991: 2888-2903.

Corpora

- BNC (https://the.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/bnc2;)
ukWaC (https://the.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/ukwac3;)

Über die Konstruktion einer hierarchischen Textverbundgesamtstruktur. Ein Beitrag zur Theorie der Wörterbuchform

Helmut Henne zum 80sten Geburtstag gewidmet

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg,
Bundesrepublik Deutschland (wiegand.h.e.oberurff@t-online.de)*

Zusammenfassung: Nach einem kurzen kritischen Einblick in die Forschung zu Textverbundgesamtstrukturen in den Jahren 1988 bis 2016 wird zunächst ein Ausschnitt aus einer Typologie von Textverbundtypen und hierarchischen Textverbundstrukturen präsentiert und kommentiert. Nach einer terminologischen Erörterung, insbesondere zum Begriff der lexikographischen Konstituente, wird als Beispiel eine hierarchische Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur so schrittweise konstruiert, dass ihre Teilstrukturen, nämlich die hierarchische rechtserweiterte Textverbundkonstituentenstruktur, ihre hierarchische konstituentenlose Vor- und Nachspannstruktur, ihre hierarchische Wörterverzeichnisstruktur und ihre hierarchische Artikeltextstrukturen getrennt konstruiert und dann zusammengefügt werden. In diesem Prozess werden eine Reihe neuer Begriffe geprägt und zugehörige Termini eingeführt, wie z.B. *Vor- und Nachspannsektor*, *hybride Vor- und Nachspannkonstituente*, *vorderer Vorspann* und *hinterer Nachspann*. Weiterhin wird überlegt, da Wörterbücher und Gebrauchsgegenstände mit lexikographischen Formeigenschaften immer mindestens eine äußere Zugriffsstruktur aufweisen, ob und wie wenigstens äußere alphabetische Hauptzugriffsstrukturen bei der Konstruktion von hierarchischen Textverbundgesamtstrukturen berücksichtigt werden können. Dazu wird abschließend ein Vorschlag gemacht.

Schlagwörter: ARTIKELSTRECKE, ARTIKELSTRECKENSTRUKTUR, ÄUSSERE ZUGRIFFSSTRUKTUR, FUNKTIONAL-POSITIONALE SEGMENTATION, HIERARCHISCHE KONSTITUENTENLOSE NACHSPANNSTRUKTUR, HIERARCHISCHE KONSTITUENTENLOSE VORSPANNSTRUKTUR, HIERARCHISCHE PARTIELL KONSTITUENTELOSE NACHSPANNSTRUKTUR, HIERARCHISCHE TEXTVERBUNDGESAMTSTRUKTUR, HIERARCHISCHE TEXTVERBUNDKONSTITUENTENSTRUKTUR, HIERARCHISCHE PARTIELL KONSTITUENTENLOSE VORSPANNSTRUKTUR, HINTERER NACHSPANN, HYBRIDE NACHSPANNKONSTITUENTE, HYBRIDE VORSPANNKONSTITUENTE, LEXIKOGRAPHISCHE KONSTITUENTE, NACHSPANN, NACHSPANNSEKTOR, NICHTFUNKTIONALE SEGMENTATION, POSITIONSGESTEUERTE GRUPPIERUNGSMETHODE, RECHTSERWEITERTE TEXTVERBUNDKONSTITUENTENSTRUKTUR, VORDERER VORSPANN, VORSPANNSEKTOR

Abstract: On the Construction of a Hierarchical Comprehensive Text Compound Structure. A Contribution to the Theory of Dictionary Structures.

After a brief critical look at research on comprehensive text compound structures in the period 1988 to 2016 an excerpt of a typology of types of text compounds and hierarchical text compound structures is presented and commented on. Following a terminological discussion, especially regarding the concept of lexicographic constituents, a hierarchical comprehensive text compound structure with right expanded text compound structure is constructed as an example step by step, in such a way that its partial structures, i.e. the hierarchical right expanded text compound constituent structure, its hierarchical constituentless front and back matter structure, its hierarchical word list structure and its hierarchical article stretch structures, are constructed separately and are then combined. In this process a number of new concepts are coined and relevant terms are introduced, e.g. *front* and *back matter sector*, *hybrid front* and *back matter constituents*, *front front matter* and *back back matter*. In addition, because dictionaries and utility objects with lexicographical structural features always display at least one outer access structure, it is considered if and how at least outer alphabetical main access structures can be considered in the construction of hierarchical comprehensive text compound structures. In conclusion a proposal is made in this regard.

Keywords: ARTICLE STRETCH, ARTICLE STRETCH STRUCTURE, BACK BACK MATTER, BACK MATTER, BACK MATTER SECTOR, FRONT FRONT MATTER, FRONT MATTER SECTOR, FUNCTIONAL-POSITIONAL SEGMENTATION, HIERARCHICAL COMPREHENSIVE TEXT COMPOUND STRUCTURE, HIERARCHICAL CONSTITUENTLESS BACK MATTER STRUCTURE, HIERARCHICAL CONSTITUENTLESS FRONT MATTER STRUCTURE, HIERARCHICAL PARTIAL CONSTITUENTLESS BACK MATTER STRUCTURE, HIERARCHICAL PARTIAL CONSTITUENTLESS FRONT MATTER STRUCTURE, HIERARCHICAL TEXT COMPOUND CONSTITUENT STRUCTURE, HYBRID BACK MATTER CONSTITUENTS, HYBRID FRONT MATTER CONSTITUENTS, LEXICOGRAPHIC CONSTITUENTS, NON-FUNCTIONAL SEGMENTATION, OUTER ACCESS STRUCTURE, POSITIONAL-GUIDED GROUPING METHOD, RIGHT EXPANDED TEXT COMPOUND CONSTITUENT STRUCTURE

1. Vorbemerkung

Dieser Beitrag versucht, einen Weg zu zeigen, wie man bei der Erforschung und Darstellung einer Struktur von initialalphabetischen Printwörterbüchern, die mit guten Gründen *hierarchische Textverbundgesamtstruktur* heißen kann, zu Fortschritten kommt.

2. Kritischer themaspezifischer Einblick in den metalexikographischen Diskurs von 1988 bis 2016

„Niemand hat die Absicht, eine Mauer zu bauen“ für die Lektüre dieses Beitrags in Form eines gründlichen und klassischen Forschungsberichts. Vielmehr besteht die Absicht, ein Fenster zu öffnen für einen kritischen (und damit auch selbstkritischen) Blick auf diejenigen Arbeiten, die sich mit der Problematik der „globalen Strukturen“ von Printwörterbüchern befasst haben. Mit dem Aus-

druck *globale Struktur* soll nicht en passant ein neuer metalexikographischer Terminus eingeführt werden; vielmehr wird der aus der germanistischen Textlinguistik der 70er und 80er Jahre stammende Ausdruck für die Gesamtstruktur von Texten als unverbindlich charakterisierende Benennung für eine textuelle Struktur verwendet, für die ein eingeführter und definierter Terminus an dieser Stelle noch nicht verfügbar ist.

Die moderne Metalexikographie ist in Frankreich entstanden. Besonders Rey-Debove 1971 verdankt die spätere Forschung den geschärften Blick auf Strukturen im Wörterbuch, wie z.B. den mit Hilfe der Termini *macrostructure* und *microstructure*. Wie aber eine globale Struktur zu erfassen, darzustellen und zu benennen ist, als deren Teilstrukturen die Makro- und die Mikrostruktur hätten gelten können, wurde nicht gefragt. Diese Frage wurde, soweit ich sehe, zum ersten Mal in Wiegand 1988 gestellt und damit zu einem Zeitpunkt, zu dem die Theorie der Wörterbuchform noch in den Kinderschuhen steckte. Der in Wiegand 1988 vorgeschlagene Terminus für die globale Struktur von Printwörterbüchern war *Hyperstruktur*. Es zeigte sich aber bald, dass sich dieser Terminus — wahrscheinlich wegen seiner Nähe zu *Hypertext* und *Hypertextstruktur* — nicht durchsetzen konnte. Auch andere haben die sich damals in der Forschung auftuende terminologische Lücke bemerkt. So formuliert z.B. Nielsen (1990: 49): “The lexicographic macrostructure is not merely the arrangement of lemmata, but a structure which applies to the dictionary as a whole”, und Hartmann und James (2001: 93) schlagen *megastructure* vor. Beide Termini haben sich ebenfalls nicht durchsetzen können, was wahrscheinlich seinen Grund darin hat, dass mit der isolierten Einführung von Termini ohne zugehörigen theoretischen Hintergrund kein wirklicher Erkenntnisfortschritt zu erzielen ist. Es sei aber erwähnt, dass *Megastruktur* in Bielińska (2010: 294) als Synonym für *Textverbundkonstituentenstruktur* verwendet wird, ein Vorschlag, dem hier nicht gefolgt wird. Hausmann und Wiegand (1989: 330ff.) nennen eine hierarchische und damit eingeschränkte Form einer globalen Struktur von Printwörterbüchern, in der weder die Makrostruktur noch eine Zugriffsstruktur eine Rolle spielen, *textual book structure* und präsentieren in Englisch (wahrscheinlich) den ersten Baumgraphen zu einer hierarchischen globalen Struktur eines deutschen Printwörterbuchs, der in Wiegand (1988: 49, Abb. 6) einen Vorläufer hat. In der Zwischenzeit hat sich der Begriff, der zum Terminus *Makrostruktur* gehört, stark verändert (vgl. Wiegand 2008: 229ff. u. Wiegand und Gouws: 78ff.). Es wird daher zu fragen sein, ob man zumindest Hauptzugriffsstrukturen nicht bei der Konstruktion von Textverbundgesamtstrukturen berücksichtigen sollte (vgl. 5.). Während *textual book structure* nach 1989 nur selten verwendet wurde, weil andere Begrifflichkeiten eingeführt wurden, spielte der erwähnte Baumgraph in seiner abstrakten Form und damit ohne direkten Bezug zu einem bestimmten Wörterbuch bei den Bemühungen, die globale Struktur von Printwörterbüchern zu konstruieren, längere Zeit eine zentrale und z.T., aus heutiger Sicht, eine fatale Rolle. In Abb. 1 ist er erneut dargestellt, und zwar in der Form, wie er sich in Kammerer und Wiegand 1998 (228, Abb. 4) findet und damit in einer um die als Identifikationsmarken dienenden Ziffern 1-11

erweiterten Form. Diese numerische Erweiterung wirkt sich auf die strukturellen Eigenschaften der Baumgraphen nicht aus.

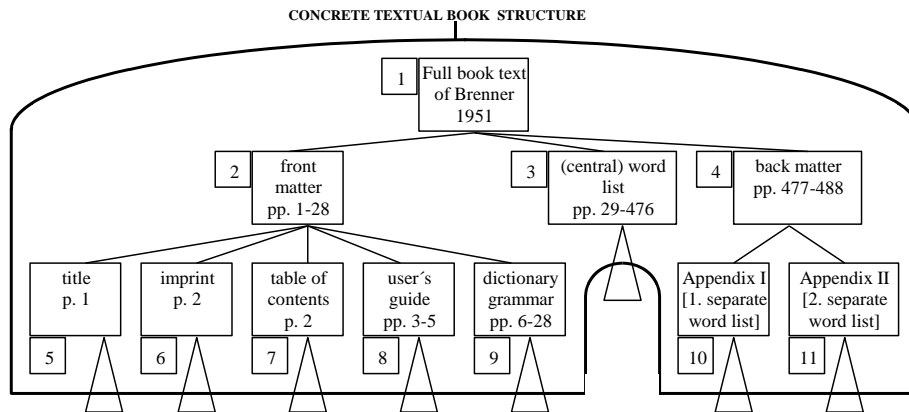


Abb. 1: Konkrete textual book structure von Brenner (1951) nach Hausmann und Wiegand (1989: 332, Abb. 36.2) in der von Kammerer und Wiegand (1998: 228, Abb. 4) um die Identifizierungsmarken 1-11 erweiterten Form; *Abkürzungen:* p. = page

Wenn man die Attribuierung des Baumgraphen, die in Abb. 1 nur als metalexikographische Etikettierung der Knoten gegeben ist, nicht berücksichtigt, sondern ausschließlich seine mathematischen Eigenschaften, dann ist der Baumgraph korrekt, weil er (a) genau einen Knoten aufweist, nämlich den Wurzelknoten, der ohne Vorgänger ist und von dem zu jedem anderen Knoten genau ein Pfad existiert, und weil (b) jeder Knoten (außer dem Wurzelknoten) genau einen Vorgänger hat, und weil schließlich (c) die Nachfolger eines Knotens von links nach rechts geordnet sind (vgl. z.B. Klabunde 1998: 32ff.). Berücksichtigt man jedoch die Attribuierung, dann gilt die Kritik von Kammerer und Wiegand 1998, die das Folgende beinhaltet: Es ist nicht berücksichtigt, dass *erstens* gilt: 5, 6, 7, 8 und 9 sind nichtfunktionale Teile von 2 und zugleich funktionale Teile von 1 sowie weiterhin *zweitens*, dass gilt: 10 und 11 sind nichtfunktionale Teile von 4 und zugleich funktionale Teile von 1 und schließlich, dass *drittens* gilt: 2, der Vorspann (front matter), und 4, der Nachspann (back matter), sind nichtfunktionale Teile von 1. Dies bedeutet: In Abb. 1 fehlen Kanten, also Knotenpaare $\langle k_1, k_2 \rangle$, die durch „ — “ repräsentiert sind, und durch „ — “ repräsentierte Knotenpaare sind Elemente unterschiedlicher Relationen vom Typ der partitiven Relation! An einem Beispiel erläutert, besagt diese Kritik: 8, die Benutzungshinweise, sind ein nichtfunktionaler Teil des Vorspanns, denn ihre genuine Funktion besteht nicht darin, dem Benutzer den Vorspann zu erklären, sondern darin, dem Benutzer Hinweise für die Benutzung des Wörterbuchs zu geben; daher bezieht sich die genuine Funktion von 8 als Teil von 2

auf 1, also das gesamte Wörterbuch, so dass 8 ein funktionaler Teil von 1 ist. Die Kritik von Kammerer und Wiegand 1998 ist jedoch nicht weitgehend genug, wie erst später gezeigt werden kann, wenn weitere Begriffe und die zugehörigen Termini zur Verfügung stehen.

3. Textverbundtypen und Typen von hierarchischen Textverbundstrukturen

Will man darlegen, wie eine hierarchische Textverbundgesamtstruktur konstruiert werden kann, muss *erstens* zunächst klar sein, was in metalexikographischer Perspektive unter einem Textverbund verstanden werden soll, wozu auch gehört, dass man einen Überblick über die wichtigsten Textverbundtypen hat. Erst danach kann man *zweitens* im Rahmen der Theorie der Wörterbuchform festlegen, was als ein Typ von hierarchischer Textverbundstruktur gelten soll und zu welchem Typ von Textverbund er gehört.

Im Kontrast zu der Auffassung, Wörterbücher seien Texte, die zu einer Textsorte gehören (vgl. z.B. Fenner 1997), wurden in Wiegand 1995 die Termini *Textträger* und *Textverbund* in die Metalexikographie eingeführt. Papierne Textträger sind Zusammenordnungen von schriftkonstituierten Texten unterschiedlicher Textsorten zu gedruckten Gebrauchsgegenständen als statischen Informationssystemen; es lassen sich unterschiedliche Typen von papiernen Textträgern unterscheiden, wie z.B. Zeitschriften, Gemeindeblätter, Programmschriften von Gymnasien, Tageszeitungen und alle gedruckten Gebrauchsgegenstände mit lexikographischen Formeigenschaften (i. S. v. Wiegand 2013), wie z.B. Telefonbücher und Reiseführer, deren gezielte selektive Nutzung mit spezifischen Suchfragen dadurch ermöglicht wird, dass sie intern geordnete und extern akzessive Daten aufweisen, und schließlich Wörterbücher aller Typen. Textträger haben sich historisch aus praktisch-pragmatischen Gründen in jeweiligen Kulturräumen relativ gleichartig entwickelt. Von den genannten Typen von Textträgern sind nur einige, die zum Typ der gedruckten Gebrauchsgegenstände mit lexikographischen Formeigenschaften gehören, wie z.B. Telefonbücher (vgl. dazu Wiegand 2013), sowie alle, die zum Typ des Wörterbuchs gehören, Textverbunde; denn für Texte, die zu einem Textverbund gehören, gilt: Sie sind inhaltlich-funktional und formstrukturell aufeinander bezogen, so dass eine verbundinterne Intertextualität gegeben ist, zu der auch gehört, dass es mindestens einen Metatext gibt, in dem die Benutzung des Textverbundes erklärt wird. Dies gilt immer für Wörterbücher, aber nicht z.B. für Tageszeitungen. Um nicht ständig von einem *lexikographischen Textverbund vom Typ des Wörterbuchs* etc. sprechen zu müssen, wird hiermit festgelegt, dass in diesem Beitrag mit *Textverbund* stets über ein Printwörterbuch gesprochen wird. Diese Ansage ist aus mehreren Gründen erforderlich. Denn in einer zeitgemäßen erweiterten Auffassung von Lexikographie, die letztere als Teilgebiet der Informationswissenschaft auffasst (vgl. Wiegand 2013a), so dass u.a. auch gedruckte Gebrauchsgegenstände mit lexikographischen Formeigenschaften (i.S.v.: Wiegand 2013) zu ihrem Gegenstandsbereich gehören, treten auch

andere Textverbunde als Wörterbücher auf, wie z.B. Telefon- und Faxbücher (vgl. Wiegand 2013: 293ff.).

In Abb. 2 wird nachfolgend ein Typologigraph zu einem Ausschnitt aus einer Typologie von hierarchischen Textverbundstrukturen präsentiert, anhand dessen zugleich der zugehörige parallele Ausschnitt von Textverbunden, die Wörterbücher sind, erschließbar ist.

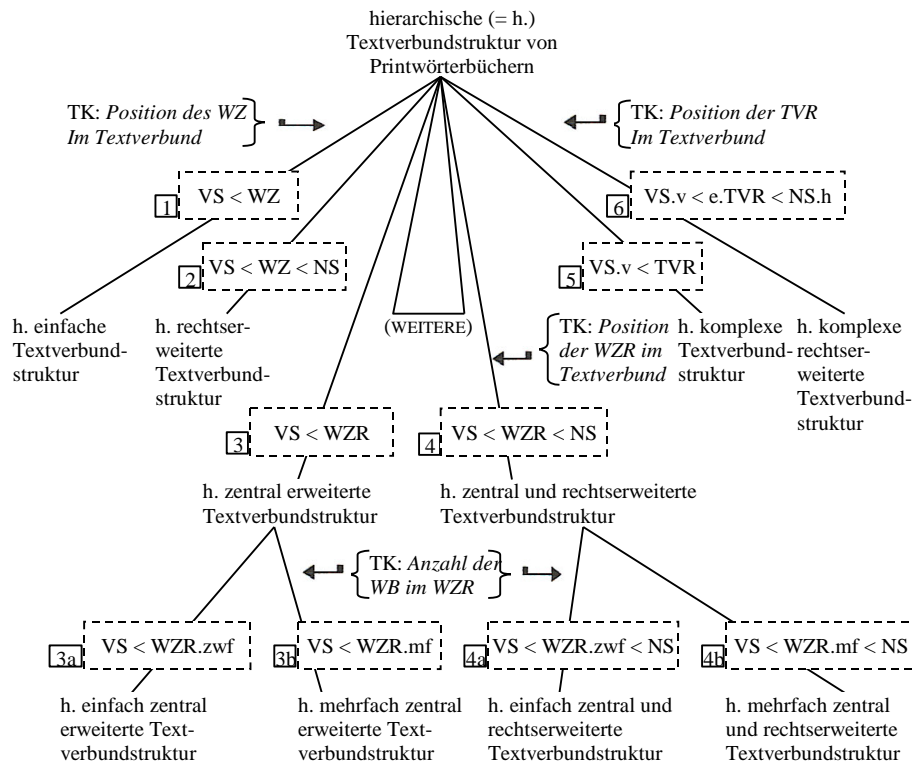


Abb. 2: Typologigraph zu einem Ausschnitt aus einer Typologie von hierarchischen Textverbundstrukturen von Printwörterbüchern, anhand dessen zugleich die zugehörigen Textverbundtypen erschlossen werden können; *Abkürzungen:* VS = Vorspann; WZ = Wörterverzeichnis; NS = Nachspann; WZR = Wörterverzeichnisreihe; VS.v = vorderer Vorspann; TVR = Textverbundreihe; e.TVR = eingebettete Textverbundreihe; NS.h = hinterer Nachspann; WZR.zwf = zweifache WZR (i.S.v. Wörterverzeichnisreihe, bestehend aus zwei WZ); WZR.mf =mehrfache WZR (i.S.v. Wörterverzeichnisreihe, bestehend aus mehr als zwei WZ); TK = Typologiekriterium. Vorläufer von Abb. 2 sind die Abb. 4-8 in Wiegand 2001: 146 und Figura 3-5 in Wiegand und Fuentes Morán 2009: 45 sowie die dazugehörige englische Version in Wiegand und Gouws 2013: 62

In der Theorie der Wörterbuchform wird nicht nur mit Artikelstrukturschemata (vgl. WLWF-1: 592 u. Wiegand 2010: 313f), sondern u.a. auch mit Textverbundstrukturschemata zur summarischen Kurzdarstellung von Strukturen gearbeitet; in letzteren wird das Klassensymbol TV für die Klasse der Textverbunde sowie die Klassensymbole für Teile von Textverbunden, wie z.B. VS, WZ etc. verwendet, sowie das Symbol „:“ für *besteht aus* und das Symbol „<“ für *geht voraus*. Beispiele für Textverbundstrukturschemata sind: TV:VS < WZ und TV:VS < WZ < NS. In Abb. 2 werden die definitiven Teile von Textverbundstrukturschemata, also die, die auf das Symbol „:“ folgen, verwendet, um den Textverbundtyp (= TV-Typ) kriterial zu kennzeichnen, zu dem der Typ der Textverbundstruktur gehört, der erhältlich ist, wenn man das zugehörige Typologiekriterium TK auf den Ausgangstyp anwendet. Wendet man z.B. das Typologiekriterium *Position des WZ im Textverbund* auf den Ausgangstyp der hierarchischen Textverbundstruktur an, sind der Typ 1, also der Typ der hierarchischen einfachen Textverbundstruktur erhältlich, und weil das Typologiekriterium eine zweifache Ausprägungsstruktur aufweist (vgl. dazu Wiegand 1998: 845ff.), auch der Typ 2, also der hierarchischen rechts-erweiterten Textverbundstruktur. Zum Typ 1 gehört das Textverbundstrukturschema (= TV-Str-S): TV:VS < WZ, und zum Typ 2 das TV-Str-S: TV:VS < WZ < NS. Nach diesen exemplarischen Erläuterungen dürfte deutlich sein, dass anhand des Typologiegraphen in Abb. 2 zu einem Ausschnitt aus der Typologie der hierarchischen Textverbundstrukturen eindeutig der zugehörige Ausschnitt aus einer Typologie von Textverbunden angegeben werden kann. Der Typ 1 ist deswegen der *einfache Textverbund*, der auch *basaler Textverbund* heißt, weil ein „Wörterbuch“ ohne Vorspann (VS) nur ein Wörterbuchfragment ist.

Zu jedem in Abb. 2 genannten Textverbundtyp können zahlreiche Wörterbücher als Beispiele angegeben werden; damit das Literaturverzeichnis nicht zu lang wird, werden nur wenige Beispiele genannt. Weiterhin kann zu jedem Typ von Textverbundstruktur ein allgemeines Textverbundstrukturbild nach formalen Prinzipien konstruiert werden; aus Platzgründen werden hier nur einige präsentiert.

- (1) TV-Typ 1 mit dem TV-Str-S: TV:VS < WZ; Beispiele: HWDG 1984, Kluge 1995, Köster 1969 (vgl. 5.), Agricola 1992.
- (2) TV-Typ 2 mit dem TV-Str-S: TV:VS < WZ < NS; Beispiele: Paul 2002 (vgl. 4.), Wahrig 2006.

Das allgemeine Textverbundstrukturbild für den Typ der hierarchischen rechts-erweiterten Textverbundstruktur hat die Form in Abb. 3.

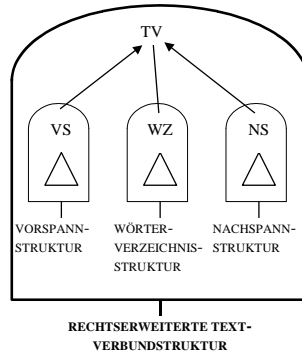


Abb. 3: Allgemeines Textverbundstrukturbild zum Typ der hierarchischen rechtserweiterten Textverbundstruktur mit dem Textverbundstrukturschema: $TV : VS < WZ < NS$; *Darstellungskonventionen:* „ $x \text{ — } y$ “ bedeutet so viel wie x ist eine Textverbundkonstituente von y ; „ $x \rightarrow y$ “ bedeutet so viel wie x ist ein nichtfunktionaler Teil von y .

- (3) TV-Typ 3 mit dem TV-Str - S: $TV : VS < WZR$; zum TV-Typ 3 gehören zwei Untertypen, die durch die Anwendung des Typologiekriteriums *Anzahl der WZ in der WZR* auf den TV-Typ 3 gegeben sind, nämlich TV-Typ 3a mit dem TV-Str-S: $TV : VS < WZR.zwf$ sowie der TV-Typ 3b mit dem TV-Str-S: $TV : VS < WZR.mf$; Beispiel: Dultz 1965.

Das allgemeine Textverbundstrukturbild für den Typ der hierarchischen einfach zentral erweiterten Textverbundstruktur (zum TV-Typ 3a) findet sich in Abb. 4.

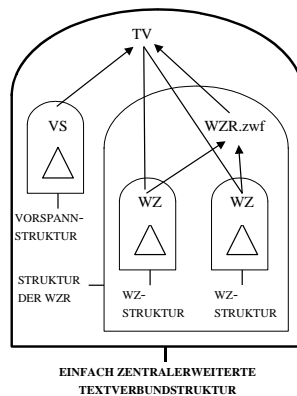


Abb. 4: Allgemeines Textverbundstrukturbild zum Typ der hierarchischen einfach zentral erweiterten Textverbundstruktur mit dem Textverbundstrukturschema: $TV : VS < WZR.zwf$; *Darstellungskonventionen:* wie in Abb. 3.

- (4) TV-Typ 4 mit dem TV-Str-S: $TV:VS < WZR < NS$; Beispiele: Wittstock und Kauczor 1988, Telling 1988.
- (5) TV-Typ 5 mit dem TV-Str-S: $TV:VS.v < TVR$; Beispiele: Duden-GFWB.
- (6) TV-Typ 6 mit dem TV-Str-S: $TV:VS.v < e.TVR < NS.h$. Beispiele: Kienle 1965, Kämpfert 1997.

Das allgemeine Textverbundstrukturbild für den Typ der hierarchischen komplexen rechtserweiterten Textverbundstruktur findet sich in Abb. 5.

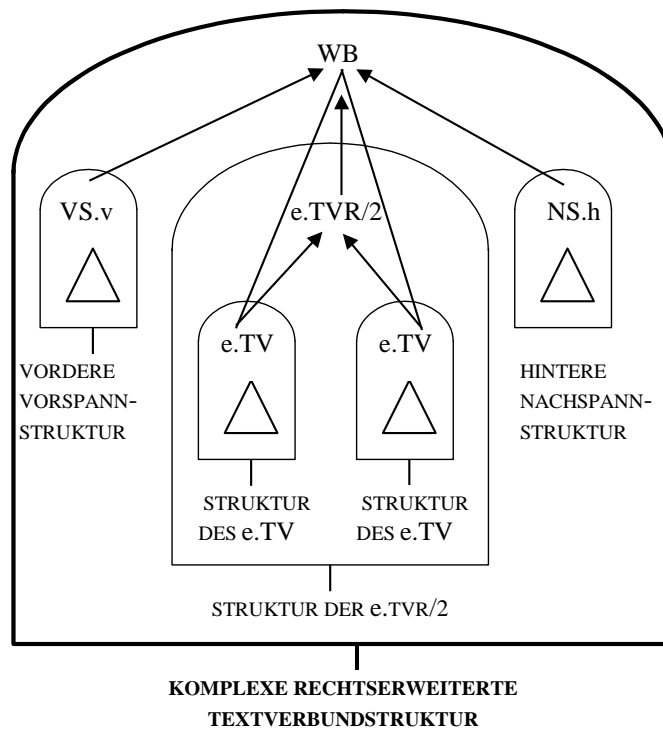


Abb. 5: Allgemeines Textverbundstrukturbild zum Typ der hierarchischen komplexen rechtserweiterten Textverbundstruktur mit dem Textverbundstrukturschema: $VS.v < e.TVR < NS.h$; *Abkürzungen:* $e.TVR/2$ = eingebettete Textverbundreihe, die aus zwei Textverbunden besteht. *Darstellungskonventionen:* wie in Abb. 3.

Der Typologiegraph in Abb. 2 stellt nur einen Typologieausschnitt dar; er hat daher einen Teil mit dem Hinweis: WEITERE. Einige dieser in Abb. 2 nicht berücksichtigten Textverbundtypen sind die folgenden:

- (7) Der wörterverzeichnisintern erweiterte Textverbund, der dann gegeben ist, wenn das Wörterverzeichnis einen Einschub aufweist: Im deutlichen Unterschied zu einem eingelagerten Binnentext, der eine Artikelstrecke erweitert, deren unmittelbare Textkonstituente er ist, ist ein Einschub (E) eine Textverbundkonstituente, die das Wörterverzeichnis aufspaltet, so dass ein vorderer Wörterverzeichnisteil (WZT.v) und ein hinterer Wörterverzeichnisteil (WZT.h) gegeben ist. Dementsprechend ist eine wörterverzeichnisintern erweiterte Textverbundstruktur erhältlich mit dem Textverbundschemata $TV:VS < WVT.v < E < WZT.h$. Ein Beispiel ist: OSACD 2006.
- (8) Der wörterverzeichnisintern und rechtserweiterte Textverbund, zu dem das Textverbundschemata $TV:VS < WVT.v < E < WZT.h < NS$ gehört; ein Beispiel ist CALD 2008.

Mit den Typen (7) und (8) dürfen die beiden folgenden Textverbundtypen (9) und (10) nicht verwechselt werden, die häufig bei biskopalen Wörterbüchern auftreten, nämlich

- (9) der Textverbundtyp mit diskontinuierlicher zweifacher Wörterverzeichnisreihe (WZ[E]R.zwf), in dem die beiden Wörterverzeichnisse durch einen Einschub getrennt sind, der weder ein Vor- noch ein Nachspann ist. Das zugehörige Textverbundschemata lautet: $TV:VS < WZ[E]R.zwf$.
- (10) Der rechtserweiterte Textverbund mit diskontinuierlicher zweifacher Wörterverzeichnisreihe, für deren Einschub ebenfalls gilt, dass er weder ein Vor- noch ein Nachspann ist; das zugehörige Textverbundschemata lautet: $TV:VS < WZ[E]R.zwf. < NS$. Beispiele sind: Hueber DaF 2006, MWALED 2008 und PONS 2006.

Weiterhin treten Textverbundtypen auf, für die eine sprachliche Bezeichnung zu lang ist, um als Terminus funktionieren zu können. Ein Beispiel ist Sellner 1997. Zu diesem Wörterbuch gehört das folgende Textverbundschemata: $TV:VS.v < WZR/6 < TV.e < NS.h$ (mit WZR/6 = Wörterverzeichnisreihe, bestehend aus sechs Wörterverzeichnissen und mit TV.e = einfacher Textverbund). Als Bezeichnung für die Textverbundstruktur gilt dann: *Textverbundstruktur vom Typ: $TV:VS.v < WZR/6 < TV.e < NS.h$* . Ein weiteres Beispiel ist Mackensen 1956. Zu diesem Wörterbuch gehört das folgende Textverbundschemata: $TV:VS.v < TVR/4 < NS.h$ (mit VS.v = vorderer Vorspann; TVR/4 = Textverbundreihe, bestehend aus vier Textverbunden). Als Bezeichnung für die Textverbundstruktur gilt dann: *Textverbundstruktur vom Typ: $TV:VS.v < TVR/4 < NS.h$* .

4. Zur Konstruktion von hierarchischen Textverbundgesamtstrukturen

Nach dem Überblick über die zentralen Textverbundtypen und zugehörigen Typen von hierarchischen Textverbundstrukturen im vorausgehenden dritten

Abschnitt sollten wohl folgende Feststellungen akzeptabel sein:

(1) Eine hierarchische Textverbundgesamtstruktur kann sinnvoller Weise nur relativ zu einem Typ von Textverbundstruktur konstruiert werden, so dass es offensichtlich nicht den einen Typ von Textverbundgesamtstruktur geben kann, sondern nur mehrere Typen; daher liegt es nahe, die Typennamen für die Typen von Textverbundgesamtstrukturen — soweit dies möglich ist — an die Typennamen von Textverbundstrukturen anzuschließen; dann sind z.B. folgende hierarchische Textverbundgesamtstrukturen unterscheidbar:

- hierarchische Textverbundgesamtstruktur mit einfacher Textverbundstruktur
- hierarchische Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur
- hierarchische Textverbundgesamtstruktur mit zentral und rechtserweiterter Textverbundstruktur usw. (vgl. Abb. 2).

Hierarchische Textverbundgesamtstruktur ist dann ein hyperonymer Terminus für die Termini der gerade genannten Typen von hierarchischen Textverbundgesamtstrukturen.

(2) Wenn von einer Textverbundgesamtstruktur irgendeines Typs gesprochen wird, dann darf das keinesfalls so verstanden werden, als wären alle bekannten Strukturen, die zu einem Textverbund als Teilstrukturen gehören können, von der Textverbundkonstituentenstruktur bis zu den Strukturen der Angaben berücksichtigt. Derart hypertrophe Monsterstrukturen wären weder sinnvoll noch übersichtlich darstellbar. Sinnvoll ist die Konstruktion von Textverbundgesamtstrukturen nur dann, wenn sie dem Zweck dienen, eine gute Übersicht über die Distribution zentraler lexikographischer Textdatentypen und deren Gruppierung zu hierarchischen Teilstrukturen des Textverbundes zu gewährleisten gemeinsam mit den Datendistributionsstrukturen (i.S.v. Bergenholtz, Tarp und Wiegand 1999: 1778ff.). Sind sie zweckgerecht konstruiert, sind sie daher bei der Planung von größeren Wörterbüchern nützlich sowie insbesondere bei der Beurteilung, ob die gesamte Datenorganisation im Textverbund relativ zu den angesetzten Wörterbuchfunktionen angemessen und benutzerfreundlich ist.

Zwar gibt es bereits ältere Versuche, hierarchische Textverbundgesamtstrukturen zu konstruieren (vgl. z.B. Wiegand 2000: 214 u. 2001: 137), aber die Konstruktionsprinzipien sowie die Analysekonventionen wurden nicht explizit genug formuliert, und die Terminologie ist teilweise schwankend und nicht hinreichend spezifisch. Es sind daher, bevor anhand eines Beispiels eine Textverbundgesamtstruktur schrittweise konstruiert wird, Vorklärungen verschiedener Art erforderlich.

Damit die begriffliche und damit die terminologische Vielfalt hier nicht zu groß wird, sei zunächst der empirische Gegenstandsbereich auf die initial-alphabetischen Wörterbücher eingeschränkt. Weiterhin sei klargestellt, dass

nichttypographische Strukturanzeiger, wie z.B. Zugriffsstrukturanzeiger **A**, **B**, ..., **Z** vor bzw. zwischen Artikelstrecken, im Unterschied z.B. zu Wiegand (2001: 137, Abb. 4-3) nicht berücksichtigt werden. Darüber hinaus muss in Erinnerung gerufen werden, dass es zwei terminologisch deutlich unterschiedene Typen von lexikographischen Konstituenten gibt, nämlich den Typ der Textverbundkonstituente und den der Textkonstituente, so dass *lexikographische Konstituente* oberbegrifflich verwendet wird. Die Konstituenten beider Typen haben gemeinsam, dass ihnen eine bestimmte individuelle Form, genau eine Position sowie mindestens eine Funktion im zugehörigen (nächst größeren) textuellen Ganzen zukommt. Nicht gemeinsam ist ihnen das Folgende: Bei den Textkonstituenten wird zwischen unmittelbaren und mittelbaren Textkonstituenten unterschieden. Bei den Textverbundkonstituenten ist das terminologisch unangemessen (oder: unüblich, unpraktisch etc.). Es ist aber keineswegs selbstverständlich, denn alle Typen von lexikographischen Konstituenten sind das Ergebnis einer korrekten Anwendung der Methode der funktional-positionalen Segmentation. Daher sei hier *expressis verbis* folgende terminologische Konvention (K) festgelegt: K_1 : Eine Unterscheidung zwischen unmittelbaren Textverbundkonstituenten und mittelbaren Textverbundkonstituenten n -ter Stufe (mit $n \geq 1$) ist nicht vorgesehen. K_1 hat folgende, in der Forschung sozusagen stets stillschweigend eingehaltene, aber deswegen keineswegs selbstverständliche terminologische Konsequenzen: Die Konstituenten einer Textverbundkonstituente (wie z.B. die von Benutzungshinweisen oder Abkürzungsverzeichnissen) gelten nicht als mittelbare Textverbundkonstituenten erster Stufe. Insbesondere gilt bei der Konstruktion von Textverbundgesamtstrukturen: Die Konstituenten eines Wörterverzeichnisses gelten nicht als mittelbare Textverbundkonstituenten (erster Stufe), sondern als unmittelbare Wörterverzeichnis-konstituenten, die *Artikelstrecken* heißen. Die typverschiedenen Konstituenten einer Artikelstrecke, wie z.B. Wörterbuchartikel und funktionale Artikelteilstrecken, wie z.B. Artikelnester und Artikelnischen sowie eingelagerte Binnentexte, gelten nicht als mittelbare Textverbundkonstituenten (zweiter Stufe), sondern als unmittelbare Artikelstreckenkonstituenten. Schließlich gelten alle Artikelkonstituenten, die auch *Artikeltextkonstituenten* heißen, wie z.B. Angaben und Angabetexte, nicht als Textverbundkonstituenten (dritter Stufe), sondern als Artikelkonstituenten. Es muss unbedingt berücksichtigt werden, dass alle lexikographischen Konstituenten, unabhängig davon, zu welchen Typen sie gehören, das Ergebnis einer korrekten Anwendung der Methode der funktional-positionalen Segmentation sind, und zwar in ihrer nichtexhaustiven Variante, weil nichttypographische Strukturanzeiger oben ausgeschlossen wurden.

4.1 Konstruktion einer hierarchischen Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur

Im Folgenden wird eine hierarchische Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur in ihrer abstrakten Ausprägung konstruiert. Die Konstruktion erfolgt in der Weise, dass nacheinander die hierarchischen

Teilstrukturen der genannten Struktur gebildet und dann zusammengefügt werden. Der Typ der hierarchischen rechtserweiterten Textverbundgesamtstruktur wurde gewählt, weil in der neueren Printlexikographie rechtserweiterte Textverbunde, also solche, für die das Textverbundstrukturschema TV:VS < WZ < NS gültig ist, besonders häufig auftreten. Als Beispielwörterbuch dient Paul 2002.

4.1.1 Die hierarchische rechtserweiterte Textverbundkonstituentenstruktur

Hierarchische Textverbundkonstituentenstrukturen sind als Teilstrukturen von hierarchischen Textverbundgesamtstrukturen — und diese Einschränkung ist strukturell relevant — dadurch erhältlich, dass ein Textverbund funktional-positional so segmentiert wird, dass nur ein Segmentationschritt erfolgt, so dass nur eine einstufige Struktur erhältlich ist (vgl. u.a. Schierholz und Wiegand 2004: 2009ff; Wiegand und Fuentes Morán 2009: 25ff; Wiegand, Beer und Gouws 2013: 57ff). Während in alphabetischen Wörterbüchern die Gestaltung der Form des Wörterverzeichnisses und die der kondensierten Wörterbuchartikel relativ stark standardisiert ist (vgl. z.B. Wiegand 1997), gilt das für wörterverzeichnisexterne Textverbundteile hinsichtlich ihrer Form und Reihenfolge sowie hinsichtlich ihrem Umfang und ihrem Inhalt weniger; lediglich die Reihenfolge der Titelei, des Impressum, des Inhaltsverzeichnisses und des Vorworts folgt in der Mehrzahl der Fälle der Tradition der Buchtypographie (vgl. auch Bielińska 2010: 51f.). Daher ist nicht immer eindeutig entscheidbar, ob ein Textteil als Textverbundkonstituente gelten soll oder nicht. Bei einer solchen Entscheidung kann das Inhaltsverzeichnis hilfreich sein, wenn es nicht fehlt, wie z.B. im Duden-GFWB. Sehr häufig fehlen in den Inhaltsverzeichnissen die Übersichten, Karten oder Verzeichnisse, die sich auf dem vorderen und hinteren Vorsatzblatt befinden; so auch im PONS-DaF 2016 und im Paul 2002, der ein Vor- und einen Nachspann mit integriertem Vorsatzblatt aufweist (vgl. unten). Diese werden im letztgenannten Einbänder in den Hinweisen zur Benutzung (XII) als „Didaktisierte Übersichten auf dem vorderen und hinteren Innendeckel“ bezeichnet. Bei der Übersicht auf dem vorderen Vorsatzblatt handelt es sich um eine farbige, benutzerfreundliche Übersicht mit dem Titel „Zusammenspiel von Sachregister und Wörterbuch“. Die ebenfalls farbige Übersicht auf dem hinteren Vorsatzblatt mit dem Titel „Schlüssel zu den Wörterbucheinträgen“ erklärt — wie in der neueren Lernerlexikographie auf dem vorderen und/oder dem hinteren Vorsatzblatt üblich (vgl. z.B. PONS-DaF 2016) — anhand von Wörterbuchartikeln deren Textsegmente. Beide Übersichten gelten nachfolgend als Textverbundkonstituenten.

Im Folgenden werden die abstrakte und isomorphe konkrete rechtserweiterte Textverbundkonstituentenstruktur von Paul 2002 (= wb) als Teilstrukturen der zugehörigen Textverbundgesamtstruktur gebildet; der Strukturbildungsprozess wird in vier methodischen Schritten durchgeführt (vgl. auch Wiegand 2010: 269ff.). Im *ersten* methodischen Schritt wird wb einstufig nicht-exhaustiv funktional-positional segmentiert. Das Segmentationsergebnis wird

nachfolgend aufgelistet; dabei wird mit Hilfe des Gleichheitszeichens „=“ den Textverbundkonstituenten ein Kleinbuchstabe als Individuennamen zugeordnet, mittels dessen sie unter Verwendung des Zeichens „∈“ für die Element-Klassen-Relation ihren Textverbundkonstituentenklassen zugeordnet werden, die für die Bildung der abstrakten Textverbundkonstituentenstruktur benötigt werden.

a = (Übersicht zum) Zusammenspiel von Sachregister und Wörterbuch; a ∈ ZSW

b = Titelei; b ∈ TIT

c = erweitertes Impressum; c ∈ IMP

d = Inhaltsverzeichnis; d ∈ IV

e = Zur Einführung; e ∈ EINF

f = Hinweise zur Benutzung; f ∈ BENH

g = Abkürzungen; g ∈ ABK 1

h = Laut- und Sonderzeichen; h ∈ LSZ

i = Abgekürzt zitierte Autoren und Werke; i ∈ ABK 2

j = Sachregister — Wegweiser zum Wortschatz; j ∈ SAR

k = Wörterbuch A bis *Zypresse* [Wörterverzeichnis]; k ∈ WZ

l = Literarische Quellen; l ∈ LQ

m = Wörterbücher und weitere Hilfsmittel; m ∈ WÖHI

n = Schlüssel zu den Wörterbucheinträgen; n ∈ WBE

Im *zweiten* methodischen Schritt wird die Trägermenge für die abstrakte (a) hierarchische rechtserweiterte Textverbundstruktur von WB (wb ∈ WB) — sie heiße $M_{\text{TVKonS}}^a(\text{WB})$ — sowie die gleichmächtige Trägermenge für die konkrete (k) hierarchische rechtserweiterte Textverbundkonstituentenstruktur — sie heiße $M_{\text{TVKonS}}^k(\text{wb})$ — gebildet. Die beiden Trägermengen mit der Mächtigkeit |15| können mit Hilfe der Klassensymbole bzw. Individuennamen wie folgt notiert werden: $M_{\text{TVKonS}}^a(\text{WB}) = \{\text{WB}, \text{ZSV}, \text{TIT}, \text{IMP}, \text{IV}, \text{EINF}, \text{BENH}, \text{ABK1}, \text{LSZ}, \text{ABK2}, \text{SAR}, \text{WZ}, \text{LQ}, \text{WÖHI}, \text{WBE}\}$.

$M_{\text{TVKonS}}^k(\text{wb}) = \{\text{wb}, a, b, c, d, e, f, g, h, i, j, k, l, m, n\}$.

Im *dritten* methodischen Schritt werden auf den beiden Trägermengen je zwei zweistellige strukturprägende Relationen vom gleichen Relationstyp definiert. Auf $M_{\text{TVKonS}}^a(\text{WB})$ wird *erstens* eine reflexive, antisymmetrische und transitive Relation vom Typ der partitiven Relation — sie heiße $R_{\text{part}}^{\text{WB}}$ — definiert, auf deren Elemente (also die geordneten Paare, wie z.B. $\langle \text{ZSV}, \text{WB} \rangle$), der Relationsterm $\text{RT}_1 = \text{ist eine Textverbundkonstituente von}$ zutrifft, *zweitens* wird auf $M_{\text{TVKonS}}^a(\text{WB})$ eine Relation vom Typ der Präzedenzrelation — sie heiße R_p^{WB} — definiert, die irreflexiv (damit asymmetrisch) und transitiv ist, und auf deren Elemente (also die geordneten Paare, wie z.B. $\langle \text{ZSV}, \text{TIT} \rangle$), der Relationsterm $\text{RT}_2 = \text{geht voraus}$ zutrifft. Für $R_{\text{part}}^{\text{WB}}$ und R_p^{WB} gelten folgende Aussagen: $R_{\text{part}}^{\text{WB}} \subseteq (M_{\text{TVKonS}}^a(\text{WB}))^2$ und $R_p^{\text{WB}} \subseteq (M_{\text{TVKonS}}^a(\text{WB}))^2$ (wobei für das einfache Kreuzprodukt einer Menge M mit sich selbst M^2 geschrieben ist). $R_{\text{part}}^{\text{WB}}$ und R_p^{WB} sind

disjunkt ($R_{part}^{WB} \cap R_p^{WB} = \emptyset$) und ihre Vereinigung $R_{part}^{WB} \cup R_p^{WB}$ ordnet $M_{TVKons}^a(WB)$ exhaustiv. Das Ergebnis der Definitionsoperationen ist die abstrakte hierarchische rechtserweiterte Textverbundkonstituentenstruktur von Paul 2002, die eine Ordnungsstruktur ist. Entsprechend wird verfahren, um die isomorphe konkrete Ordnungsstruktur von Paul 2002 zu erhalten. Sowohl für die konkrete als auch für die abstrakte hierarchische rechtserweiterte Textverbundstruktur kann ein Baumgraph mit einer Baumhöhe, die den Wert 1 hat, konstruiert werden. Bildet man die beiden Baumgraphen so aufeinander ab, dass von dem Baumgraph für die konkrete Struktur nur die terminalen etikettierten Knoten sichtbar und durch eine gestrichelte Linie mit den zugehörigen Klassensymbolen im Baumgraph für die abstrakte Struktur verbunden sind, ist der Strukturgraph für den Typ der hierarchischen rechtserweiterten Textverbundkonstituentenstruktur von Paul 2002 in abstrakter und isomorpher konkreter Ausprägung gegeben, der sich in Abb. 6 findet.

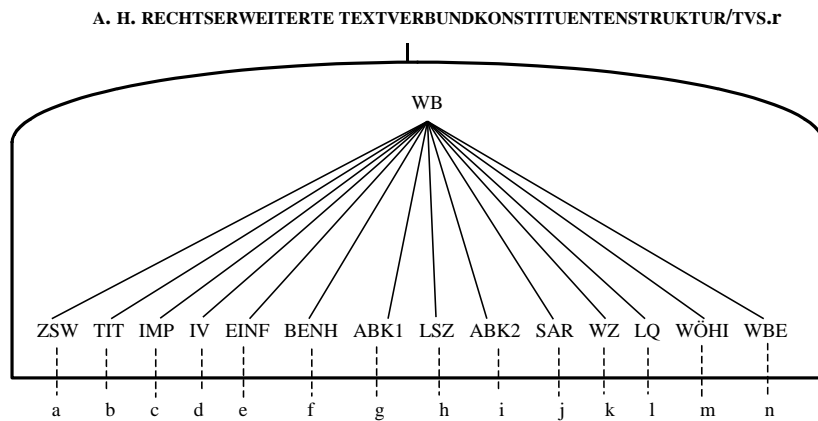


Abb. 6: Einfach kommentierter Strukturgraph zum Typ der hierarchischen abstrakten (und isomorphen konkreten) rechtserweiterten Textverbundkonstituentenstruktur als Teilstruktur der rechtserweiterten Textverbundstruktur am Beispiel von Paul 2002. *Abkürzungen:* (als Klassensymbole): ZSW = Zusammenspiel von Sachregister und Wörterbuch; TIT = Titelei; IMP = (erweitertes) Impressum; IV = Inhaltsverzeichnis; EINF = Einführung; BENH = Hinweise zur Benutzung; ABK1 = Abkürzungen; LSZ = Laut- und Sonderzeichen; ABK2 = Abgekürzt zitierte Autoren und Werke; SAR = Sachregister; WZ = Wörterverzeichnis; LQ = Literarische Quellen; WÖHI = Wörterbücher und weitere Hilfsmittel; WBE = Schlüssel zu den Wörterbucheinträgen; *andere Abkürzungen:* A. = ABSTRAKTE; H. = HIERARCHISCHE. *Darstellungskonventionen:* „x --- Y“ bedeutet soviel wie x ist eine Element von Y ($x \in Y$); /TVS.r ist zu lesen wie als Teilstruktur der rechtserweiterten Textverbundstruktur

An dieser Stelle sei kurz auf die Frage eingegangen, ob Wörterbücher, wenn sie als Textverbund und damit als eine Abfolge von Texten unterschiedlicher Textsorten betrachtet werden und wenn man ihre Textverbundgesamtstruktur konstruieren möchte, als zwei- oder als dreidimensionale Gebilde konzeptualisiert werden sollten. Bekanntlich ist in der Alltagssprache die Konzeptualisierung von Buchtexten schwankend, was sich buchgeschichtlich erklären lässt. So sagt man z.B. nicht *das Vorwort ist oben* oder *das Register ist unten*, aber in den Texten findet man *siehe oben* oder *siehe unten*, und diese Verweisungen sind nicht seitenspezifisch. Bei der strukturellen Analyse von Wörterbuchartikeln hat die explizite Berücksichtigung der textuellen Zweidimensionalität durch die Einführung von textarchitektonischen Relationen, wie z.B. der *oberhalb*- und der *unterhalb*-Relation und der Textarchitekturen die Strukturanalyse erheblich verfeinert und vertieft (vgl. z.B. Wiegand 2000a: 259ff.; 2001: 191ff; Bergenholtz, Tarp und Wiegand 1999: 1763ff.; Wiegand und Fuentes Morán 2009: 145ff.; Wiegand und Beer 2013: 269ff.). Auch bei der Analyse von Textverbundstrukturen wurden die vertikalen textarchitektonischen Relationen öfter berücksichtigt. Wenn man die zweidimensionale Konzeptualisierung für den gesamten Textverbund gelten lässt, dann gilt grundsätzlich: Alle Textverbundkonstituentenstrukturen und alle Wörterverzeichnisstrukturen lassen sich als vertikalarchitektonisch ausgebaut repräsentieren, so dass es nicht unbedingt notwendig wäre, eine solche Repräsentation in entsprechenden Strukturgraphen ausdrücklich vorzunehmen. Dennoch gilt eine Struktur nur dann als vertikalarchitektonisch ausgebaut, wenn auf ihrer Trägermenge mindestens eine vertikalarchitektonische Relation ausdrücklich definiert ist. In diesem Beitrag sei aus Platzgründen auf Textverbundarchitekturbilder ausdrücklich verzichtet (vgl. z.B. Kammerer und Wiegand 1998: 226; Wiegand und Beer 2013: 270, Abb. 8.17); nur ein allgemeines architektonisch angereichertes Textverbundkonstituentenstrukturmodell für den Typ der rechtserweiterten Textverbundkonstituentenstruktur sei in Abb. 7 präsentiert, weil es Gelegenheit gibt, auf einen weiteren Sachverhalt aufmerksam zu machen. Die in Abb. 7 dargestellte Struktur ist erhältlich, wenn man auf $M_{\text{TVKONS}}^a(\text{WB})$ nicht nur $R_{\text{part}}^{\text{WB}}$ und R_{p}^{WB} , sondern zusätzlich eine *oberhalb*-Relation $R_{\text{ob}}^{\text{WB}}$ definiert mit dem Relationsterm *x ist oberhalb von y*, mit *x* und *y* als Variablen für Klassen von Textverbundkonstituenten.

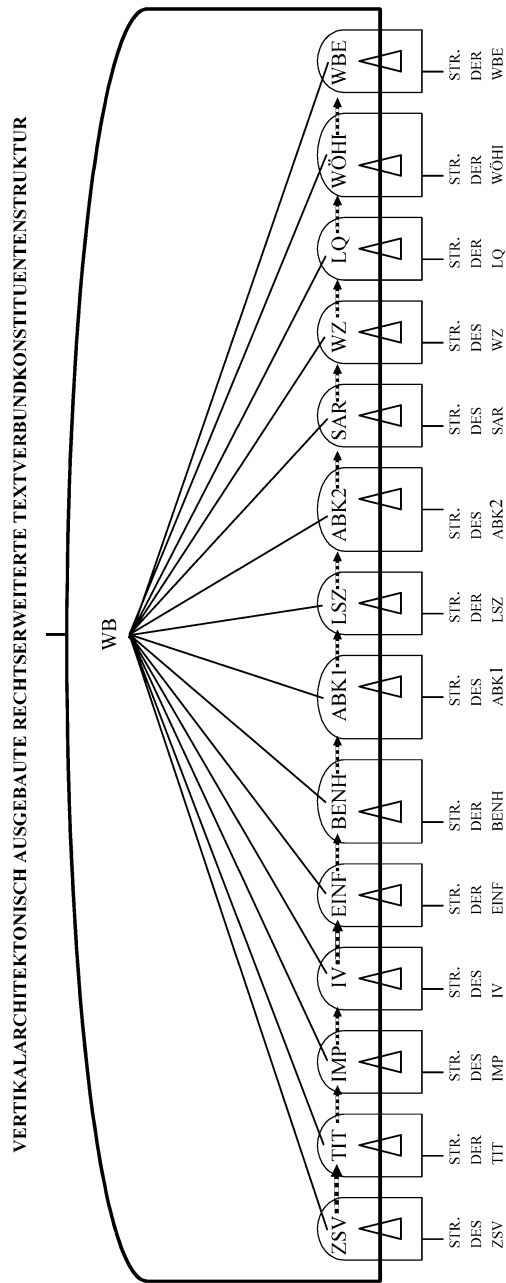


Abb. 7: Allgemeines architektonisch angereichertes Textverbundkonstituentenstrukturmodell für den Typ der rechnerweiterten Textverbundkonstituentenstruktur am Beispiel von Paul 2002; *Abkürzungen:* wie in Abb. 6 und STR. = STRUKTUR; *Darstellungskonventionen:* „x \rightarrow y“ bedeutet soviel wie x ist oberhalb von y.

Die Abb. 6 und 7 lassen deutlich erkennen, dass bei der Konstruktion einer hierarchischen Textverbundkonstituentenstruktur die hierarchischen Strukturen der einzelnen Textverbundkonstituenten keine Berücksichtigung finden. In Abb. 7 wird dieser Sachverhalt dadurch repräsentiert, dass der waagrechte Abschnitt des Strukturumrandungszeichens für die Textverbundkonstituentenstruktur die Umrandungszeichen für die hierarchischen Strukturen der 14 Textverbundkonstituenten schneidet, so dass erkennbar ist, dass nur die Textverbundkonstituenten, nicht aber deren hierarchischen Strukturen zu hierarchischen Textverbundkonstituentenstruktur gehören. Dies bedeutet zugleich, dass eine hierarchische Textverbundkonstituentenstruktur niemals hierarchische Teilstrukturen aufweist und stets eine flache hierarchische Struktur ist, zu der ein Baumgraph gehört, dessen Baumhöhe der Wert 1 zugeordnet ist und der demgemäß nur einen Wurzelknoten und terminale Knoten aufweist.

Damit taucht die Frage auf, ob nicht eine hierarchische Struktur vorgesehen ist, zu der alle hierarchischen Strukturen der einzelnen Textverbundkonstituenten als Teilstrukturen gehören. Die Antwort muss wie folgt lauten: In der Theorie der Wörterbuchform ist eine solche Struktur aus guten Gründen nicht vorgesehen. Denn *erstens* gilt: Für die Konstruktion einer hierarchischen Textverbundgesamtstruktur, die ja eine Übersicht über die relevanten Strukturen des Textverbundes ermöglichen soll, ist sie überflüssig, weil dazu nur die Struktur des Wörterverzeichnis benötigt wird, und weiterhin gilt *zweitens*: Die Analyse von Strukturen, wie z.B. die der Struktur der Benutzungshinweise oder die von Inhaltsverzeichnissen wird nicht dadurch bereichert, dass man diese Strukturen als Teilstrukturen einer übergeordneten hierarchischen Struktur betrachtet.

4.1.2 Vor- und Nachspanne, ihre Typen und Strukturen

Zu Beginn dieses Abschnittes sei an den Untertitel dieses Beitrags erinnert; es geht also nicht um Vor- und Nachspanntypen unter inhaltlichen, sondern ausschließlich um formstrukturelle Aspekte. Aus diesem Grund wird z.B. die Unterscheidung von integrierten und nichtintegrierten Vor- und Nachspanntexten nicht diskutiert. Obwohl sich die neuere Forschung seit drei Jahrzehnten und besonders in den letzten Jahren relativ ausführlich (vgl. Gouws 2002 u. 2004 u. besonders Bielińska 2010) mit dem wörterverzeichnisexternen Wörterbuchteilen beschäftigt hat — einer der Ersten, der von Wörterbuchvor- und Wörterbuchnachspann spricht, ist Herberg (1985: 134) — wurden eine Reihe von wörterbuchformbezogenen Problemen und damit Gründe für weitergehende Differenzierungen nicht wahrgenommen. Im Folgenden wird versucht, diese Probleme — soweit sie Vor- und Nachspanne betreffen — dadurch zu separieren und damit bearbeitungsfähig zu machen, dass zwei Fragen gestellt werden.

In der Theorie der Wörterbuchform gilt der Grundsatz: Als theorierelevant gelten nur solche textuellen Einheiten, die aufgrund der Anwendung

einer bekannten und ausreichend erprobten metalexikographischen Methode erhältlich sind. Die *erste Frage* lautet daher:

(1) Welche Methoden wurden angewandt um Vor- und Nachspanne zu ermitteln? Üblicherweise wurde einfach in alltagssprachlicher Manier festgestellt, dass alle Wörterbuchteile, die vor dem Wörterverzeichnis stehen, zum Vorspann, und — falls vorhanden — alle, die nach dem Wörterverzeichnis stehen, zum Nachspann gehören. Zunächst muss festgestellt werden, dass eine solche Antwort unvollständig ist, und zwar aus zwei Gründen, nämlich (a), weil ein Vorspann auch vor und ein Nachspann auch nach einer Wörterverzeichnisreihe stehen kann (vgl. z.B. Textverbundtyp 3 und 4 in Abb. 2) und (b), weil ein Vorspann, der dann *vorderer Vorspann* heißt, vor dem Vorspann einer eingebetteten Textverbundreihe stehen kann, und ein Nachspann, der dann *hinterer Nachspann* heißt, auf den Nachspann einer eingebetteten Textverbundreihe folgen kann (vgl. z.B. Textverbundtyp 5 und 6 in Abb. 2). Wörterbuchformbezogen lassen sich zwei weitere Typen von Vor- und Nachspannen unterscheiden, und zwar wenn berücksichtigt wird, ob die Vorsatzblätter textuell in den Vor- oder Nachspann integriert sind oder nicht. Man kann dann von Vor- oder Nachspannen mit integriertem oder nicht integriertem Vorsatzblatt sprechen. Die beiden letzteren Unterscheidungen gelten jedoch nicht als strukturelevant.

Die Methode, die angewandt werden muss, um Vor- und Nachspanne zu erhalten, ist eine positionsgesteuerte Gruppierungsmethode, die ausschließlich auf Textverbundkonstituenten, die vorher durch eine Anwendung der Methode der funktional-positionalen Segmentation gegeben sind, angewandt wird. Dies bedeutet, dass der Vor- und der Nachspann positionsspezifische Gruppierungen von Textverbundkonstituenten sind und nicht das Ergebnis einer funktional-positionalen Segmentation. Sie haben daher auch als Vor- und Nachspann, wenn sie als je eine zusammengehörige Gruppierung betrachtet werden, keine genuine Funktion für den gesamten Textverbund; sie sind damit keine Textverbundkonstituenten, weil sie nicht alle drei Definitionskriterien für eine lexikographische Konstituente aufweisen, sondern nur zwei, nämlich eine bestimmte Form und genau eine Position. Deswegen ist die Abb. 1 so fatal irreführend, weil sie den falschen Eindruck erweckt, als sei der Baumgraph das Ergebnis einer funktional-positionalen Segmentation in zwei Segmentationschritten und damit seine etikettierten Knoten Repräsentationen lexikographischer Konstituenten. Vorspanne gehören aber zu den Textsektoren und Nachspanne, wenn sie nicht in bestimmter Weise einteilig sind (und z.B. nur aus den Benutzungshinweisen bestehen), ebenfalls; sie sind keine lexikographischen Konstituenten. Vielmehr sind Textsektoren artikelexterne textuelle Teil eines Wörterbuchs, die durch die Anwendung von Gruppierungsmethoden gegeben sind. Sie sind selbst nicht als Ergebnis einer funktional-positionalen Segmentation erhältlich und — falls sie nicht einteilig sind — selbst nur nichtfunktional-positional segmentierbar; dabei ist es wichtig zu beachten, dass

die entscheidende Korrektheitsbedingung für die Anwendung der Methode der nichtfunktional-positionalen Segmentation lautet: Unter den erhältlichen Textsegmenten muss mindestens ein nichtfunktionales Textsegment sein, so dass das Auftreten von funktionalen Textsegmenten ebenfalls gegeben ist (vgl. z.B. Wiegand 2010: 267). Die nichtfunktionalen Textsegmente sind dann Subsektoren. Im Falle der beiden Textsektoren Vorspann und Nachspann heißen sie *Vorspann-* und *Nachspannsektoren*. Die funktionalen Textsegmente, die als Ergebnis einer nichtfunktional-positionalen Segmentation erhältlich sind, sind hybride lexikographische Konstituenten; als Teile eines Vorspanns bzw. eines Nachspanns heißen sie *hybride Vor-* oder *Nachspannkonstituenten*. Sie haben entweder eine genuine Funktion für einen Vor- oder Nachspannsektor desselben Vor- oder Nachspanns, zu dem sie gehören, oder für den gesamten Vor- oder Nachspann. Der Terminus *Vorspannsektor* kann als Hyperonym für *Vorspanntext*, *Vorspanntabelle*, *Vorspannliste*, *Vorspannverzeichnis* und *Vorspanntafel* verwendet werden. Entsprechendes gilt für *Nachspannsektor*. Dieser terminologische Hinweis kommt denen entgegen, die Tabellen, Listen, Verzeichnisse und Tafeln nicht als Texte gelten lassen möchten; seine Befolgung hat aber enge Grenzen, da man z.B. zwar von Umtexten, aber nicht von Umlisten etc. und auch nicht von Außentabellen, wohl aber von Außentexten sprechen kann etc.! Es ist daher wohl effektiver alle auftretenden Textsektoren in Vor- und Nachspannen als lexikographische Texte aufzufassen, gleichgültig, ob dies Textlinguisten passt oder nicht.

Die *zweite Frage* lautet: Gibt es Vor- oder Nachspanne, die nur aus einer Textverbundkonstituente bestehen, so dass sich ihre genuine Funktion notwendigerweise auf den gesamten Textverbund bezieht? Für die Vorspanne ist das Folgende festzustellen: Wenn man das Impressum sowie den so genannten Schmutztitel als Teile der Titelei versteht, wie das in der Buchwissenschaft üblich ist, dann muss die Antwort lauten: Ja. In der Metalexikographie gibt es jedoch keinen zwingenden Grund, das buchwissenschaftliche Verständnis vollständig zu übernehmen; vielmehr ist es hier üblich, das Impressum als eigene Textverbundkonstituente zu verstehen. Bei dieser Voraussetzung muss die Antwort lauten: Einteilige Vorspanne gibt es nicht, da jeder Vorspann mindestens eine Titelei und ein Impressum, das meistens auf der Titelfrückseite steht, aufweist und nur in sehr wenigen älteren Wörterbüchern am Fuß der Titelseite zu finden ist. Daher gibt es keine Vorspanne, die mit einer Textverbundkonstituente formidentisch sind. Ein Nachspann kann dagegen einteilig und damit mit einer Textverbundkonstituente form- und funktionsidentisch sein. Dies ist z.B. in Hessky und Iker 2005 und in Krämer und Kaehlbrandt 2011 der Fall.

Im Folgenden werden die abstrakte hierarchische konstituentenlose Vor- und Nachspannstruktur von Paul 2002 betrachtet. Die Strukturbildungsprozesse werden nur verkürzt dargestellt. Die Trägermenge für die abstrakte (a) Vorspannstruktur — sie heiße $M_{VS}^a(WB) - (M_{VS}^a(WB) \subseteq (M_{TVK_{onS}}^a(WB)))$ — kann wie folgt notiert werden:

$M_{VS}^a(WB) = \{ZSW, TIT, IMP, IV, EINF, BENH, ABK1, LSZ, ABK2, SAR\}$.

Auf $M_{VS}^a(WB)$ werden zwei strukturprägende Relationen definiert, und zwar R_{part}^{VS} mit dem Relationsterm $RT_3 = x$ ist ein Vorspannsektor von y (mit x als Variable für Vorspannsektoren und y als Variable für den Vorspann) sowie R_p^{VS} mit dem Relationsterm $RT_4 = x$ geht y voraus (mit x, y als Variablen für Textverbundkonstituenten, (die auch entweder Vorspannsektoren oder hybride Vorspannkonstituenten sind)). Für R_{part}^{VS} und R_p^{VS} gelten folgende Aussagen:

$R_{part}^{VS} \subseteq (M_{VS}^a(WB))^2$ und $R_p^{VS} \subseteq (M_{VS}^a(WB))^2$ sowie $R_{part}^{VS} \cap R_p^{VS} = \emptyset$; die Vereinigung $R_{part}^{VS} \cup R_p^{VS}$, ordnet $M_{VS}^a(WB)$ exhaustiv. Das Ergebnis der Definitionsoperationen ist die abstrakte hierarchische Vorspannstruktur, die in Abb. 8 dargestellt ist.

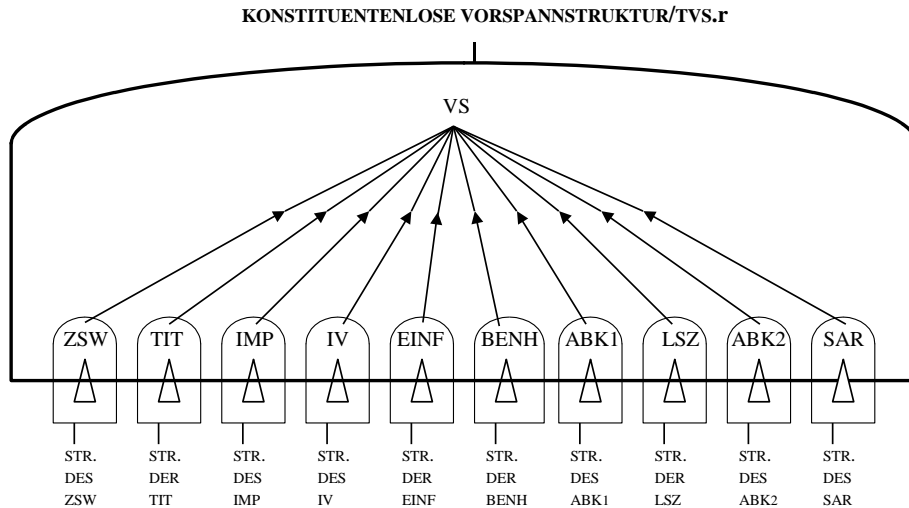


Abb. 8: Allgemeines Vorspannstrukturbild zum Typ der hierarchischen konstituentenlosen Vorspannstruktur als Teilstruktur der rechts-erweiterten Textverbundstruktur am Beispiel von Paul 2002; *Abkürzungen:* wie in Abb. 7 und VS = Vorspann; *Darstellungskonventionen:* „ $x \longrightarrow y$ “ bedeutet so viel wie x ist ein Textsektor von y ; /TVS.r ist zu lesen wie als Teilstruktur der rechtserweiterten Textverbundstruktur.

Die abstrakte Nachspannstruktur von Paul 2002 ist ebenfalls konstituentenlos. Sie ist auf dem gleichen methodischen Weg erhältlich wie die abstrakte hierarchische Vorspannstruktur; ihr allgemeines Strukturbild findet sich in Abb. 9.

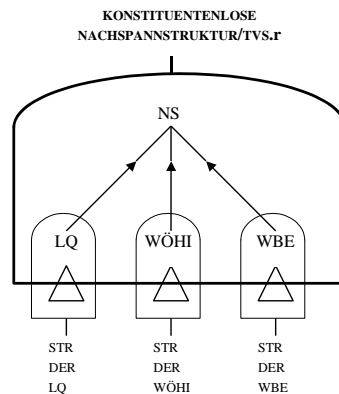


Abb. 9: Allgemeines Nachspannstrukturbild zum Typ der hierarchischen konstituentenlosen Nachspannstruktur am Beispiel von Paul 2002; *Abkürzungen:* wie in Abb. 7; *Darstellungskonventionen* wie in Abb. 8

Paul 2002 gehört zu den rechtserweiterten Textverbunden, zu deren Vor- und Nachspann konstituentenlose hierarchische Vor- und Nachspannstrukturen gehören (vgl. Abb. 8 u. 9); dies gilt z.B. auch für Wahrig 2006, Brenner 1951, LDW 1955 und COED 2009. Wörterbücher, deren Vorspann und/oder Nachspann sowohl Vorspannsektoren als auch hybride Vorspannkonstituenten bzw. Nachspannsektoren und hybride Nachspannkonstituenten aufweisen, sind seltener. Beispiele sind PONS-DaF 2016 und CALD 2008, die diesbezüglich nun ausschnittsweise analysiert werden.

Zunächst wird der Vorspann von PONS-DaF 2016 betrachtet. Es handelt sich um einen Vorspann mit integriertem Vorsatzblatt. Er besteht aus den Textverbundkonstituenten (1) bis (9), die nachfolgend genannt und mit Hilfe von „∈“ und eines Klassensymbols einer Klasse von Textverbundkonstituenten zugeordnet werden.

- (1) Inhalt und Aufbau der Einträge; (1) ∈ IAE
- (2) Das intelligente Online-Wörterbuch [...]; (2) ∈ ONWB
- (3) Titelei; (3) ∈ TIT
- (4) Impressum; (4) ∈ IMP
- (5) Inhaltsverzeichnis; (5) ∈ IV
- (6) Vorwort; (6) ∈ VW
- (7) Hinweise für die Benutzung; (7) ∈ BENH
- (8) Inhaltsverzeichnis zum Regelteil zur Rechtschreibung; (8) ∈ IVRR
- (9) Regelteil zur Rechtschreibung; (9) ∈ RR

Die Textverbundkonstituenten (1) bis (7) und (9) sind Vorspannsektoren; (8) ist eine hybride Vorspannkonstituente, deren Titel mit ihrer ersten Seitenzahl 25 in (5), dem Inhaltsverzeichnis zum Textverbund, genannt wird, so dass die

erwähnte Seitenzahl für die erste Seite von (9), nämlich 25, ein Element der vorspanninternen numerischen äußeren Zugriffsstruktur zum gesamten Textverbund ist, die als senkrecht verlaufender Teil des Inhaltsverzeichnisses (5) gegeben ist, was den äußeren Zugriff auf (9), den Regelteil, sehr erleichtert. (8) gehört zu den hybriden Vorspannkonstituenten, deren genuine Funktion sich nicht auf den gesamten Vorspann, sondern nur auf einen Vorspannsektor, nämlich auf (9), bezieht.

Da nicht alle vorspannzugehörigen Textverbundkonstituenten des Vorspanns von PONS-DaF 2016 Textsektoren sind, gehört die zugehörige Vorspannstruktur zum Typ der partiell konstituentenlosen Vorspannstruktur. Deren Bildungsprozess wird hier nicht beschrieben; ein relevanter Unterschied zu dem skizzierten Strukturbildungsprozess einer konstituentenlosen Vorspannstruktur besteht darin, dass auf der Trägermenge einer partiell konstituentenlosen Vorspannstruktur zwei unterschiedliche partitive Relationen definiert werden, und zwar eine mit dem Relationsterm $RT_3 = x \text{ ist ein Vorspannsektor von } y$ (mit x als Variable für Vorspannsektoren und y als Variable für Vorspanne) sowie eine Relation mit dem Relationsterm $RT_4 = x \text{ ist eine hybride Vorspannkonstituente von } y$ (mit x als Variable für Vorspannkonstituenten und y als Variable für Vorspanne). Die abstrakte hierarchische partiell konstituentenlose Vorspannstruktur, die zum Vorspann (mit integrierten Vorsatzblatt) von PONS-DaF 2016 gehört, findet sich in Abb. 10.

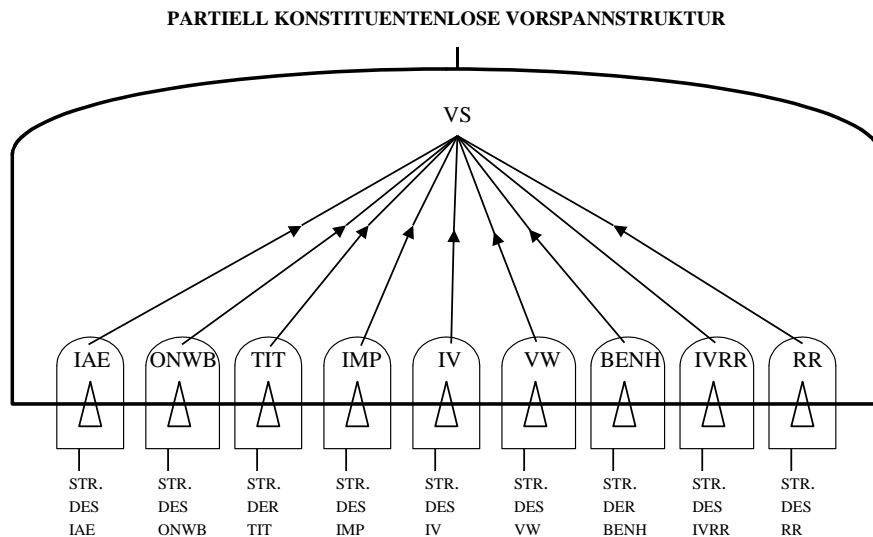


Abb. 10: Allgemeines Vorspannstrukturbild zum Typ der hierarchischen partiell konstituentenlosen Vorspannstruktur (mit integriertem Vorsatzblatt) am Beispiel von PONS-DaF 2016; *Abkürzungen:* IAE = Inhalt und Aufbau der Einträge; ONWB = Online-Wörterbuch; TIT =

Titel; IMP = Impressum; IV = Inhaltsverzeichnis; VW = Vorwort; BENH = Benutzungshinweise; IVRR = Inhaltsverzeichnis zum Regelteil zur Rechtschreibung; RR = Regelteil zur Rechtschreibung; *Darstellungskonventionen*: „ $x \rightarrow y$ “ bedeutet soviel wie x ist ein Vorspannsektor von y ; „ $x \dashrightarrow y$ “ bedeutet soviel wie x ist eine hybride Vorspannkonstituente von y

Im Folgenden wird der Nachspann (mit nichtintegriertem Vorsatzblatt) von CALD 2008 ausschnittsweise betrachtet. Er weist ein Nachspanninhaltsverzeichnis auf, in dem sich eine senkrecht verlaufende alphanumerische Nachspannzugriffsstruktur mit 16 alphanumerischen Zugriffstextelementen findet. Das Nachspanninhaltsverzeichnis (IV) gehört zu den hybriden Nachspannkonstituenten, deren genuine Funktion sich auf den gesamten Nachspann bezieht. Auf die Nachspannkonstituente folgen 15 Nachspannsektoren. Ein nicht vollständig ausgeführtes allgemeines Nachspannstrukturmodell findet sich in Abb. 11.

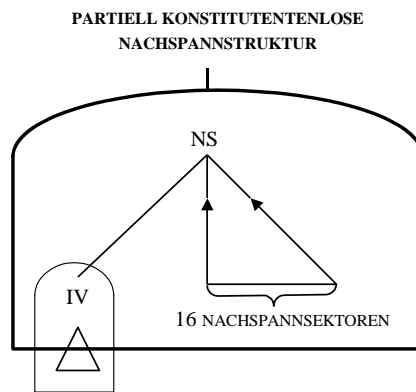


Abb. 11: Nicht vollständig ausgeführtes allgemeines Nachspannstrukturmodell zum Typ der hierarchischen partiell konstituentenlosen Nachspannstruktur (ohne integriertes Vorsatzblatt) am Beispiel von CALD 2008; *Abkürzungen*: wie in Abb. 9 und NS = Nachspann; *Darstellungskonventionen*: wie in Abb. 9

4.1.3 Hierarchische rechtserweiterte Textverbundstrukturen

In diesem Abschnitt wird auf dem Weg zur Konstruktion der abstrakten hierarchischen Textverbundgesamtstruktur von Paul 2002 als deren Teilstruktur die abstrakte hierarchische rechtserweiterte Textverbundstruktur dadurch konstruiert, dass sie mit bereits bekannten abstrakten hierarchischen Teilstrukturen der Letzteren (vgl. Abb. 6, 8 und 9) zusammengefügt werden. Dies geschieht zunächst auf der Ebene der Strukturdarstellung. Das Ergebnis findet sich in Abb. 12.

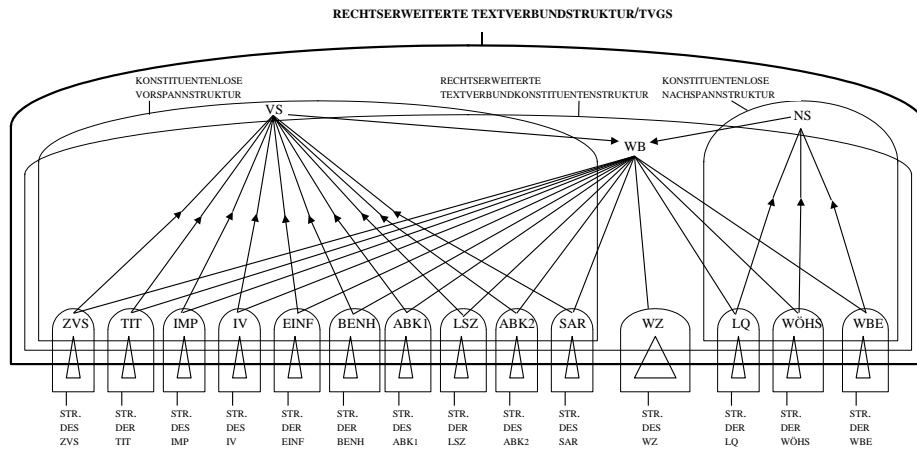


Abb. 12: Allgemeines Textverbundstrukturbild zum Typ der hierarchischen rechteckigen Textverbundstruktur als Teilstruktur der hierarchischen Textverbundgesamtstruktur am Beispiel von Paul 2002; *Abkürzungen:* wie in Abb. 6, 8 u. 9. *Darstellungskonventionen:* „x—y“ bedeutet soviel wie x ist eine Textverbundkonstituente von y; „x → y“ bedeutet soviel wie x ist ein nichtfunktionaler Teil von y. /TVGS ist zu lesen wie als Teilstruktur der Textverbundgesamtstruktur

Es sei darauf aufmerksam gemacht, dass der Strukturgraph in Abb. 12 kein Baumgraph ist, weil alle Textverbundkonstituenten außer WZ nicht gerade einen Vorgänger haben sondern zwei. Auch die in Abb. 12 dargestellte Struktur ist als Ergebnis eines regulären Strukturbildungsprozesses erhältlich. Man muss dann zur Bildung der abstrakten Struktur lediglich die Trägermenge $M_{TVKons}^a(WB)$ um VS und NS erweitern, so dass die erweiterte (er) Trägermenge $erM_{TVKons}^a(WB)$ mit einer Mächtigkeit von |17| gegeben ist und auf dieser eine Präzedenzrelation definieren sowie eine partitive Relation mit dem Relationsterm $RT_1 = x \text{ ist eine Textverbundkonstituente von } y$ und eine weitere partitive Relation mit dem Relationsterm $RT_5 = x \text{ ist ein nichtfunktionaler Teil von } y$.

Je nach dem Ziel der Untersuchung kann man bei der Bildung einer Textverbundstruktur auch die zweigeteilte textuelle Rahmenstruktur berücksichtigen sowie Vollständigkeitsgrade einführen (vgl. Kammerer und Wiegand 1998: 232ff.). Dies ist im hiesigen Kontext jedoch nicht erforderlich, in dem es vor allem darauf ankommt, eine übersichtliche Textverbundgesamtstruktur zu erhalten und die Grundregeln für deren Konstruktion darzustellen.

4.1.4 Hierarchische Wörterverzeichnisstrukturen

Für die Textverbundkonstituente Wörterverzeichnis (= WZ) in Abb. 12 gilt, im Unterschied zu allen anderen: Im Rahmen der Konstruktion einer hierarchischen

Textverbundgesamtstruktur wird WZ in einem weiteren Segmentationsritt nichtexhaustiv funktional-positional segmentiert. Mit der Wahl dieser Segmentationsmethodenvariante (i.S.v. Wiegand 2005: 217f. u. 2010: 267f.) ist zugleich festgelegt, dass die nichttypographischen Strukturanzeiger, die unmittelbare Textkonstituenten des Wörterverzeichnisses sind, wie z.B. die äußeren Zugriffsstrukturanzeiger **A, B, ..., Z** oberhalb der Artikelstrecken — im Unterschied zu Wiegand (2001: 137, Abb. 4-3) — nicht berücksichtigt werden, sondern nur die Artikelstrecken. Diese sind unmittelbare Textkonstituenten von Wörterverzeichnissen in initialalphabetischen Wörterbüchern. Ihre Anzahl wird durch das Zugriffsalphabet (i.S.v. Wiegand 2005a: 208ff.) bestimmt. In Paul 2002 sind es 26 Artikelstrecken. Die abstrakte hierarchische Wörterverzeichnisstruktur von Paul 2002 kann demgemäß nach dem Durchlaufen eines regulären Strukturbildungsprozesses wie in Abb. 13 dargestellt werden.

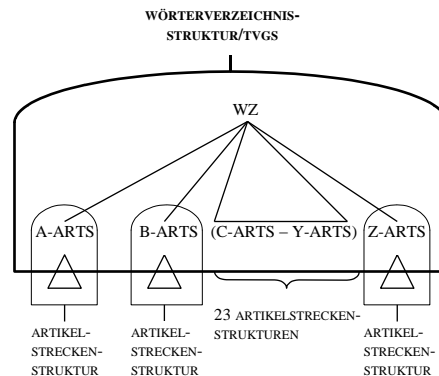


Abb. 13: Nicht vollständig ausgeführtes allgemeines Wörterverzeichnisstrukturbild zum Typ der hierarchischen Wörterverzeichnisstruktur als Teilstruktur der Textverbundgesamtstruktur am Beispiel von Paul 2002; *Abkürzungen:* ARTS = Artikelstrecke; *Darstellungskonventionen:* /TVGS ist zu lesen wie als Teilstruktur der Textverbundgesamtstruktur.

4.1.5 Hierarchische Artikelstreckenstrukturen

Bei den Artikelstrecken lassen sich zahlreiche Typen unterscheiden. Im Paul 2002 liegen einfache artikelhomogene nestalphabetische Artikelstrecken vor. Die Artikelstrecken sind einfach, weil es keine eingelagerten Binnentexte gibt; sie sind artikelhomogen (kurz: homogen), weil nur einfache und Verweisartikel als unmittelbare Artikelstreckenkonstituenten auftreten und keine Synopseartikel, und sie sind schließlich nestalphabetisch, weil es Lemmata gibt, die nicht alphabetisch einsortiert sind, so dass Artikelnester und Wörterbuchartikel in sehr zahlreichen unterschiedlichen Reihenfolgen aufeinander folgen können, was zur Folge hat, dass die hierarchische Artikelstreckenstruktur mit einem

vollständigen Baumgraph nicht übersichtlich dargestellt werden kann. Demzufolge hat das allgemeine Artikelstreckenstrukturbild die Form in Abb. 14.

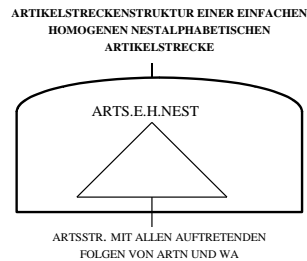


Abb. 14: Nicht vollständig ausgeführtes allgemeines Artikelstreckenstrukturbild zum Typ der hierarchischen einfachen homogenen nestalphabetischen Artikelstreckenstruktur am Beispiel von Paul 2002; *Abkürzungen:* ARTS.E.H. NEST = einfache homogene nestalphabetische Artikelstrecke; ARTN = Artikelnest; WA = Wörterbuchartikel

Bei Wörterbüchern mit einfachen homogenen glattalphabetischen Artikelstrecken, wie z.B. Agricola 1992, hat das allgemeine Artikelstreckenstrukturbild die Form in Abb. 15.

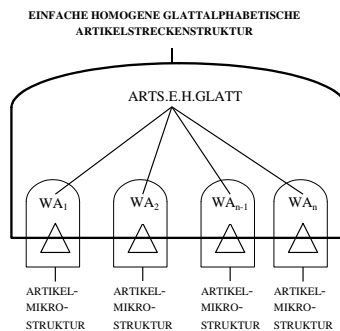


Abb. 15: Allgemeines Artikelstrukturbild zum Typ der hierarchischen einfachen homogenen glattalphabetischen Artikelstreckenstruktur am Beispiel von Agricola 1992; *Abkürzungen:* ARTS.E.H.GLATT = einfache homogene glattalphabetische Artikelstrecke; *Leseanweisungen* für die unteren Indices am Klassensymbol WA, nämlich „1“, „2“, „n-1“ und n: die Indices zählen nicht etwa Unterklassen von WA oder verschiedene WA-Klassen, denn es gibt stets nur gerade eine Klasse der Wörterbuchartikel (bezogen auf eine Grundmenge). Vielmehr ist zu lesen: Wenn M eine Menge von n Wörterbuchartikeln (WA) ist, die unmittelbare Artikelstreckenkonstituenten ein und derselben Artikelstrecke sind, und wenn M eine Teilmenge von WA ist, dann gilt, dass auf M eine lineare Struktur (i.S.v. einer totalen

Ordnung) definiert ist, die zu einer endlichen Reihe natürlicher Zahlen der Form $1, 2, \dots, n-1, n$ isomorph ist. Diese Isomorphie zeigt sich in den Indices von WA.

Bei Abb. 15 ist zu beachten, dass auf der Menge aller Wörterbuchartikel einer glattalphabetischen Artikelstrecke stets eine *oberhalb*-Relation definierbar ist. Wie z.B. auch im Fall der Textverbundkonstituentenstruktur in Abb. 6, ist die Definition hier nicht ausdrücklich vorgenommen, so dass die textarchitektonische Relation in Abb. 15 nicht repräsentiert ist.

4.1.6 Die hierarchische Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur von Paul 2002

Im Folgenden wird der Strukturbildungsprozess für die hierarchische Textverbundgesamtstruktur von Paul 2002 dadurch abgeschlossen, dass die in Abb. 12, 13 und 14 präsentierten Teilstrukturen zusammengefügt werden. Der Prozess der Zusammenfügung wird lediglich auf der Darstellungsebene dadurch vorgenommen, dass sein Ergebnis, die abstrakte hierarchische Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur, in Form eines allgemeinen Textverbundgesamtstrukturbildes in Abb. 16 präsentiert wird.

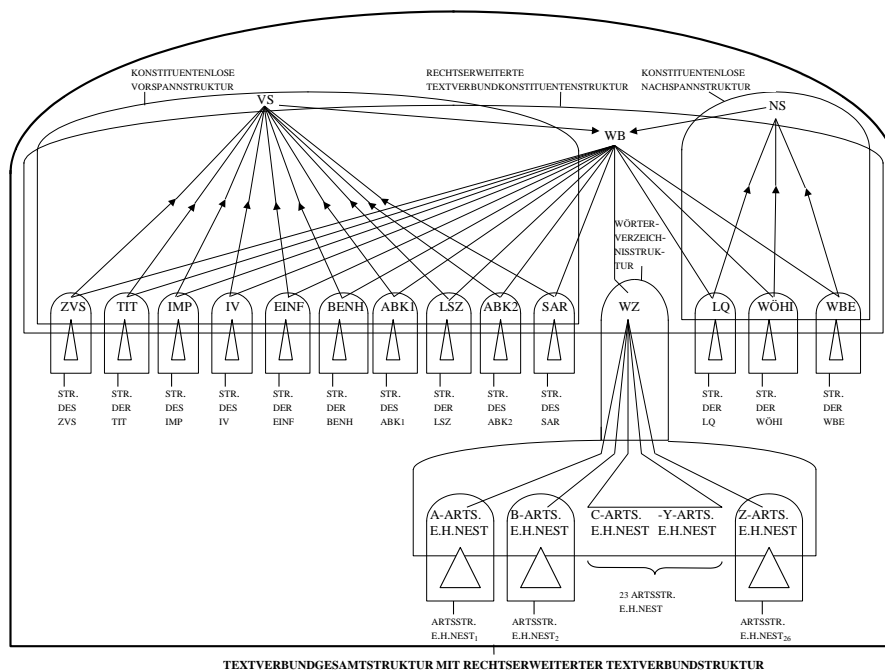


Abb. 16: Allgemeines Textverbundgesamtstrukturbild zum Typ der hierarchischen Textverbundgesamtstruktur mit rechtserweiterter Textver-

bundstruktur am Beispiel von Paul 2002; *Abkürzungen*: wie in Abb. 12, 13 und 14 und A-ART.S.E.H.NEST = einfache homogene nest-alphabetische A-Artikelstrecke; ARTSSTR = Artikelstreckenstruktur; ARTSSTR.E.H.NEST = Artikelstreckenstruktur zu einer einfachen homogenen Artikelstrecke

5. Zur Berücksichtigung von äußeren Zugriffsstrukturen bei der Konstruktion von hierarchischen Textverbundgesamtstrukturen

Es steht wohl außer Frage, dass das Vorhandensein von äußeren Zugriffsstrukturen (vgl. zu diesen z.B. Wiegand 2005a, Wiegand und Beer 2013a) eines der wichtigsten Definitionskriterien darstellt, wenn eine Definition von Printwörterbüchern oder von Gebrauchsgegenständen mit lexikographischen Formeigenschaften formuliert werden soll. Ein Nachdenken darüber, ob bei der Konstruktion von Textverbundgesamtstrukturen auch äußere Zugriffsstrukturen Berücksichtigung finden sollten, ist daher erforderlich. Dabei kann an einen liegen gebliebenen ersten Versuch in Wiegand (2001: 137) angeknüpft werden, in dem Artikelstreckenzugriffsstrukturen als Teilstrukturen von äußeren Hauptzugriffsstrukturen berücksichtigt wurden.

Mit dem Gebrauch eines Terminus, in dem *Textverbundgesamtstruktur* als Komponente auftritt, wie z.B. im Terminus *Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur*, werden m.E. berechnete dahingehende Erwartungen erzeugt, die es notwendig erscheinen lassen, dass äußere Zugriffsstrukturtypen bei der Konstruktion von Textverbundgesamtstrukturen zu berücksichtigen sind. Es entsteht daher die Frage, wie dies in angemessener und möglichst effektiver Weise geschehen kann. Eine Voraussetzung ist, dass man über eine hinreichend exhaustive Typologie von äußeren Zugriffsstrukturen mit einer zugehörigen Terminologie verfügt. Diese Voraussetzung ist mit Wiegand 2008 erfüllt (vgl. auch Wiegand und Beer 2013a). Nicht zu empfehlen ist — es sei denn, es liegen spezifische didaktische Zwecke vor — eine Kombination von Strukturgraphen mit Veranschaulichungen (i.S.v. Wiegand 2010: 318ff.) für die Darstellung zu verwenden. Zugriffsstrukturen in Printwörterbüchern sind nichthierarchische Strukturen. Für ihre (auch in formaler Hinsicht) einwandfreie Berücksichtigung gibt es m.E. zwei Möglichkeiten:

(1) man spezifiziert geeignete Knotenetikette. Für kundige Leser reicht z.B. die Etikettierung der terminalen Knoten in Abb. 16 bereits aus, denn er kann aus einem Knotenetikett, wie z.B. „A-ARTS.E.H.NEST (für: einfache homogene nestalphabetische A-Artikelstrecke) schließen, dass die Hauptzugriffsstruktur von Paul 2002 nestalphabetisch ist. Um diesen Sachverhalt zu verdeutlichen, kann man das Knotenetikett „WZ“ auch zu „WZ.NEST“ spezifizieren und demgemäß von einem *nestalphabetischen Wörterverzeichnis* sprechen (aber **nicht** von einer *nestalphabetischen Wörterverzeichnisstruktur*!). Wenn man Textverbund-

gesamtstrukturen auch terminologisch weiter unterscheiden möchte, kann man den Zugriffstyp der Hauptzugriffstruktur berücksichtigen. Paul 2002 wäre dann ein Textverbund mit folgendem Typennamen: *Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur und nestalphabetischer Hauptzugriffstruktur*.

Auch die peripheren äußeren Zugriffstrukturen können bei der Konstruktion von Textverbundgesamtstrukturen durch die Spezifizierung von Etiketten für die Textverbundkonstituenten berücksichtigt werden. Für Paul 2002 hätte man dann folgende fünf Etikette für Textverbundkonstituenten mit peripheren Zugriffstrukturen: ABK1.sia, ABK2.sia, SAR.sia, LQ.sia und WÖHL.sia (mit sia = striktinitialalphabetisch). Man kann dann von *Textverbundgesamtstruktur mit rechtserweiterter Textverbundstruktur mit vollständigem äußeren Zugriffsprofil* sprechen. Äußere Zugriffsprofile bilden den zentralen Teil von Datenakzessivitätsprofilen (i.S.v. Wiegand 2008: 301ff.).

(2) Wenn das Wörterverzeichnis eines Wörterbuchs glattalphabetisch ist, kann man die Hauptzugriffstruktur auch auf andere Weise bei der Konstruktion der hierarchischen Textverbundgesamtstruktur berücksichtigen, was im Folgenden am Beispiel von Köster 1969 in stark geraffter Form demonstriert werden soll. Köster 1969 (= wb \in WB) ist ein einfacher Textverbund; demgemäß gehört zu ihm eine einfache Textverbundstruktur (vgl. Abb. 2). Die nichtexhaustive funktional-positionale Segmentation dieses einfachen Textverbundes erbringt vier Textverbundkonstituenten; diese werden nachfolgend genannt, und die Klassensymbole der zugehörigen Textverbundklassen werden ihnen in runden Klammern zugeordnet.

Titelei (TIT)

Impressum (IMP)

Hinweise zur Benutzung (BENH)

Glattalphabetisches Wörterverzeichnis (WZ.GLATT).

TIT, IMP und BENH werden zum Vorspann VS gruppiert. Dann wird WZ.GLATT nichtexhaustiv funktional-positionale segmentiert, so dass 26 Artikelstrecken gegeben sind, die zum Typ der einfachen homogenen glattalphabetischen Artikelstrecken gehören; sie werden der Klasse ARTS.E.H.GLATT zugeordnet. Schließlich werden alle Artikelstrecken nichtexhaustiv funktional-positionale segmentiert, so dass alle Wörterbuchartikel von Köster 1969 gegeben sind. Nachdem alle genannten Segmentationschritte ausgeführt sind, schließen sich die weiteren regulären Schritte einer Strukturbildung an. Das Ergebnis ist die Textverbundgesamtstruktur mit einfacher Textverbundstruktur und glattalphabetischer Hauptzugriffstruktur, die in Abb. 17 dargestellt ist.

turen in der Theorie der Wörterbuchform in konkreter und abstrakter Ausprägung durch die Anwendung theoriezugehöriger Methoden erhältlich ist, ist eine textuelle Struktur eines als Textverbund aufgefassten initialalphabetischen Printwörterbuchs, für die *erstens* gilt: sie weist obligatorisch folgende Teilstrukturen auf: eine Textverbundstruktur mit festlegbaren Vollständigkeitsgrad, bestehend aus einer Textverbundkonstituentenstruktur und einer Vorspannstruktur, eine Wörterverzeichnisstruktur sowie mehrerer Artikelstreckestrukturen, deren Anzahl durch das Zugriffsalphabet bestimmt wird, und für die *zweiten* gilt: sie weist fakultativ mindestens eine Hauptzugriffsstruktur auf.

Ergänzend zu dieser Antwort muss festgestellt werden: Sie bezieht sich auf den basalen Typ von Textverbundgesamtstrukturen, den Typ der einfachen Textverbundgesamtstruktur (vgl. Typ 1 in Abb. 2). Für jeden anderen Typ muss die Antwort typspezifisch erweitert werden.

Literatur

- Agricola, Erhard.** 1992. *Wörter und Wendungen. Wörterbuch zum deutschen Sprachgebrauch.* Überarb. Neufassung der 14. Aufl. unter Mitwirkung von Herbert Görner und Ruth Kufner. Mannheim [...]: Dudenverlag.
- Bergenholtz, Henning, Sven Tarp und Herbert Ernst Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, Lothar, Hartwig Kalverkämper und Herbert Ernst Wiegand. In Verbindung mit Christian Galinski und Werner Hüllen. (Hrsg.). 1999. *Fachsprachen. Languages for Special Purposes. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft.* 2. Halbbd. 1762–1832, Handbücher zur Sprach- und Kommunikationswissenschaft 14(2). Berlin/New York: Walter de Gruyter.
- Bielińska, Monika.** 2010. *Lexikographische Metatexte. Eine Untersuchung nichtintegrierter Außentexte in einsprachigen Wörterbüchern des Deutschen als Fremdsprache.* Danziger Beiträge zur Germanistik 32. Frankfurt [etc.]: Peter Lang.
- Brenner, E.** 1951. *Deutsches Wörterbuch.* 2. Aufl. Wunsiedel/Oberfranken: Leitner & Co.
- CALD 2008 = Cambridge Advanced Learner's Dictionary.** 3. Aufl. Cambridge: Cambridge University Press.
- COED 2009 = Concise Oxford English Dictionary.** 11. Aufl. Bearb. von Catherine Soanes und Angus Stevenson. Oxford: Oxford University Press.
- Duden-GFWB 2003 = Duden. Das große Fremdwörterbuch. Herkunft und Bedeutung der Fremdwörter.** 3., überarb. Aufl. Hrsg. u. bearb. von der Dudenredaktion. Mannheim [etc.]: Duden.
- Dultz, Wilhelm.** 1965. *Fremdwörterbuch. Gebrauch und Bedeutung von Fremdwörtern.* Berlin: Frankfurt/Wien:Ullstein.
- Fenner, Kirsten.** 1997. Von Text zu Text: die Textsorte Wörterbuch als Vermittlerin bei der Rezeption und Produktion von Texten. Ein benutzerorientierter Ansatz. *Lexicographica* 13: 169-197.
- Gouws, Rufus H.** 2002. The Outer Texts in the *De Gruyter Wörterbuch Deutsch als Fremdsprache.* Wiegand, Herbert Ernst (Hrsg.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuchs Deutsch als Fremdsprache:* 471-484. *Lexicographica. Series Maior* 110. Tübingen: Max Niemeyer.

- Gouws, Rufus H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- 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: Mouton de Gruyter.
- Hartmann, Reinhard R.K. und Gregory James**. 2001. *Dictionary of Lexicography*. 2. Aufl. London/New York: Routledge.
- 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. *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie* [...]: 328-360. *Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1. Berlin/New York: Walter de Gruyter. [Wieder abgedruckt in Hartmann, R.R.K. (Hrsg.). 2003. *Lexicography. Critical Concepts. Vol. III: Lexicography, Metalexigraphy and Reference Science*: 207-254. London/New York: Routledge].
- Herberg, Dieter**. 1985. Zur Funktion und Gestaltung von Wörterbucheinleitungen. Hyldgaard-Jensen, Karl und Arne Zettersten (Hrsg.). 1985. *Symposium on Lexicography II. Proceedings of the Second International Symposium on Lexicography May 16-17, 1984 at the University of Copenhagen*: 133-154. *Lexicographica. Series Maior* 5. Tübingen: Max Niemeyer.
- Hessky, Regina und Bertalan Iker**. 2005. *Német-Magyar/Magyar-Német tanulósztár érettségizőknek és nyelvvizsgázóknak*. Szeged: Grimm Kiadó.
- Hueber DaF** 2006 = *Hueber Wörterbuch Deutsch als Fremdsprache. Learner's Dictionary. German-English. English-German. Deutsch-Englisch. Englisch-Deutsch*. Ismaning: Hueber.
- HWDG** 1984 = *Handwörterbuch der deutschen Gegenwartssprache in zwei Bänden*. Von einem Autorenkollektiv unter Leitung von Günter Kempcke. Berlin: Akademie-Verlag.
- Kammerer, Matthias und Herbert Ernst Wiegand**. 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica: International Annual for Lexicography* 14: 224-238.
- Kämpfert, Peter**. 1997. *Französisch im Küstenplatt. Ein sprachgeschichtliches Wörterbuch aus dem Lande Hadeln*. Neuhaus (Oste): Eigenverlag.
- Kienle, Richard von**. 1965. *Fremdwörterlexikon*. München: Keyser'sche Verlagsbuchhandlung.
- Klabunde, Ralf**. 1998. *Formale Grundlagen der Linguistik. Ein Arbeitsbuch*. Narr Studienbücher. Tübingen: Narr.
- Kluge, Friedrich**. 1995. *Etymologisches Wörterbuch der deutschen Sprache*. Bearb. v. Elmar Seebold. 23. erw. Aufl. Berlin/New York: Walter de Gruyter.
- Köster, Rudolf (Hrsg.)**. 1969. *DBG-Lexikon der deutschen Sprache. Wörterbuch für Rechtschreibung, Silbentrennung, Aussprache, Bedeutungen, Synonyme, Phraseologie, Etymologie* unter Mitarbeit von Harald Hahmann, Heribert Hartmann und Franz Mehling. Berlin [etc.]: Ullstein GmbH.
- Krämer, Walter und Roland Kaehlbrandt**. 2011. *Lexikon der schönen Wörter*. München/Zürich: Piper.
- LDW** 1955 = *Langenscheidts Deutsches Wörterbuch. Rechtschreibung und Grammatik*. Begründet von August Vogel. Neubearb. von Arthur Busse und Mitarbeit von Richard Pekrun. Berlin-Schöneberg: Langenscheidt KG.
- Mackensen, Lutz**. 1956. *Das große ABC. Ein Lexikon zur deutschen Sprache*. Bearb. v. Lutz Mackensen in Verbindung mit Robert Müller-Sternberg und Wolf Seidl. Düsseldorf: August Bagel Verlag.

- MWALED 2008** = *Merriam-Webster's Advanced Learner's English Dictionary*. Springfield, Massachusetts: Merriam-Webster.
- Nielsen, Sandro**. 1990. Lexicographic Macrostructures. *Hermes* 4: 49-66.
- OSACD 2006** = *Oxford South African Concise Dictionary*. 2. Aufl. Bearb. von der Dictionary Unit for South African English. Oxford: Oxford University Press.
- Paul, Hermann**. 2002. *Deutsches Wörterbuch. Bedeutungsgeschichte und Aufbau unseres Wortschatzes*. 10., überarb. u. erw. Aufl. von Helmut Henne, Heidrun Kämper und Georg Objartel. Tübingen: Max Niemeyer.
- PONS-DaF 2016** = *PONS Kompaktwörterbuch Deutsch als Fremdsprache mit Online-Wörterbuch*. Stuttgart: PONS GmbH.
- Rey-Debove, Josette**. 1971. *Étude linguistique et sémiotique des dictionnaires français contemporains*. (Approaches to Semiotics 13). The Hague/Paris: Mouton.
- Schierholz, Stefan J. und Herbert Ernst Wiegand**. 2004. Die Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK). Eine neue Konzeption der linguistischen Fachlexikographie und ihre computergestützte Praxis. *Lexicographica* 20: 164-264.
- Sellner, Alfred**. 1997. *Fremdsprachliche Redewendungen im Alltag. Sprichwörter, Floskeln, Phrasen, Formeln, Zitate, Sentenzen*. Wiesbaden: VMA-Verlag.
- Telling, Rudolf**. 1988. *Französisch im deutschen Wortschatz. Lehn- und Fremdwörter aus 8. Jahrzehnten*. 2. Aufl. Berlin: Volk und Wissen.
- Wahrig, Gerhard**. 2006. *Wahrig. Deutsches Wörterbuch*. Hrsg. v. Renate Wahrig-Burfeind. Mit einem Lexikon der Sprachlehre. 8., vollständig neu bearb. Aufl. Gütersloh/München: Wissen Media Verlag.
- Wiegand, Herbert Ernst**. 1988. Wörterbuchartikel als Text. Harras, Gisela (Hrsg.). 1988. *Das Wörterbuch. Artikel und Verweisstrukturen*. Jahrbuch 1987 des Instituts für deutsche Sprache: 30-120. Sprache der Gegenwart (LXXIV) Düsseldorf: Schwann.
- Wiegand, Herbert Ernst**. 1995. Lexikographische Texte in einsprachigen Lernerwörterbüchern. Kritische Überlegungen anlässlich des Erscheinens von Langenscheidts *Großwörterbuch Deutsch als Fremdsprache*. Popp, Heidrun (Hrsg.). 1995. *Deutsch als Fremdsprache. An den Quellen eines Faches. Festschrift für Gerhard Helbig zum 65. Geburtstag*: 463-499. München: Judicium.
- Wiegand, Herbert Ernst**. 1997. Printed Language Dictionaries and their Standardization: Notes on the Progress toward a General Theory of Lexicography. Hock, Hans Heinrich (Hrsg.). 1997. *Historical, Indo-European, and Lexicographical Studies. A Festschrift for Ladislav Zgusta on the Occasion of his 70th Birthday*. 319-380. Trends in Linguistics. Studies and Monographs 90. Berlin/New York: Mouton de Gruyter.
- Wiegand, Herbert Ernst**. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd. Mit 159 Abbildungen im Text. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst**. 2000. Über tabellarische Wörterverzeichnisse und deren Wörterbuchartikel. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 16: 212-234.
- Wiegand, Herbert Ernst**. 2000a. Über Suchbereiche, Suchzonen und ihre textuellen Strukturen in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. Wiegand, Herbert Ernst (Hrsg.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*. 233-301. Lexicographica. Series Maior 102. Tübingen: Niemeyer.

- Wiegand, Herbert Ernst.** 2001. Sprachkontaktwörterbücher. Typen, Funktionen, Strukturen. Iglá, Birgit, Pavel Petkov und Herbert Ernst Wiegand (Hrsg.). 2001. *Theoretische und praktische Probleme der Lexikographie. 1. Internationales Kolloquium zur Wörterbuchforschung am Institut Germanicum der St. Kliment Ohridski-Universität, Sofia. 7. bis 8. Juli 2000*: 115-224. Germanistische Linguistik 161–162. 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.** 2005a. Über die Datenakzessivität in Printwörterbüchern. Einblicke in neuere Entwicklungen einer Theorie der Wörterbuchform. *Lexikos* 15: 196-230.
- Wiegand, Herbert Ernst.** 2008. Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. *Lexicographica* 24: 209-315.
- Wiegand, Herbert Ernst.** 2010. Zur Methodologie der Systematischen Wörterbuchforschung: Ausgewählte Untersuchungs- und Darstellungsmethoden für die Wörterbuchform. *Lexicographica* 26: 249-330.
- Wiegand, Herbert Ernst.** 2013. Gedruckte Gebrauchsgegenstände mit lexikographischen Formeigenschaften. *Lexicographica* 29: 285-307.
- Wiegand, Herbert Ernst.** 2013a. Lexikographie und Angewandte Linguistik. *Zeitschrift für angewandte Linguistik*: 13-39.
- Wiegand, Herbert Ernst und Sandra Beer.** 2013. Textual Architectures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard und Herbert Ernst Wiegand (Hrsg.). 2013: 253-273.
- Wiegand, Herbert Ernst und Sandra Beer.** 2013a. Access Structures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard, Herbert Ernst Wiegand (Hrsg.). 2013: 110-149.
- Wiegand, Herbert Ernst, Sandra Beer und Rufus H. Gouws.** 2013. Textual Structures in Printed Dictionaries: An Overview. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard und Herbert Ernst Wiegand (Hrsg.). 2013: 31-73.
- Wiegand, Herbert Ernst und Rufus H. Gouws.** 2013. Macrostructures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard, Herbert Ernst Wiegand (Hrsg.). 2013: 73-110.
- Wiegand, Herbert Ernst und M^a Teresa Fuentes Morán.** 2009. *Estructuras lexicográficas. Aspectos centrales de una teoría de la forma del diccionario*. Granada: Ediciones TragaCanto.
- Wittstock, Otto und Johannes Kauczor.** 1988. *Latein und Griechisch im deutschen Wortschatz. Lehn- und Fremdwörter altsprachlicher Herkunft*. 5. Aufl. Berlin: Verlag Volk und Wissen (DDR).
- WLWF-1.** 2010 = Wörterbuch zur Lexikographie und Wörterbuchforschung, Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen/With English Translations of the Outer Texts and Definitions as well as Equivalents in nine Languages. Hrsg. u. Bearb. von/Ed. and compiled by Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer, Werner Wolski unter Mitarbeit von/with the Collaboration of Ekaterina Butina-Koller (Russisch/Russian), Rute Costa (Portugiesisch/Portuguese), M^a Teresa Fuentes Morán (Spanisch/Spanish), Laura Giacomini (Italienisch/Italian), Rufus H. Gouws (Afri-

kaans; Englisch/English), Franz Josef Hausmann/Maria Hegner (Französisch/French), Regina Hessky/Zita Hollós (Ungarisch/Hungarian), Pavel Petkov (Bulgarisch/Bulgarian), Giovanni Rovere (Italienisch/Italian), Stefan J. Schierholz (Portugiesisch/Portuguese), Maria Smit (Englisch/English). 1. Bd. Systematische Einführung/Vol. 1 Systematic Introduction. A–C. Berlin/New York: Walter de Gruyter.

How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions

Sascha Wolfer (*wolfer@ids-mannheim.de*)
and

Carolin Müller-Spitzer (*mueller-spitzer@ids-mannheim.de*)
*Institute for the German Language (Institut für Deutsche Sprache),
Mannheim, Germany*

Abstract: Wiktionary is increasingly gaining influence in a wide variety of linguistic fields such as NLP and lexicography, and has great potential to become a serious competitor for publisher-based and academic dictionaries. However, little is known about the "crowd" that is responsible for the content of Wiktionary. In this article, we want to shed some light on selected questions concerning large-scale cooperative work in online dictionaries. To this end, we use quantitative analyses of the complete edit history files of the English and German Wiktionary language editions. Concerning the distribution of revisions over users, we show that — compared to the overall user base — only very few authors are responsible for the vast majority of revisions in the two Wiktionary editions. In the next step, we compare this distribution to the distribution of revisions over all the articles. The articles are subsequently analysed in terms of rigour and diversity, typical revision patterns through time, and novelty (the time since the last revision). We close with an examination of the relationship between corpus frequencies of headwords in articles, the number of article visits, and the number of revisions made to articles.

Keywords: USER-GENERATED CONTENT, ONLINE DICTIONARY, WIKTIONARY, REVISION, EDIT, FREQUENCY, COLLABORATION, WISDOM OF THE CROWD

Zusammenfassung: Wie viele Menschen sind in einer 'Crowd', und was tun sie? Quantitative Analysen der Revisionen im englischen und deutschen Wiktionary. Wiktionary gewinnt immer mehr an Einfluss in vielen linguistischen Bereichen wie bspw. NLP und Lexikographie. Es hat das größte Potential, ein ernsthafter Wettbewerber für die Vertragslexikographie und akademische Lexikographie zu werden. Allerdings wissen wir wenig über die "Crowd", die für den Inhalt von Wiktionary verantwortlich zeichnet. Im vorliegenden Artikel wollen wir einige ausgewählte Fragen bearbeiten, die sich auf groß angelegte Kooperationsarbeit an Online-Wörterbüchern beziehen. Wir verfolgen dabei einen quantitativen Ansatz und verwenden die kompletten Historien des englischen und deutschen Wiktionarys als Datenbasis. Wir zeigen, dass — im Vergleich zur kompletten Autorenbasis des Wiktionarys — nur sehr

wenige Autoren für die überwältigende Mehrheit der Revisionen in beiden Wiktionarys verantwortlich sind. Im Folgenden vergleichen wir diese Verteilung mit der Verteilung der Revisionen über alle Artikel. Dann werden die Artikel hinsichtlich Gründlichkeit und Diversität, typischen Revisionsmustern in der Zeit sowie der Neuigkeit (Zeit seit der letzten Revision) untersucht. Wir schließen mit einer Analyse des Zusammenhangs zwischen Korpusfrequenz des Stichworts, der Anzahl der Seitenaufrufe des Artikels und der Anzahl der Revisionen des Artikels.

Stichwörter: NUTZERGENERIERTE INHALTE, ONLINEWÖRTERBUCH, WIKTIONARY, REVISION, ÜBERARBEITUNG, FREQUENZ, KOLLABORATION, SCHWARMINTELLIGENZ

1. Introduction

There is an on-going debate about whether collaboratively constructed dictionaries have the potential to become serious competitors for publisher-based and academic dictionaries (Hanks 2012, Meyer and Gurevych 2012, Rundell 2012). The most promising candidate currently available is Wiktionary — the dictionary project of the Wikimedia foundation. Wikimedia's main project, Wikipedia, has already proven its potential to cover large proportions of user needs in terms of encyclopaedic knowledge; at least if we use page view statistics as an indicator for user satisfaction: On 2016-05-23, the English Wikipedia alone registered over 5,600,000 page views per hour. Studies suggest that the Wikipedia community "takes issues of quality very seriously" (Stvilia et al. 2008)¹. But Wikipedia's success does not necessarily imply that the same foundation's dictionary project is going to have a comparably major impact on the global dictionary landscape. But Wiktionary is obviously used by many people² for a wide array of linguistic needs. And, as we elaborate in the next section, Wiktionary content is also widely used as a scientific resource.

Our main focus is not on investigating the quality of Wiktionary content (see relevant literature in the next section) but more on the processes that shape Wiktionary. It is, in our opinion, essential to get to know the crowd behind Wiktionary a little better in order to gain insights into its composition and processes. This information can help us paint a more detailed picture of "the crowd", which in turn will help us to research Wiktionary and its implications for lexicography as a whole. A good starting point is the revision (or edit) history that determines the state of Wiktionary. Keep in mind, though, that even as you are reading this article, Wiktionary is changing. We can only look at a specific snapshot at a specific point in time. However, even if Wiktionary might look different now than it did at the time of writing this article, we are confident that general principles regarding crowd composition and behaviour that are not subject to sudden change but evolve over much longer periods of time can be deduced. With data files supplied by the Wikimedia foundation (which will be described later in the article), we can consult the complete edit history of all available Wiktionary language editions. And we might identify some general underlying principles regarding *how many people* revise articles *when*

and *how*.

Many of the Wiktionary processes are run automatically using so-called bots. These revisions can also be found in our dataset. We are, however, primarily interested in non-automatic and non-minor revisions. As we will elaborate later, these revisions primarily shape the dictionary and might involve editorial choices.

A note on how we refer to the individual elements of "the crowd": Usually, the term "users of a dictionary" is reserved for the recipients of a dictionary. This terminology contrasts the users and the lexicographers as authors of a dictionary. However, this distinction is not that easy for dictionaries containing user-generated content such as the Wiki-based Wiktionary. Meyer and Gurevych (2012: 271-272) mainly use the term "users" for the authors of Wiktionary, the "Wiktionarians". Lew (2014) no longer considers users of the Web 2.0 (which Wiktionary is as an example of) as "passive recipients of packaged content". Rather, "they actively contribute to the creation and provision of self-made content. This double capacity of newly empowered users can be aptly captured in the neologism *prosumer*, which is a blend of *producer* and *consumer*." (Lew 2014: 1). Every individual who accesses a page in Wiktionary can choose to contribute to the dictionary at any given time by clicking the "edit" button in the upper right corner of the browser. Therefore, we acknowledge that "web users' social roles become blurred" (*ibid.*). However, to be as clear as possible, we will use the term "author" for people involved in revision processes and "user" to refer to people who looked something up in Wiktionary.

The remainder of the article is structured as follows: In the next section, we will introduce some related work about the quality of Wiktionary. We will also introduce several scientific applications of Wiktionary content. Section 3 will deal with data preparation and pre-processing of the history files. Also, we will introduce some basic statistics related to revisions, especially the relationship between automatic and minor revisions within our datasets. The main section of the article is Section 4. Initially, we will investigate the number of authors and, more importantly, the distribution of revisions among authors (4.1). In Section 4.2, we will highlight some of the core editing processes in the two language editions. Here, several questions are of interest: How are revisions distributed over entries (and, in the same vein, the relationship between rigour and diversity)? When is the crowd most active? Are there typical chronological revision patterns? How old are the entries (*i.e.*, how long are the phases during which no-one revises articles)? Is there a relationship between revision frequency, number of visits and corpus frequency of the headword? In Section 5, we will provide a summary including some closing remarks.

2. Related work

In terms of content quality of in Wiktionary, there is disagreement. Meyer and Gurevych (2012), amongst others, refer to the "wisdom of the crowd" (Surowiecki

2005) phenomenon and express their hope that it makes up for the potential "lack of lexicographic experience" (Meyer and Gurevych 2012: 271). They further state that "[c]ollaboratively constructed lexicons are continually updated by their community" which "yields a steeply increasing coverage of words and word senses. [...] An important characteristic of collaborative lexicography is that the large number of authors has the ability to express the actual use of language [...]." (Meyer and Gurevych 2012: 259). Hanks (2012) is not that optimistic. While acknowledging positive aspects like "imaginative use [...] of multimedia hypertext" (p. 81) which he sees as a "model for the electronic dictionary of the future" (p. 82), he also states that "[i]n the English Wiktionary, the etymologies are taken from or based on those in older dictionaries; as are definitions, which are extremely old-fashioned and derivative, taking no account of recent research in either cognitive linguistics or corpus linguistics" (p. 78). As stated above, we will not compare the quality of Wiktionary entries to professionally edited dictionaries in this article. Instead, we want to gain a more detailed insight into the revision processes in two Wiktionary editions: The English language edition, which is the largest in terms of number of pages/articles (at the time of writing this article, the English Wiktionary had over 4,250,000 entries), and the German language edition, which is rather small in comparison (the German Wiktionary had approx. 430,000 entries at the time this article was written, which is roughly a tenth of the English Wiktionary, and ranks 14 among all language editions)³. As previously mentioned, it is essential to shed some light on the crowd that is behind Wiktionary to discuss the position of Wiktionary in the lexicographic landscape. Another reason why we want to get to know the crowd is the wide-spread application of Wiktionary as a data source for a range of scientific applications, including works in the field of natural language processing (NLP) like sense definitions (Henrich, Hinrichs and Vodolazova 2011), semantic relatedness (Zesch et al. 2008), synonymy networks (Navarro et al. 2009), pronunciation extraction (Schlippe et al. 2010), idiom identification (Muzny and Zettlemyer 2013) and many more. Other areas of application are sentiment analysis (Chesley et al. 2006) and analyses of vocabulary difficulty (Medero and Ostendorf 2009). In addition, some parts of Wiktionary entries are integrated into other lexicographic resources, for example in bilingual dictionaries (Lindemann 2014). As can be seen from this (doubtlessly incomplete) list, Wiktionary has become an increasingly important resource, which contrasts with the little knowledge we actually have of the processes that create(d) it. We hope that this article can contribute to a better understanding of these processes.

Some questions we want to answer using Wiktionary revision histories have been similarly dealt with by other researchers, but mainly with respect to the Wikimedia foundation's largest project, Wikipedia (Greenstein and Zhu 2012, Poderi 2009, Stein and Hess 2008, Wilkinson and Huberman 2007). Since Wiktionary uses the same platform, we can transfer some of the ideas that have been previously applied to Wikipedia. In Section 4.2, we refer to an idea by Lih (2004)

who sees Wikipedia as "the largest example of participatory journalism to date" and compares a set of benchmark articles to Wikipedia articles that have been cited in the press. Although this question is not directly relevant for the topic of this article, Lih introduces the notions of *diversity* and *rigour*⁴. We will define and use these concepts in Section 4.2 to compare entries within one Wiktionary language edition and also to compare the two language editions to one another.

In Section 5, we will summarise our findings and discuss some implications for the lexicographic landscape.

3. Data preparation

3.1 Downloading and pre-processing Wikimedia history files

Revision data were extracted from the edit history data dumps available from the Wikimedia Foundation⁵ on 2015-08-07. The files are available in XML format and have to be parsed with a SAX parser due to their size. We used an R (R Core Team 2015) script and the XML package (Lang 2013) to implement the parser. The complete edit history files were converted to CSV files with the following information associated with each revision: (1) the title of the revised page, (2) the revision timestamp, (3) the name of the author who carried out the revision, (4) whether the revision was flagged as minor, and (5) the comment the author added to the revision. It is not trivial to decide whether a revision was made automatically (i.e., by a bot) or by a human. After consultation with contributors to the German Wiktionary, we used two lists of user names provided by the English⁶ and German⁷ Wiktionary language edition. These lists contain all users that are flagged as bots, which is the standard procedure to identify users as bots in Wikimedia products. The consultation with the contributors showed that this is — if not perfect — the most reliable way to identify automatic revisions. In the next step, we also tagged all revisions as automatic that were associated with a comment containing the strings "autoedit" or "clean up". Results concerning the distribution of automatic and non-automatic revisions are summarized, amongst others, in Table 1.

In this article, we only look into the revision history of the English and German Wiktionary language editions. Since it takes time and computational effort to parse the history files, we compiled CSV files for eight language editions of Wiktionary and provide them to the scientific community to conduct analyses for other language editions and/or expand the presented set of analyses. CSV revision history files for English, Malagasy, French, Russian, Polish, German, Chinese, and Spanish are available under <http://dx.doi.org/10.7910/DVN/TYLQBN>. The languages are ordered according to the number of revisions. All files contain the revision history as of August 2015. If readers are interested in creating CSV files for another language edition and/or newer history files, the respective R scripts are also available under <http://dx.doi.org/>

10.7910/DVN/TYLQBN. We would greatly appreciate if any output generated by these scripts were also to be made available to the scientific public.

3.2 Minor and automatic revisions

Whenever authors revise a page in Wiktionary, they can flag their revision as minor. This is a signal to others "that only superficial differences exist between the current and previous versions [of the page]. Examples include typographical corrections, formatting and presentational changes, and rearrangements of text without modification of its content. A minor edit is one that the editor believes requires no review and could never be the subject of a dispute" (https://en.wikipedia.org/wiki/Help:Minor_edit, last access on 2015-11-13). Although it is not perfectly safe to rely on authors checking the respective box if they only made minor adjustments to the page, we use this information in our analyses, simply because it is the most reliable source of information available. The help page cited above is quite clear about when a revision should be flagged as minor, and we believe that, since both the English and German Wiktionary have existed for more than ten years, authors are most likely acquainted with the use of the check box now.

Table 1: Date of first revision of a non-meta page, number of total revisions, revisions of meta pages as well as automatic and minor revisions for the English and German Wiktionary edit history

	English Wiktionary	German Wiktionary
Date of first revision of non-meta page	2001-09-27 ("zen")	2002-04-24 ("Metamorphose")
Total revisions	32,767,004	4,362,860
Revisions of meta pages	2,343,380 (7.15%)	398,930 (9.14%)
Automatic revisions	21,693,943 (66.2%)	2,831,858 (64.9%)
Minor revisions	23,736,184 (72.4%)	2,991,574 (68.6%)

Table 1 shows some key figures of minor, automatic and meta page revisions in both Wiktionary editions. Meta pages are all pages that are not "normal" entries. We identified meta pages by a colon in the page name. Note that there is not only one type of meta page. There may be talk pages associated with each article, user pages for each user, and user talk pages that are associated with each user page. Also, there are several meta pages that are not associated with other pages, such as help and question pages, several special pages and so on. Actually, the term "meta page" is not used for these kinds of pages in the Wiktionary. In the context of this article, meta page can be translated as "not an entry page".

Meta pages are frequently revised during the early stages of a new Wikimedia project so as to set up all relevant pages. This is reflected by the devel-

opment of shares of meta page revisions. In the English Wiktionary, for example, 29.5% of the first 10,000 revisions were meta page revisions. Compared to the overall value of 7.15% (see third row in Table 1), this share is quite high. The first non-meta revision of the English Wiktionary edition was approximately half a year earlier than in the German Wiktionary.

In the English Wiktionary, 7.5 times more revisions were made than in the German Wiktionary. Later in this article, we will also deal with the number of authors who contribute to the two Wiktionary editions. There, we will also see that many more authors work on the English Wiktionary than on the German Wiktionary.

An important piece of information missing from Table 1 is the cross-combination of automatic/human and minor/non-minor revisions. This information is available in Table 2. In both language editions, minor non-automatic revisions constitute around two thirds of all revisions. Automating tedious repetitive processes with bots obviously only makes sense if a lot of revisions can be performed with the bot program. So, this high share of automatic revisions is no surprise. The revisions we will be primarily interested in in this article are the ones which fall into the lower right cells of the contingency tables: non-minor, human revisions. In both language editions, these revisions account for around a fifth of all revisions. The large values for χ^2 (both above 1 million and hence not included here) and ϕ suggest that the relationships between the automation of revisions and the fact that they are flagged as minor in Table 2 are strong and did not occur by chance.

Table 2: Frequencies, percentages, and marginal sums for automatic/human vs. minor/non-minor revisions of non-meta pages in the English and German Wiktionary editions. Association measure for English: $\phi = .671$, for German $\phi = .658$

English Wiktionary			German Wiktionary				
	minor	non-minor	total		minor	non-minor	total
automatic	20,034,791 65.9%	1,187,568 3.90%	21,222,359 69.8%	automatic	2,518,446 63.5%	240,180 6.06%	2,758,626 69.6%
human	2,900,133 9.53%	6,301,132 20.7%	9,201,265 30.2%	human	324,032 8.17%	881,272 22.2%	1,205,304 30.4%
total	22,934,924 75.4%	7,488,700 24.6%	30,423,624 100%	total	2,842,478 71.7%	1,121,452 28.3%	3,963,930 100%

4. Results and discussion

4.1 How many people constitute a crowd?

The OED editorial staff comprises around 125 members⁸. This is without a doubt a very high number of lexicographic experts working on a group of dic-

tionaries and presumably the biggest editorial dictionary worldwide. Just for comparison: the staff at Duden, the most prominent publisher's dictionary for German, consists of eight permanent lexicographers⁹. It is safe to consider these people lexicographic experts as well. Nonetheless, we have to ask ourselves whether it might also be a good idea to use a (freely available) dictionary that can be potentially edited by *all* speakers of a language with access to the Internet? We should consider another question before making a judgement on this matter: How many people actually constitute the crowd behind Wiktionary? Meyer and Gurevych (2012: 272) already refer to the Zipf-like nature of the distribution of revisions over authors. This means that the vast majority of authors make very few revisions and that a small group of authors (compared to the overall number of registered authors) are responsible for the majority of revisions (just like the most frequent word types in a language accumulate the vast majority of word tokens). This statement is certainly correct as we will see in a moment. However, we want to clarify what this means in terms of numbers. In this section, we will only consider non-automatic, non-minor revisions of non-meta pages, because these are the types of revisions that really shape the dictionary and add the main content to entries. For the English Wiktionary, this applies to 6,301,132 revisions. The German Wiktionary history file contains 881,272 such revisions.

Both Wiktionary language editions allow visitors to revise articles without registration. The revision is then associated with the current IP address of the respective visitor. In the revision history of the English Wiktionary, 750,055 (11.9%¹⁰) revisions were made by unregistered authors. Unregistered authors made 144,002 revisions (16.3%) in the German Wiktionary. One might object that these are too many revisions to exclude from the analyses. However, there are certain problems associated with the analysis of IP-based data. The most crucial problem is that, for IP addresses, we cannot be sure if one person always edits with only one IP and, on the other hand, whether one IP always identifies one user¹¹.

We consider data from 36,958 registered authors for the English Wiktionary and 6,111 authors for the German Wiktionary. An author needed at least one non-minor revision to a non-meta page to enter the analysis. As can be seen in Table 3, the total number of non-minor revisions is distributed very unevenly among the authors. In the English Wiktionary, 33 authors are responsible for over half of all non-minor revisions to non-meta pages. In the German Wiktionary, only 14 authors made the majority of revisions. This obviously means that the tails of these distributions are very long, i.e. there are many authors with only few revisions. In the English Wiktionary, almost half of all registered authors (44.3%) only made one revision. It is similar in the German Wiktionary: There, 42.3% of all registered authors only made one revision.

Table 4 is a transformation of Table 3, i.e. the number of revisions is used as the starting point. It includes shares for at least 2, 5, 10, 100, and 1000 revisions respectively. In the English Wiktionary, an author needs to have made

1,346 revisions to be included in the top 1% of authors in terms of revisions¹². In the German Wiktionary, this threshold lies at 1,571 revisions. Given that the maximum number of revisions an author made is 237,600 for the English Wiktionary (user "Equinox"¹³) and 57,432 for the German Wiktionary ("Dr. Karl-Heinz Best"), this is a rather low value and shows how extreme this distribution behaves when it approaches its extremes.

Table 3: Number of top ranked authors in terms of non-minor revisions to non-meta pages and the associated shares of all those revisions. Included ranks differ after 100 for the different language editions to accommodate for the different numbers of authors. Also included are the minimum numbers of authors needed to reach over 50 and 95 percent of all revisions

Number of top ranked authors, English Wiktionary	Share of all revisions in the English Wiktionary	Number of top ranked authors, German Wikt.	Share of all revisions in the German Wiktionary
10	26.8%	10	42.8%
33	50.2%	14	50.2%
50	58.3%	50	80.6%
100	71.1%	100	89.2%
1000	94.9%	262	95.0%
1026	95.0%	500	97.1%
5000	98.6%	1000	98.3%
10000	99.2%	2000	99.1%

Table 4: Share of authors with at least n revisions in the English and German Wiktionary

Number of revisions <i>n</i>	Share of authors in the English Wiktionary	Share of authors in the German Wiktionary
at least 2	55.7%	58.0%
at least 5	25.5%	28.4%
at least 10	15.8%	17.4%
at least 100	4.27%	4.63%
at least 1000	1.20%	1.37%

The figures we presented in this section are informative if we want to get an impression of the revision distribution over users. We can, however, take another perspective by conducting a thought experiment. We compare the number of users to the sizes of the editorial staff of two dictionary publishers in the respective countries. The OED is one of the most well-known publishers for the English language and the Duden for German respectively. What would it mean for the revisions in Wiktionary if we transfer the respective staff sizes to

the number of Wiktionary authors?

We can compare the OED's staff size by checking the number of revisions of the English Wiktionary for the top 125 authors. Or to put it differently: We check, what percentage of all revisions the top 125 authors made. The value is 74.8%. This could be considered an impressive figure. However, also note that we would lose 1,400,313 human, non-minor revisions in the English Wiktionary if we were to exclude all authors not in the top 125 authors. At Duden, eight permanent employees work on entries. The top 8 authors contributing to the German Wiktionary made 38.1% of all revisions. If we would exclude all revisions from rank 8 downwards, we would lose 456,725 revisions.

4.2 Revision processes

Distribution of revisions over entries

As we saw in the previous section, revisions are very unevenly distributed among authors. From the article perspective, we can check how revisions are distributed among the entries in the two Wiktionary language editions accordingly. We are primarily interested in pages related to entries, so we excluded all meta pages from this analysis. Also, we only consider human, non-minor revisions to keep analyses from the article perspective as comparable as possible to the analyses from the author perspective. This leads to a data set with 1,983,023 article pages from the English Wiktionary and 228,869 article pages from the German Wiktionary.

Table 5: Number of top ranked articles in terms of non-minor revisions to non-meta pages and the associated shares of all revisions. Included ranks differ after 100 for the different language editions to accommodate for the different numbers of articles. Also included is the minimum numbers of articles needed to reach over 50 percent of all revisions

Top ranked articles in the English Wiktionary	Share of all revisions in the English Wiktionary	Top ranked articles in the German Wiktionary	Share of all revisions in the German Wiktionary
10	0.166%	10	0.310%
50	0.532%	50	0.990%
100	0.891%	100	1.60%
1,000	4.60%	1,000	7.61%
10,000	17.0%	10,000	30.8%
100,000	42.5%	26,203	50.0%
163,072	50.0%	100,000	82.6%
500,000	71.0%	200,000	96.7%
1,000,000	84.4%		
1,500,000	92.3%		

Table 5 shows the respective distribution. It is obviously less extreme than the distribution over authors. Let us simply compare the share of revisions the 100 most edited articles received: In both Wiktionary editions, the top 100 edited articles received less than 2% of all revisions. For authors (see Table 3), this is a very different picture: The top 100 authors in the English Wiktionary are responsible for over 70% of all revisions, in the German Wiktionary, the respective value is almost 90%.

One might argue that the distributions over authors and articles cannot be compared in such a way because there are far more articles than authors in both Wiktionaries. A way to work around this is presented in Figure 1. On the *x*-axis, we included more and more percentages of articles or authors. On the *y*-axis, we calculated the share of revisions these authors made or these articles received for each step. The grey line shows us how the graph would look if revisions were perfectly evenly distributed over articles and authors (e.g. that 50% percent of authors made exactly 50% of all revisions). We can see that all distributions deviate clearly from the grey line and that there are slight differences between the language editions.

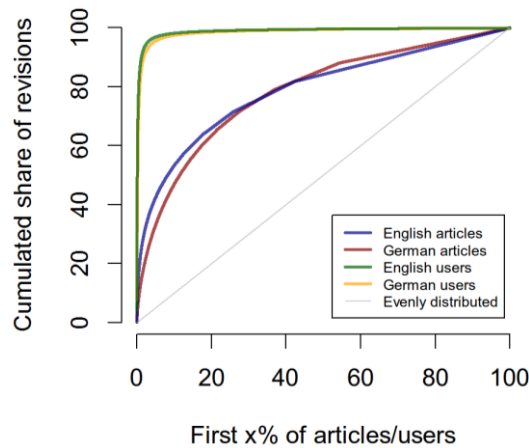


Figure 1: Top x% of articles/authors in the English and German Wiktionary editions and their associated cumulated share of revisions. The straight grey line symbolises a hypothetical distribution where each article and author receives the same share of revisions

The clearest difference is between the distributions for articles and authors, though. We need to include a much larger share of articles to reach a specific amount of revisions. For example, to register 80% of all revisions, we only need to include 0.474% of all authors (which corresponds to 175 authors) but we need to include 39.3% of all articles (which corresponds to 780,185 articles). In summary, we have to acknowledge that revisions are distributed very un-

evenly over authors **and** articles, but the distribution among authors is more extreme than for articles.

From a lexicographic point of view, this might be interpreted as a good sign. Although only few authors are responsible for the vast majority of revisions in both Wiktionaries, this does not imply that only very few articles receive preferential treatment. These few authors distribute their work in the dictionary over a wide range of articles. On the other hand, this may also support the impression that not many articles are very sophisticated. We will elaborate on this question in the following section by introducing the terms rigour and diversity.

Rigour and diversity

As briefly mentioned in the introduction, Lih (2004) used the terms rigour and diversity in a study of Wikipedia articles. Rigour is the variable we analysed in the previous subsection: the total number of edits for an article. "The assumption is that more editing cycles on an article provides [sic] for a deeper treatment of the subject or more scrutiny of the content" (Lih 2004: 8). We already saw that some articles are edited more rigorously (in this sense) than others. The idea behind the concept of diversity is that "[w]ith more editors, there are more voices and different points of view for a given subject" (ibid.). Given this operationalisation, the more unique authors contributed to an article, the more diverse it should be. Following Lih (2004), both concepts can be interpreted as the higher, the better. Although the concepts were originally applied to Wikipedia articles, they can just as easily be applied to our Wiktionary datasets. Again, we are only taking non-minor revisions to non-meta pages from registered authors into consideration. Given the definitions of diversity and rigour, the value for diversity may never exceed the value for rigour. If 100 different authors make 100 revisions to an article, diversity and rigour are equal (100). As soon as at least two of the 100 revisions are made by the same author, the value for diversity decreases while the value for rigour remains constant. The more interesting question then is by how much the data points for the different articles deviate from the baseline, which is defined as rigour and diversity being equal. In Figure 2, all articles (non-meta pages) from both Wiktionaries are plotted. The two language editions are colour-coded. The location of an article is defined by its diversity (x -axis) and rigour (y -axis).

Several conclusions can be drawn from the graph: (1) Articles in the English Wiktionary are more diversified, both in terms of diversity and rigour. This might simply be due to the fact that there is much more potential for an article to get higher values of diversity and rigour in the English Wiktionary, because there are 7.16 times more revisions and 6.05 times more authors in the English than in the German Wiktionary. However, there are also 8.66 times more articles in the English than in the German Wiktionary. So one might argue that there is also much more "ground to cover" by revisions and authors in the Eng-

lish Wiktionary. (2) The bi-variate distribution of articles in terms of diversity and rigour clearly differs from the baseline. This suggests that more than one author normally contributes to an article. (3) The "upward bend" of the overall pattern suggests that the relationship between diversity and rigour is not strictly linear. The more the rigour of an article increases, the less pronounced the increase in diversity. The comparison of two regression models predicting rigour by diversity suggests that the quadratic trend explains additional variance compared to only the linear trend¹⁴. This means that the relationship "more rigour means more diversity" is weaker for articles with very high rigour values.

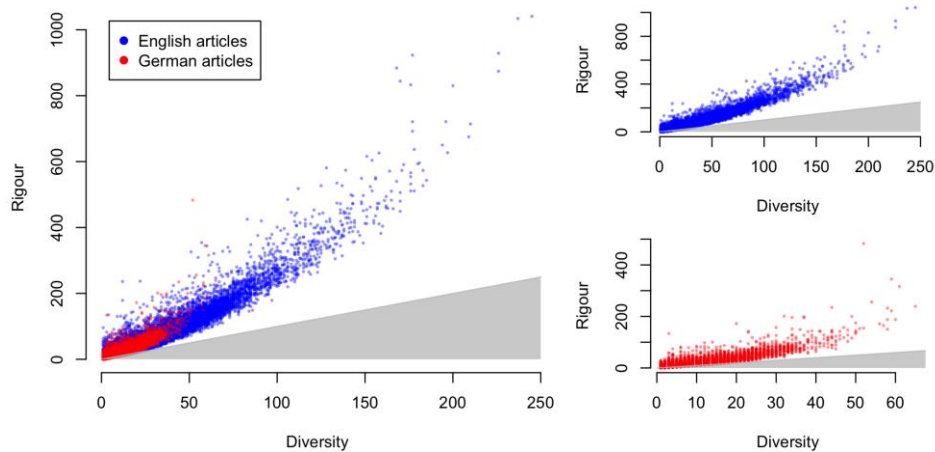


Figure 2: Diversity (number of unique registered and contributing authors) and rigour (number of revisions) of the articles in the English and German Wiktionary language editions (both plotted on top of each other in the left panel and separate plots for both languages on the right). The grey area symbolizes the area of the graphs where rigour is lower than diversity. No data point may fall in this area. The data point for "water" (English Wiktionary) was excluded from the graph for better readability. Its coordinates are (2284, 231)

Although Figure 2 spans a broad range of diversity and rigour values, we also have to point out that the majority of articles in both Wiktionaries score rather low on these scales. Visually, this is represented by the very dense data point cluster in the lower left region of the graph. The mean diversity of articles is 2.09 (English) and 2.59 (German). Mean rigour ratings are 3.18 (English) and 3.85 (German). Due to the skewness of the distributions, the respective medians¹⁵ are slightly lower. Which conclusions can we draw for the revision processes in Wiktionary given these results? As might not be apparent in Figure 2, the majority of articles in both Wiktionaries were edited by a single author

only¹⁶ (English: 67.3%, German: 53.0% of all articles). Would we apply a four (or more) eyes principle to ensure the quality of Wiktionary articles, these articles might not be sufficient. However, to evaluate this question further, we could also take discussion pages into consideration. It might well be the case that potential problems in articles are expressed on the accompanying discussion page and corrected by the one author who edited the associated article.

Typical revision patterns over time per article

We are dealing with over 13 years of history for two online dictionaries. So, it would be a shame to not take time into consideration as well. Two potentially interesting questions are: Is there a "typical" revision pattern over time for articles? And how long are typical "idle periods" of articles, meaning the time spans in which no-one works on the article (as measured by non-minor, non-automatic revisions)? The first question is not easy to answer because it is not clear how to operationalise the critical concepts "typical" and "revision pattern". We propose a visualisation to gain a first insight into revision histories of individual articles. On the x -axis, we plot the time since article creation (technically, article creation is the first revision). On the y -axis, we plot the cumulated share of revisions of this article on the specific day. The path one line takes through time can be called the "revision trajectory" of an article. Whenever long horizontal lines appear in the revision trajectory, no-one worked on the article for a long time. Vertical lines within such a revision trajectory indicate a period when the article was revised repeatedly in rapid succession. To keep the plot visually manageable, we only include the top 1000 revised articles from the English Wiktionary.

Figure 3 shows these revision trajectories. There is an accumulation of trajectories on the middle diagonal of the graph. This represents a steady revision process with revisions coming in quite regularly. The randomly chosen article "chicken", symbolised by the green line in Figure 3, is an example for this. There are no clear phases when "chicken" receives many revisions (which would lead to steeper lines) or when it receives very few revisions (which would lead to a more horizontal trajectory). The trajectories for "water" and "cool" are quite similar: Both revision histories start off slowly. Then, both articles receive many revisions in rapid succession. The "revision sprint" sets in later for "water" than for "cool"¹⁷. After that, the slope of the trajectory of "water" is slightly higher than for "cool" and there is also a minor revision sprint at around 4000 days since article creation. A few other, more extreme trajectories are also visible in the upper left region of the graph. These are articles that received a lot of revisions early in their history and are then only revised very infrequently — supposedly because the community thinks they are finished or no-one takes care of them anymore. So, we think the quantitative view of this data shows no intriguing patterns. It might leave room for further research to

examine groups of articles that share a similar revision pattern more qualitatively in order to find out whether these groups of headwords share particular (linguistic) properties.

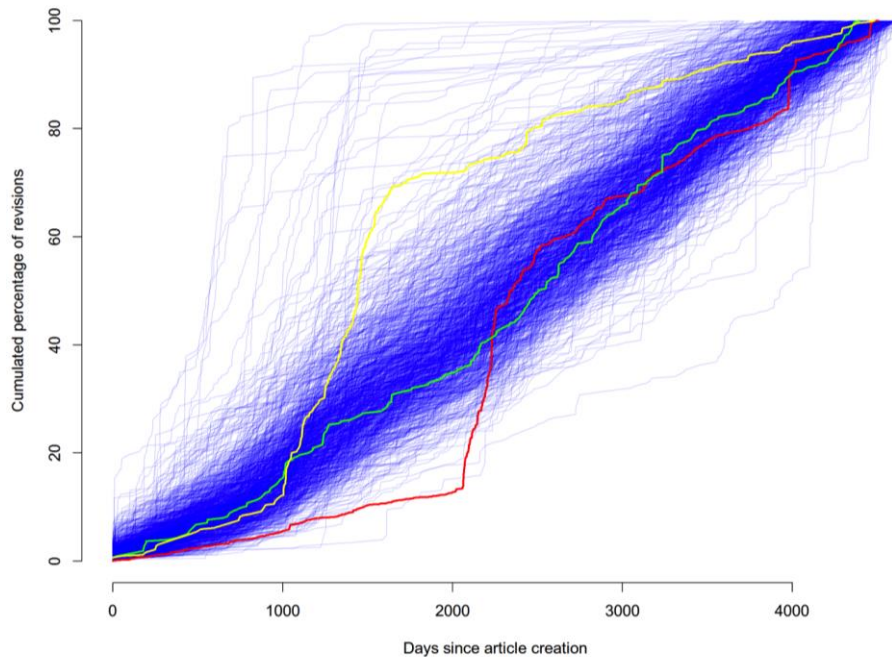


Figure 3: Cumulated percentage of revisions in relation to the days since article creation for the 1000 most revised articles from the English Wiktionary. Each line ("revision trajectory") represents one article. Three revision trajectories are highlighted. The red line symbolises "water", the yellow line "cool" and the green line "chicken". The latter two were chosen randomly. "Water" is the article that received the most revisions in the English Wiktionary

How old are the entries?

The advantage online dictionaries have over printed dictionaries is the flexibility to adapt an entry within a very short period of time because the authors do not have to wait for the next edition of the printed book. This advantage should be especially relevant for Wiktionary as well as all other Wikimedia products. After all, "wiki" is Hawaiian for "fast". So, we would expect that articles are regularly revised and we do not expect a large number of articles that remain unrevised for a very long time. A fast and easy way to assess the typical idle

period of articles is to take the current state of articles and calculate how long ago the last revision was concluded. We can then calculate the means or medians of these time spans for both Wiktionaries. The mean (median) time span during which an article is not revised in the German Wiktionary is 257 (143) days. For the English Wiktionary this time span is longer: 441 (386) days.

Another way to approach this is to start from a given number of days x (e.g. 183 days, half a year). We then count backwards starting from 2015-08-07, the day the history dump was created, until we arrive at the day x before 2015-08-07 (in our example, this would be 2015-02-05). We then extract the number of articles n_x that have not been revised within this time span. Again, we only consider human, non-minor revisions. If we divide n_x by the total number of articles, we arrive at the percentage of all articles that have not been revised in the last x days, starting from 2015-08-07. In the example, 1,443,141 articles have not been revised in the time between 2015-02-05 and 2015-08-07 ($n_{183} = 1,443,141$). These are 72.8% of all articles. In the German Wiktionary, n_{183} is only 106,189. However, there are also fewer articles in the German Wiktionary. Nevertheless, these are only 46.4% of all articles in the German Wiktionary. If we now vary x from 1 to 4134 days (which is the longest time during which an article was not revised) and calculate each n_x and the respective share of articles, we can plot x and the share in one graph. Figure 4 is the result of this process. The further we go back in time, the fewer articles were not revised within this time span. There are certain differences between the two language editions, though.

The values decrease faster for the German Wiktionary and also get to the "bottom line" much faster than the values for the English Wiktionary. On the one hand, this is not that surprising because there are fewer articles in the German Wiktionary. So, it should be easier to cover a large share of these articles with revisions. However, one should also consider the number of authors in this argumentation. Given that the community shaping the English Wiktionary is considerably larger than that of the German Wiktionary, we would have expected patterns here that are more similar. The rather steep decreases, especially in the German Wiktionary are another interesting observation found in Figure 4. For example, the line drops sharply between 500 and 600 days, i.e. many articles were revised in a comparably short time. We do not have a final explanation for this. It could be due to the German Wiktionary's official 10th "birthday" on 2014-05-01¹⁸ that motivated authors to revise a lot of articles. The pattern visualised in Figure 4 is also reflected by the values in Table 6. There, we change perspective from non-revision to revision: How many articles have been revised in the last x years? In this way, we see that the majority of articles from the German Wiktionary were revised in the last 6 months while for the English Wiktionary, only roughly a quarter of the articles were revised in the same time span.

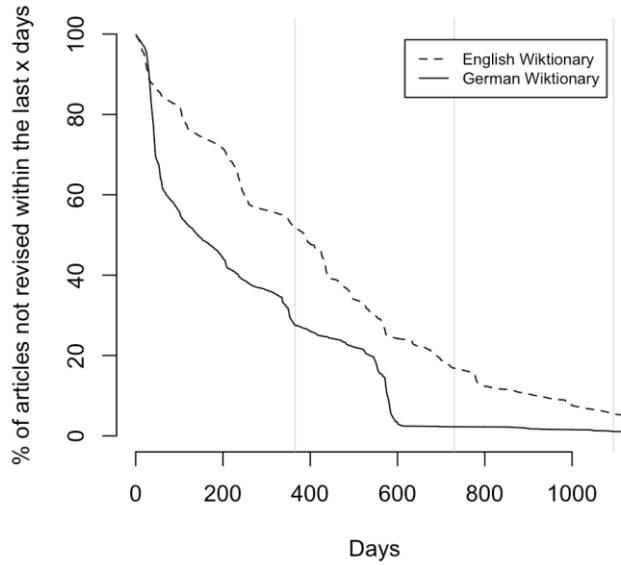


Figure 4: Share of articles that have not been revised in the last x days in relation to the time span x, starting from 2015-08-07 counting backwards in time. Vertical grey lines indicate one, two, and three years

Table 6: Share of articles that were revised within the last x years in the English and German Wiktionary language editions

Time span x	Share of articles in the English Wiktionary	Share of articles in the German Wiktionary
½ year	27.2%	53.5%
1 year	47.8%	72.4%
2 years	83.1%	97.7%
3 years	94.5%	98.9%

We can also look at a few of the articles that have not been revised in a very long time. In the German Wiktionary, for example, there are 1,758 articles that have not been revised in the last 5 years. These contain many articles for non-German words but also entries for German words. Interestingly, most of these articles are permanent redirections (e.g., "mit Mann und Maus untergehen" → "Mann und Maus"; "m.E." → "m. E."; "labio-dental" → "labiodental"; "jenes" → "jener") or very short articles only indicating alternative ("gewerbsmässig", "einigermassen") or old spellings ("deplaciert", "deplaziert"). All these articles have a "reference article" they are linked to (the redirection target, another spelling or the correct spelling) that can be updated. So, it makes sense that there are indeed a few articles that have not been revised in a very long time.

The relationship between revisions, number of visits and corpus frequency

As we have previously shown (Müller-Spitzer et al. 2015), articles with high frequency headwords are visited more frequently in the German Wiktionary. In this article, we introduce another variable to this relationship: the number of revisions an article receives during the history of Wiktionary. It would make sense for the authors of Wiktionary to revise articles that are relevant to a broad public. Or put differently, the question we want to answer is whether articles that are revised more frequently are also visited more frequently. To answer this question, we enrich the article-based data set with the number of visits based on the Wikimedia log files available from <https://dumps.wikimedia.org/other/pagecounts-raw/> (last accessed on 2015-12-02). We use the number of visits during 2014 for this analysis. As corpus frequency measures for the German Wiktionary, we use word form frequencies based on the German Reference Corpus/Deutsches Referenzkorpus (DeReKo, Kupietz et al., 2010). For the English Wiktionary, we used a frequency list based on the Google Books 2012 unigram data for both American and British English¹⁹. We found 147,205 (64.3%) of all German Wiktionary headwords in the DeReKo word form frequency list. Keep in mind that there are also articles for non-German words in the German Wiktionary. These words are most likely not included in the frequency list and are thus excluded from this analysis. The same holds for the English Wiktionary. There, we found 594,075 (24.7%) headwords in the frequency list.

All three variables of interest (number of revisions, number of visits, corpus frequency of the headword) are correlated²⁰, so we have to find a way to "disentangle" the relationships between them. One way to achieve this is to divide the data into a given number of equal sized portions. Previously, we divided the data into ten equally sized portions in terms of frequency (Müller-Spitzer et al. 2015: 16). In statistical terminology, we are looking at frequency deciles. We can then look at each frequency decile and concentrate on the other two variables we are interested in: number of visits and number of revisions. Frequency deciles are aligned on the *x*-axes in Figure 5 (English Wiktionary) and Figure 6 (German Wiktionary). On the *y*-axes, the mean number of visits for the articles in this frequency decile and the respective revision category (colour of bars) is recorded. The revision category of an article is defined by the percentile rank the respective article takes in relation to all other articles. The revision category "top 5%" contains all articles that are among the top 5% in terms of revisions (127,316 articles in the English Wiktionary; 11,509 articles in the German Wiktionary). Accordingly, the revisions category "top 10%" contains all articles that are among the top 10% articles in terms of revisions, excluding those already associated with the "top 5%" category — so this category should more precisely be named "top 5-10% of articles" (169,089 articles in the English Wiktionary; 11,656 articles in the German Wiktionary). All other articles fall into the category "bottom 90%" (2,106,436 articles in the English Wiktionary and 205,704 articles in the German Wiktionary).

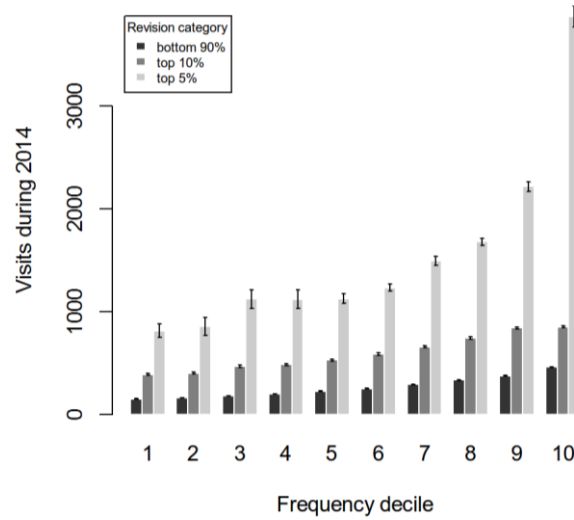


Figure 5: Visits to the English Wiktionary in 2014 (y-axis) in relation to corpus frequency decile of the headword (x-axis) and the revision category (colour of bar)

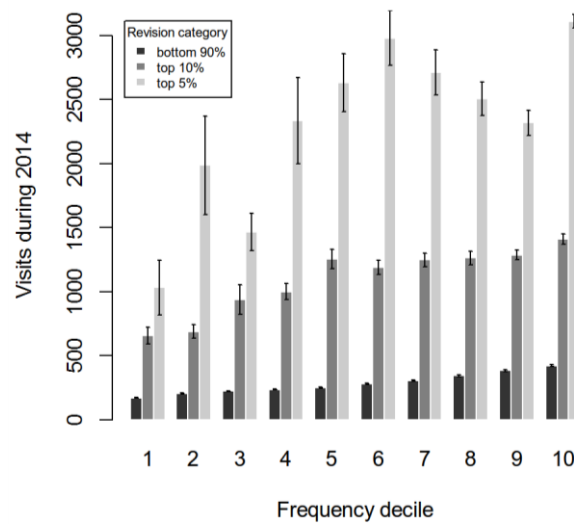


Figure 6: Visits to the German Wiktionary in 2014 (y-axis) in relation to corpus frequency decile of the headword (x-axis) and the revision category (colour of bar)

Both Figure 5 and Figure 6 exhibit a similar pattern. In all frequency deciles, articles that fall into the top 5% of articles in terms of revisions are also visited more often. This difference does not only hold true in comparison to the bottom 90%, but also in comparison to the 5% of articles below the top 5% (category name "top 10%"). The effect of corpus frequency on look-up frequency is visible for the English Wiktionary. Especially the top 5% of revised articles show a clear tendency to be visited much more often in the frequency deciles 6 to 10. This effect is not that clear in the German Wiktionary. While there is still an overall effect of corpus frequency on look-up frequency, this effect is not consistent for the top 5% of revised articles. However, the effect we are primarily interested in holds true: Articles that are revised more frequently are also looked up more frequently. One could object that an article page has to be accessed to be edited. So, each time someone revises an article, she or he also has to visit this article. It could be tested whether this fact is responsible for the observed pattern by subtracting the number of revisions in 2014 from the number of visits in 2014 and repeat the above analysis. We do not report the respective figures because they look almost the same as Figure 5 and Figure 6. The only visible change is in the absolute number of visits (y -axis). The overall pattern, however, remains stable.

5. Summary

Traditionally, dictionaries are compiled by publishers or by academic projects financed by the public. Such dictionaries are written by lexicographic experts. Wiktionary, as a collaborative, non-profit dictionary is based on voluntary work. The idea behind such collaborative joint activities is that the professionalism of some is substituted by the collective intelligence of many. In this paper, we attempted to evaluate the question whether there are really many people working on Wiktionary and whether it is therefore the right wording to speak of a 'wisdom of crowds' phenomenon.

We saw that the distribution of lexicographic effort, as operationalised by the number of revisions, is heavily biased for both authors and articles. Few authors do the majority of the work and the majority of revisions are distributed among few articles, with the latter distribution being less extreme. These distributions are also reflected by the bivariate distribution of diversity and rigour, where we saw that a few articles rank high on both scales while the vast majority of articles, however, score rather low on both scales. In terms of quality, this is problematic. In an ideal world, most articles would have high values for rigour and diversity. We would recommend that the Wiktionary community focuses their efforts on low-ranking articles in terms of rigour and diversity. Maybe, articles with very high numbers of visits would be a good starting point.

Although both author participation and article revisions are so unevenly distributed, most of the articles in Wiktionary are not very old in terms of the

most current revision. This might be interpreted as a good sign, but it also means that there does not seem to be an effective organisation of revisiting older articles to check whether they might need updating. Rather, we can observe certain points in time during which "revision sprints" seem to take place — both on the level of the whole dictionary, but also for individual articles.

Finally, we saw that frequency in language-use (as measured by corpus frequency of the headword), consultation frequency and revision frequency are heavily interdependent. The most frequent words in language are visited and revised more often than words that play a more marginal role in every-day language-use. This might be seen as a form of consumer-orientation (work on those words that people are looking up). However, this practice (which is most likely not a conscious one) is also associated with the risk of losing track of less frequent words.

On the basis of this data, there are, in our view, two possible answers to the question whether many people are working on Wiktionary:

No, there are not many people: Anyone involved in dictionary projects knows that 30 people working irregularly and part-time are not enough to write a good and comprehensive general dictionary. If a dictionary team were that small, either most of the sense disambiguation information, paraphrases, collocations and typical phrases would be copied from other lexicographic resources and only a small part of the vocabulary would be described from scratch, or only a small part of the vocabulary would be elaborated in an innovative way and, correspondingly, the dictionary would only contain this part of the vocabulary. As Wiktionary contains as many headwords as the big general dictionaries, the common strategy seems to be to integrate copyright-free data from older dictionaries and to complement them whenever necessary. From that point of view, Wiktionary is in no way comparable to Wikipedia. On the basis of the revision data, we can surely say that Wiktionary will not surpass previous lexicographic works on a wide basis in terms of comprehension of vocabulary, state-of-the-art semantic description or innovative forms of presentation. This would require a bigger crowd working on a dictionary. Accordingly, Wiktionary will not replace publishers' and academic dictionaries in terms of content and quality.

Yes, there are many people: When Wiktionary was launched there was considerable doubt that the project would attract any kind of community. The line of reasoning was that it clearly may be a pleasure to share own knowledge about a field of research, a kind of sport or a historical person with others and to use the opportunity to spread this knowledge via Wikipedia. However, who wants to be an expert, for example, on the *word* 'marmalade' or 'tree' and spend time elaborating dictionary entries? At most neologisms or special parts of the vocabulary seemed to be a promising field to attract voluntary work. These expectations were not fulfilled. The revision data clearly shows that 'normal' highly frequent words are being revised, not only neologisms etc. And there

are at least 30-50 people who regularly work on the English resp. the German Wiktionary.

What does that mean for the relation between Wiktionary and academic or publishers' lexicography? In our view, the best relation would be not that of competitors but of two actors in the field that benefit from cross-pollination of ideas and to see Wiktionary — from the point of academic and publishers' lexicography — as one more possibility to disseminate lexicographic data. However, we are very aware of the fact that this idea ignores the problem of financing lexicographic enterprises. If the user base of Wiktionary grows partially through the integration of established lexicographic work (see Rundell 2015, for elaboration on this issue) that was financed by others, it is hard to convince an enterprise or a public institution to keep playing the game. However, one fact is clear from our analyses of the revision data: We cannot expect Wiktionary to become a better dictionary on a wide basis than established dictionaries. In consequence, if no professional work is conducted on dictionaries, Wiktionary will be no long-term compensation. On the other hand, it is a pleasure to see that there is a language-interested community that works on dictionaries voluntarily. Is this not also a sign for the relevance of dictionaries?

Finally, we would like to stress that Wiktionary is a resource of major interest for usage research because all the lexicographic content, the user/author statistics and revision data is freely available at all times. It is very hard to receive such comprehensive data from any other (especially commercially oriented) dictionary projects.

6. Endnotes

1. However, there are also studies suggesting that "the number of active contributors in Wikipedia has been declining steadily for years" (Halfaker et al. 2013: 664). The authors of the study argue that "several changes the Wikipedia community made to manage quality and consistency in the face of massive growth in participation have ironically crippled the very growth they were designed to manage" (ibid.). Note that this refers to the contributors of Wikipedia and not to the recipients.
2. The English language edition of Wiktionary was visited approx. 980 million times in 2014 and 600 million times in 2015. The German language edition was visited approx. 156 million times in 2014 and 97 million times in 2015. Please note that a new page count definition came into effect after April 2015. What might look like a decrease from 2014 to 2015 is an effect of this new definition. Extensive statistics on all Wikimedia projects are available from <https://stats.wikimedia.org/wiktionary/EN/Sitemap.htm> [last accessed on 2016-02-04].
3. Note that the German Wiktionary is rank 6 in terms of number of revisions (4,710,263 revisions compared to 40,266,646 revisions in the English Wiktionary). This information is available from <https://meta.wikimedia.org/wiki/Wiktionary/Table> [last accessed on 2015-11-12].
4. In British English, rigor and rigour have different meanings. In BE, "rigour" has the meaning we want to imply, rigor has a medical meaning. In contrast to Lih (2004), we are using "rigour" to avoid confusion.

5. The most actual data dumps are available under <https://dumps.wikimedia.org/backup-index.html> [last accessed at 2016-06-08].
6. The list of users with the bot flag for the English Wiktionary is available under <https://en.wiktionary.org/wiki/Special:ListUsers/bot> [last accessed at 2015-11-16].
7. The list of users with the bot flag for the German Wiktionary is available under <https://de.wiktionary.org/w/index.php?title=Spezial%3ABenutzer&username=&group=bot> [last accessed at 2015-11-16].
8. Information taken from <http://public.oed.com/the-oed-today/staff-of-the-oxford-english-dictionary/> [last accessed on 2015-11-11].
9. Personal e-mail communication with a Duden editorial staff employee.
10. Percentages apply to non-automatic, non-minor revisions of non-meta pages.
11. Concerning this argument, there might be objections that we cannot be sure about registered authors, either. For example, the user account "Wamito" in the German Wiktionary is used by (at least) two persons as indicated on the associated user page. However, we are confident that this is an exception. We are not aware of any way to assess this automatically for all user accounts.
12. The top 1% of authors in the English Wiktionary consists of 369 authors and 61 authors in the German Wiktionary.
13. As an anonymous reviewer points out, this "sounds as [sic] an incredibly high number of revisions for one single user". However, the author's page in Wiktionary (<https://en.wiktionary.org/wiki/User:Equinox>, last accessed on 2016-05-23) does not suggest that she or he makes heavy use of automated revision processes.
14. The model comparison tests whether the residual sum of squares decreases significantly if the quadratic term is included into the predictor structure. Both F and X^2 tests detect highly significant improvements of model fits for both the English and German Wiktionary if the quadratic trend is included.
15. The median is the value which divides a distribution into two equally sized halves (i.e. there is an equal number of data points below (or equal) and above the median). Compared to the mean, the median value is more robust in presence of extreme values ("outliers").
16. Note that we excluded non-registered authors from the analysis which might be considered "unfair" in this context. Consequently, our measurements represent the lower bound of diversity and rigour values.
17. Note that "later" has to be interpreted relative to the creation date of the respective article. In Figure 3, $x = 0$ could be a different absolute point in time for each of the thousand articles we plotted.
18. There are revisions in the German Wiktionary history file that are over two years older than the "official" birthday on 2004-05-01. We decided to keep these earlier revisions because they also contribute to the current state of the German Wiktionary.
19. Google Books frequency data is available from <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> [last accessed on 2015-12-02].
20. Spearman correlations for the English Wiktionary are: corpus frequency vs. revisions: $\rho = .381$; corpus frequency vs. visits: $\rho = .477$; visits vs. revisions: $\rho = .273$. Spearman correlations for the German Wiktionary are: corpus frequency vs. revisions: $\rho = .313$; corpus frequency vs. visits: $\rho = .331$; visits vs. revisions: $\rho = .616$.

7. References

- Chesley, P., B. Vincent, L. Xu and R.K. Srihari.** 2006. Using Verbs and Adjectives to Automatically Classify Blog Sentiment. Nicolov, N., F. Salvetti, M. Liberman and J.H. Martin (Eds.). 2006. *Computational Approaches to Analyzing Weblogs: Papers from the 2006 Spring Symposium*: 27-29. (Technical Report SS-06-03.) Menlo Park, CA: AAAI Press.
- Greenstein, S. and F. Zhu.** 2012. *Collective Intelligence and Neutral Point of View: The Case of Wikipedia*. Working Paper, National Bureau of Economic Research. <http://www.nber.org/papers/w18167>.
- Halfaker, A., R.S. Geiger, J.T. Morgan and J. Riedl.** 2013. The Rise and Decline of an Open Collaboration System. How Wikipedia's Reaction to Popularity is Causing its Decline. *American Behavioral Scientist* 57(5): 664-688.
- Hanks, P.** 2012. Corpus Evidence and Electronic Lexicography. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 57-82. Oxford: Oxford University Press.
- Henrich, V., E. Hinrichs and Vodolazova, T.** 2011. Semi-Automatic Extension of GermaNet with Sense Definitions from Wiktionary. *Proceedings of the 5th Language and Technology Conference (LTC 2011), Poznań, Poland, November 25–27, 2011*: 126-130.
- Lang, D.T.** 2013. XML: Tools for Parsing and Generating XML within R and S-Plus. <http://CRAN.R-project.org/package=XML>.
- Lew, R.** 2014. User-Generated Content (UGC) in English Online Dictionaries. Abel, A. and A. Klosa (Eds.). 2014. *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*: 9-30. Mannheim: Institut für Deutsche Sprache. (OPAL — Online publizierte Arbeiten zur Linguistik 4/2014.
- Lih, A.** 2004. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Resource. *Proceedings of the 5th International Symposium on Online Journalism, April 16–17, 2004, University of Texas at Austin, USA*. <https://online.journalism.utexas.edu/papers.php?year=2004>.
- Lindemann, D.** 2014. Creating a German–Basque Electronic Dictionary for German Learners. *Lexikos* 24: 331-349.
- Medero, J. and M. Ostendorf.** 2009. Analysis of Vocabulary Difficulty Using Wiktionary. *Proceedings of the ISCA International Workshop on Speech and Language Technology in Education (SLaTE–2009)*. Warwickshire, England.
- Meyer, C.M. and I. Gurevych.** 2012. Wiktionary: A New Rival for Expert-built Lexicons? Exploring the Possibilities of Collaborative Lexicography. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 259-291. Oxford University Press.
- Müller-Spitzer, C., S. Wolfer and A. Kopleinig.** 2015. Observing Online Dictionary Users: Studies Using Wiktionary Log Files. *International Journal of Lexicography* 28(1): 1-26. doi:10.1093/ijl/ecu029.
- Muzny, G. and L. Zettlemoyer.** 2013. Automatic Idiom Identification in Wiktionary. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), 18–21 October 2013 Seattle, Washington, USA*: 1417-1421. Stroudsburg, USA: Association for Computational Linguistics.
- Navarro, E., F. Sajous, B. Gaume, L. Prévot, H. ShuKai, K. Tzu-Yi, P. Magistry and H. Chu-Ren.** 2009. Wiktionary and NLP: Improving Synonymy Networks. *ACL Workshop on The People's*

- Web Meets NLP: Collaboratively Constructed Semantic Resources*, August 2009, Singapore: 19-27. Stroudsburg, USA: Association for Computational Linguistics.
- Poderi, G.** 2009. Comparing Featured Article Groups and Revision Patterns Correlations in Wikipedia. *First Monday* 14(5). <http://firstmonday.org/ojs/index.php/fm/article/view/2365/2182>.
- R Core Team.** 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- 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. http://www.euralex.org/elx_proceedings/Euralex2012/pp47-92%20Rundell.pdf.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Schlippe, T., S. Ochs and T. Schultz.** 2010. Wiktionary as a Source for Automatic Pronunciation Extraction. Kobayashi, T., K. Hirose, S. Nakamura. 2010. *11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, Makuhari, Chiba, Japan, September 26–30, 2010: 2290-2293. International Speech Communication Association (ISCA). <http://dblp2.uni-trier.de/db/conf/interspeech/interspeech2010.html>
- Stein, K. and C. Hess.** 2008. Viele Autoren, gute Autoren? Eine Untersuchung ausgezeichneter Artikel in der deutschen Wikipedia. Alpar, P. and S. Blaschke (Eds.). 2008. *Web 2.0: Eine empirische Bestandsaufnahme*: 107-129. Wiesbaden: Vieweg+Teubner.
- Stvilia, B., M.B. Twidale, L.C. Smith and L. Gasser.** 2008. Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology* 59(6): 983-1001.
- Surowiecki, J.** 2005. *The Wisdom of Crowds*. New York: Anchor Books.
- Wilkinson, D.M. and B.A. Huberman.** 2007. Cooperation and Quality in Wikipedia. *WikiSym '07: Proceedings of the 2007 International Symposium on Wikis*, Montreal, Canada, October 21–23, 2007: 157-164. New York: ACM Press.
- Zesch, T., C. Müller and I. Gurevych.** 2008. Using Wiktionary for Computing Semantic Relatedness. Fox, D. and C.P. Gomes (Eds.). 2008. *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, Chicago, Illinois, July 13–17, 2008: 861-867. Menlo Park, California: AAAI Press.

Comparing Dictionary-induced Vocabulary Learning and Inferencing in the Context of Reading

Di Zou, *English Language Centre, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (dizoudaisy@gmail.com)*

Abstract: This research examines dictionary-induced vocabulary learning and inferencing in the context of reading. One hundred and four intermediate English learners completed one of two word-focused tasks: reading comprehension and dictionary consultation, and reading comprehension and inferencing. In addition to performing the tasks, some subjects reported their thinking processes either during or after the completion of the tasks, and those who did not were tested both immediately and one week later for their learning of target words. The results show that dictionary-induced vocabulary learning was significantly more effective than inferencing. The researcher explains such results in terms of theories of the degree of elaboration and connectionist models, and suggests that the provision of a number of various aspects of knowledge about a target word is very facilitative for word learning.

Keywords: DICTIONARY, DICTIONARY-INDUCED VOCABULARY LEARNING, INFERENCE, WORD LEARNING, ELABORATION, CONNECTIONIST, CONNECTIONS, TASK, WORD KNOWLEDGE, EFFECTIVENESS

Opsomming: Vergelyking van woordeboek-gemotiveerde aanleer en -afleiding van woordeskat in die leeskonteks. Hierdie navorsing ondersoek woordeboek-gemotiveerde aanleer en -afleiding van woordeskat in die leeskonteks. Eenhonderd-en-vier intermediaire Engels leerders het een van twee woordgefokusde opdragte voltooi: leesbegrip en woordeboekkonsultasie, en leesbegrip en betekenisafleiding. Buiten die uitvoer van die take, het party proefkonlyne ook tydens of ná voltooiing van die take verslag gedoen van hul denkprosesse, en dié wat dit nie gedoen het nie is onmiddellik, asook een week later, getoets vir die aanleer van tekenwoorde. Die resultate toon dat woordeboek-gemotiveerde aanleer van woordeskat beduidend effektiewer was as betekenisafleiding. Die navorser verduidelik hierdie resultate aan die hand van teorieë van die graad-van-uitbreidingsmodel en die konneksionistiese model, en stel voor dat die voorsiening van 'n aantal verskillende kennisaspekte van 'n trefwoord die aanleer van woorde baie vergemaklik.

Sleutelwoorde: WOORDEBOEK, WOORDEBOEK-GEMOTIVEERDE AANLEER VAN WOORDESKAT, BETEKENISAFLEIDING, AANLEER VAN WOORDE, UITBREIDING, KONNEKSIONISME, KONNEKSIES, OPDRAG, WOORDKENNIS, EFFEKTIEWEIT

1. Introduction

As vocabulary is one of the crucial components of a language, an increasing amount of research on second language acquisition has concentrated on vocabulary acquisition. Word knowledge generally involves knowledge of the form, meaning and use, and dictionary consultation and inferencing are two of the most frequently investigated approaches to the learning of word meaning (Nation 2001).

1.1 Dictionary-induced Vocabulary Learning

A great deal of controversy exists over the facilitative effects of dictionary consultation. On the one hand, some researchers argue that using a dictionary while reading leads to inefficient learning (Chan 2011), or extra time spent on looking up words was unlikely to be used efficiently (Knight 1994). It does not always improve comprehension (Koyama and Takeuchi 2004); or looking up too many words without careful consideration led to shallow processing and short-term retention (De Ridder 2002). Prichard's (2008) explanations for such results are that dictionary use was facilitative for less proficient learners because it can help them reach the comprehension threshold; proficient learners, however, already knew enough words in the passages to enable comprehension, so the benefits were reduced. After examining dictionary assistance for learners' identification and retention of collocations, Laufer (2011) concluded that learners sometimes had challenge of identifying the correct verbs, but often they regarded dictionary consultation unnecessary.

On the other hand, many researchers believe that dictionary consultation assists reading comprehension and promotes word knowledge development. Luppescu and Day (1993), Hulstijn, Hollander and Greidanus (1996), Abraham (2008), Dziemianko (2010), Chen (2011), Liu and Lin (2011), Chiu and Liu (2013), and Nation (2015) all found significant facilitative effects of dictionary consultation for word learning. Dictionary use is important for learning the correct use (Chan 2012), meanings (Lew 2012, Coffey 2011), and spellings of target words (Liu et al. 2014). It is beneficial to learners whose language proficiency is neither too high nor too low (Prichard 2008).

1.2 Inferencing

Inferencing, which asks learners to guess the meaning of an unknown word according to its context, is closely associated with vocabulary learning (Nassaji 2004) and has its basis in the availability of linguistic and non-linguistic cues in the text (Hu and Nassaji 2012). The facilitative effects of inferencing on word learning have been noted by many previous studies, for example, Hulstijn (1992), Nation (2001), Pulido (2009), Nassaji and Hu (2012), Carpenter et al.

(2012), and Webb and Chang (2015). Huang and Lin (2014) examined three approaches to glossing and also found that the inference-gloss-gloss condition was more effective than full glossing condition. Nevertheless, Sadoski, Goetz and Rodriguez (2000) argued that although inferencing led to good immediate recall of word meanings, it promoted little retention. Mondria (2003) also pointed out that inferencing was considerably more time-consuming but did not lead to better learning than the meaning-given method.

Compared to dictionary-induced vocabulary learning, inferencing is believed to be less effective in promoting word learning (e.g. Knight 1994, Liu et al. 2014, Nation 2015). However, the involvement load hypothesis, which argues that the retention of words is conditional upon three factors in a task: need, search and evaluation, simply defines search as the attempt to find the meaning or form of target words and allocates the same involvement load to various approaches to it (Laufer and Hulstijn 2001). It is therefore necessary to conduct a direct test on the effectiveness of different approaches to search from the perspective of the involvement load hypothesis, the results of which may shed light into further development of this hypothesis.

1.3 Objectives and Research Questions of this Research

This research aims at two objectives: to examine whether dictionary-induced vocabulary learning and inferencing would promote initial learning and retention of target words with similar effectiveness, and to inspect how they would differ from each other with regard to the processing of target words by learners.

Two questions are raised accordingly. (1) Does reading comprehension and dictionary consultation promote word learning with similar effectiveness as reading comprehension and inferencing? (2) What are the reasons for any similarity or difference detected?

2. Method

2.1 Tasks

The present research investigated two tasks. Task 1: reading comprehension and dictionary consultation asked the subjects to read a text, look up the ten underlined words in the dictionary and complete ten multiple-choice questions. The *Longman Dictionary of Contemporary English, 5th Edition (2009)* was used because firstly it is a monolingual dictionary and meets the requirement of eliminating the influence of translation from Chinese in the task, and secondly the participants in this research noted that it was one of their most commonly used dictionaries.

Task 2: Reading comprehension and inferencing, as demonstrated in Appendix 1, required the subjects to read the same text, infer the meanings of the ten underlined words according to their contexts, write down the inferred meanings in the margin of the text, and complete three multiple-choice questions. These questions were different from those for Task 1 in that they were designed to inquire about the general gist of the whole text, rather than specifically require knowledge of the target words, because the inferencing exercises had already imposed the search for the meanings of those words.

Each of the participants in this research was randomly asked to complete one of the two tasks and strictly follow the requirements. As the task completion time was inherent to the nature of the tasks, I did not attempt to make the completion time the same for different learners doing different tasks so as to avoid deliberate manipulation of the completion time for the two tasks. Also, although these two tasks are different in respect to the multiple-choice questions and the way the target words were processed by the participants, these differences relate closely to and result from the diverse nature of the two approaches to search that are induced by the two tasks. The main significant difference between the two tasks, therefore, can be regarded as the two approaches to search: dictionary consultation and inferencing.

2.2 Subjects

A total of one hundred and four non-English major freshmen at Tsinghua University participated in the research. All of them had learnt English for at least ten years and had obtained scores ranging from 425 to 450 out of 710 in a Chinese national English proficiency exam named College English Examination Band 4 three months before participating in this study. Scores of 450 in this exam are approximately equivalent to a 5.5 score for IELTS.

2.3 Procedure

During task completion, eight subjects doing Task 1 and six doing Task 2 were randomly selected to answer introspective questionnaires to reflect their thinking process while doing their assigned tasks (see Table 1). Also, six subjects doing Task 1 and eight doing Task 2 were selected to do think-aloud to record how they tackled the tasks. These participants were allowed to report their thinking process either in English (the target language) or Chinese (their first language).

After task completion, six subjects doing Task 1 and six doing Task 2 were randomly selected to take part in retrospective interviews about their thinking processes during task completion. 31 subjects doing Task 1 and 33 doing Task 2 were tested immediately to measure their initial learning of the target words. One week later, the subjects were unexpectedly tested again to examine their

retention of the target words in a delayed posttest. During these posttests, the subjects were not allowed to refer to the task materials they had just been working on.

		Task 1: Dictionary consultation	Task 2: Inferencing
During task	Introspective questionnaires	8	6
	Think-aloud protocols	6	8
After task	Retrospective interviews	6	6
	Posttests	31	33

Table 1: Allocation of subjects to tasks

2.4 Materials

Materials for the two tasks include mainly the reading text and the target words. The text "Coping with Procrastination" which was also used in Zou's (2016) study was selected as the original text as experience in procrastination was likely to be similarly familiar to all participants. Working within the context of this text, ten target words (*divulge, renege, taunt, lassitude, trait, apprehensive, assiduous, indispensable, ostensible and pernicious*) were selected. Words with both concrete and abstract meanings were included, with the purpose of removing any possible concreteness effect (Salsbury et al. 2011, Mestres-Missé et al. 2014). Also, the words selected had only a single meaning for the part of speech in which it occurred in the reading text.

A pretest, which asked the 104 participants to write down meanings of the target words and generate original sentences using these words, was conducted three weeks before the experiment, the results of which showed that all participants had very little pre-knowledge of the target words.

2.5 Assessment and Scoring

In the immediate and delayed posttests, the participants were provided with a list of the target words and asked to write down the meanings of the words, either in English or Chinese, and generate original English sentences using them. The scoring criteria of the answers to the meanings of the target words had its foundation on those developed by Hulstijn and Laufer (2001) and Keating (2008). (1) A score of 0 was given to an answer with a completely incorrect meaning (e.g. the answer "last long" for the meaning "lassitude"). (2) A half score of 0.5 was given to a semantically acceptable equivalent of the target word (e.g. "negative" for "pernicious"). And (3) a full score of 1 was given to an

answer with a comparable meaning to that of the target word (e.g. "reveal" for "divulge").

Regarding the sentences, the criteria used in the studies of Paribakht and Wesche (1993), Hulstijn and Laufer (2001), and Zou (2016) were employed. (1) A score of 0 was given to a sentence that had a completely inappropriate semantic context for the target word (e.g. "I'll never forget how badly he traited me."). (2) A half score of 0.5 was given to a sentence that had an appropriate semantic context for the target word but the target word was used ungrammatically (e.g. "The students who have let days off should now apprehensive about their scores of this exam."). And (3) a full score of 1 was given to a sentence with an appropriate semantic context for the target word and the target word was used grammatically (e.g. "Her contribution is indispensable for the success of this project.").

The subjects' answers in both posttests were blindly and separately scored by the author and another trained rater. In cases where it was difficult to decide on an appropriate score, the two raters negotiated the score in collaboration until consensus was reached. Such cases were, however, very rare (fewer than 2%). Pearson's r for the inter-rater reliability was .94 for the immediate posttest and .92 for the delayed posttest.

2.6 Analysis of the Self-reporting Data

All self-reporting data were typed up and transcribed into Word documents. As the subjects answered questionnaires in either Chinese or English, and the two self-reporting protocols were conducted in either language, the transcriptions were correspondingly in either language. The Chinese transcriptions were also translated into English for reporting. This stage of data transcription assisted the researcher to not only gain a holistic view of all qualitative data collected but also identify distinct features of the data.

The next stage of data coding involved three main steps: (1) skimming the questionnaire answers and transcriptions to note the elements that the subjects found facilitative for word learning; (2) reviewing relevant literature to identify theories that provided explanations for the facilitative elements; and (3) re-examining the data and paying close attention to the types of thinking processes involved so as to identify reasons for the facilitative effects of these thinking processes and the extent to which these processes promoted word learning. Regarding the thinking patterns of the subjects, I read the transcripts very carefully to identify any expressions that were noted and repeated by different participants in similar contexts.

3. Results

The descriptive statistics for the two posttests of the 64 subjects, as displayed in

the following Table 2, demonstrated that the mean scores of subjects doing Task 1: reading comprehension and dictionary consultation was 12.11 in the immediate posttest and 8.98 in the delayed posttest. And the mean scores of subjects doing Task 2: reading comprehension and inferencing was 7.91 in the immediate posttest and 5.45 in the delayed posttest. The results also showed that subjects obtained great gains of knowledge about the target words after doing these two tasks, given that their pre-knowledge of these words was almost zero.

	N	Pretest		Immediate posttest		Delayed posttest	
		M	SD	M	SD	M	SD
Task 1: Dictionary consultation	31	0.16	0.37	12.11	4.22	8.98	4.99
Task 2: Inferencing	33	0.18	0.39	7.91	4.70	5.45	3.95

Table 2: Descriptive statistics of the subjects' scores in the pretest, immediate and delayed posttests

To further test whether dictionary consultation was significantly more effective than inferencing, two independent samples tests were applied, following the standard practices of Hulstijn and Laufer (2001), Keating (2008), and Eckerth and Tavakoli (2012). Statistically significant differences were identified both in the Welch test results concerning the immediate posttest ($t = 3.75, df = 62, p < .01, Cohen's d = .94$) and in those concerning the delayed posttest ($t = 3.14, df = 62, p < .01, Cohen's d = .79$).

4. Discussion

With an attempt to figure out how dictionary-induced vocabulary learning differs from inferencing, the thinking processes of the subjects doing these two tasks were analyzed from two perspectives: the diverse degrees of elaboration entailed by the two tasks and the diverse complexity of connections that make up the networks of target words.

4.1 Degree of Elaboration

Depth of processing examines different types or levels of information processing from a qualitative perspective, arguing that the chance of a piece of infor-

mation being stored in long-term memory depends on the depth with which it is initially processed rather than the length of time that it is held in short-term memory (Craik and Lockhart 1972). Elaboration of encoding, within each qualitative type or any one level of processing depth, investigates information processing from a quantitative perspective (Craik and Tulving 1975). It is the degree to which each type of processing has been enriched during encoding, and refers to the richness or extensiveness of information processing (Craik 2002).

Both dictionary consultation and inferencing involve semantic processing of information about target words, as both entail attempts to find out diverse information of a target word, thus they should be at the same level of processing depth. However, they differ in their degrees of elaboration in that dictionaries provide richer information about target words and hence dictionary-induced vocabulary learning is likely to entail a greater degree of elaboration than inferencing, when a learner consults different features of a word in the dictionary. Empirical support for this assessment can be found in the subjects' answers to the introspective questionnaires, transcripts of think-aloud protocols and retrospective interviews, extracts of which will be provided accordingly as follows.

Dictionaries always provide various aspects of knowledge about words, including such information as pronunciation, part of speech, definition, sample sentences, usage labels (e.g. formal, informal, taboo, etc.), sense relations, and so on. With the advantage of being provided with more information, most participants, who did reading comprehension and dictionary consultation in this experiment, made full use of the available information and processed as much as possible in order to facilitate their comprehension of the target words and the reading text. The following think-aloud transcripts of two subjects doing reading comprehension and dictionary consultation demonstrate that in searching a word's meaning through dictionary consultation, the subjects paid close attention to different aspects of information about the target words, such as its form, spelling, pronunciation, part of speech, definition, and sample sentences:

Or are you picturing that others taunt ... t-a-u-n-t, a-u-n, a-u-n-t ... taunt ... verb, VN, nontransitive ... to try to make somebody angry or upset by saying unkind things about them, laughing at their failures, The other kids continually taunted him about his size ... Taunt ... Or are you picturing that others taunt you about your accent?

You might delay preparing for an oral presentation, because you are apprehensive ... a-p-p-r-e ... apprehensive ... about, of something, that ... worried or frightened that something unpleasant may happen. I was a little apprehensive about the effects of what I had said... You have no reason to be apprehensive of the future ... similar to worried ... apprehensive, worried ... because you are apprehensive that you will not be able to remember the entire speech.

Illustrations like these can be found in many other think-aloud transcripts of subjects doing reading comprehension and dictionary consultation (see Zou, 2012, for details). These findings may be different from the observations of some researchers who noted that dictionary users were selective in what they read. However, it is worth pointing out that the majority of the participants in their studies were not Chinese learners of English. Studies that were conducted among Chinese learners, nevertheless, found that diverse aspects of knowledge of target words were consulted by the participants. For example, Chen (2010) reported consultation of pronunciation; Chen (2011) and Chan (2012) noticed consultation of usage information.

In contrast to the above findings concerning dictionary consultation, learners who inferred a word's meaning based on a reading text normally have no access to many aspects of knowledge about target words. Beyond inferring the grammatical category and possible meaning of a target word in the given context, it is unlikely that they can acquire any definite knowledge about the words' pronunciation, derivatives, or other collocational possibilities other than that in the sole example in the reading text. It is clearly showed, in the interview transcripts of the subjects who did reading comprehension and inferencing, that they were only able to infer a word's meaning using the limited context provided by the reading text, and that inferring any information other than the meaning was difficult if not impossible. The following transcript is one piece of evidence:

I don't know how the words should be pronounced or used, because such information is not given. I even haven't thought about this, because it is not demanded by the task. I do not know anything else about these words other than their meanings in the given contexts, because nothing else is provided. I haven't thought how the words can be pronounced or used with other words in other contexts. It's very hard for me to know, or infer, such information based on the limited information given.

Supporting evidence of the assessment that subjects who did dictionary consultation processed more information about the target words than those who did inferencing can also be found in the subjects' answers to the introspective questionnaire for Task 1: reading comprehension and dictionary consultation. The following Table 3 demonstrates the percentages of various types of information that were consulted by the eight subjects who answered the questionnaires while consulting the dictionary. The percentage was calculated by dividing the sum of items being consulted by the sum of items provided in the dictionary. For example, a total of 10 definitions for the ten target words were provided, so the sum is 80 for the eight participants (10×8). As the results of the questionnaires showed that all 80 definitions had been consulted, the percentage of definitions being consulted was thus 100% ($80/80$). Similarly, as a total of 40 synonyms were provided and 38 of them had been consulted, the percentage of synonyms being consulted was 95% ($38/40$).

	Percentage of information being consulted
Definition	100% (80/80)
Synonym	95% (38/40)
Part of speech	84% (67/80)
Example sentence or phrase	83% (60/72)
Derivatives	75% (24/32)
Pronunciation	74% (59/80)
Usage information	58% (46/80)
Others	20% (16/80)

Table 3: Percentages of information being consulted

In summary, the percentages of different types of information being consulted were respectively 100% for definitions, 95% for synonyms, 84% for parts of speech, 83% for example sentences or phrases, 75% for derivatives, and 74% for pronunciation (see Table 3). Only the rate of usage information was slightly lower than 60%. It is therefore evident that overall, a substantial amount and range of dictionary information were used by subjects doing reading comprehension and dictionary consultation. This is conducive to word learning although it may be due to reactivity effects.

In addition to the above arguments, another significant facilitative effect on word learning of the processing of rich information about target words, which can also be referred to as extensive elaboration on target words, is its assistance with the comprehension of words meanings. Supporting evidence for this can be found in the interview transcripts of the subjects who did reading comprehension and dictionary consultation. They show that information like synonyms, example phrases, sentences and collocations all contributed to their comprehension of word meanings. When subjects had difficulty in thoroughly understanding the definitions given by the dictionary, the processing of information additional to the definition was helpful in respect to the provision of valuable clues. Many subjects were acutely aware of the assistance provided by the rich information about the target words made available to them through the dictionary entries. Supporting evidence of these arguments can be found in the following transcripts:

I find synonyms for target words very useful. Synonyms can help me to grasp the most important feature of a word's meaning very quickly. Um, and they are very short, just one word, so are easy to be remembered. Take 'indispensable' as

an example. I understood and memorized its meaning the moment I noticed its synonym 'meticulous'.

I consulted the example phrases because I still felt uncertain about what 'trait' exactly means after reading its definition. So I ... I read the example phrases and their definitions so as to better grasp the word's meaning, and this helped my comprehension. The phrases drew my attention on the word 'personality', and helped me to realize that 'trait' is closely related to 'personality'.

I consulted the example sentences because they provide several different contexts in which the target words can be used. This can give me a more comprehensive understanding of the words.

In contrast, subjects who inferred word meanings were given quite a limited range of information for processing, and the available information was basically restricted to linguistic clues in the text and non-linguistic clues from their personal world knowledge. They felt handicapped by the insufficiency of useful information and often found it difficult to get an exact idea about the meaning of a target word. Moreover, since they were not given any specific information about the word's usage, nor were required to pay attention to this aspect, they did not attempt to obtain any information of this type. Supportive evidence of these can be identified in the following transcripts:

I think pernicious refers to something negative. Because here the writer explains that if you procrastinate, you will not have time to do satisfying work, and this will make other people unhappy and get a bad impression of you. And it can even damage your friendships. All of these talk about something bad ... But I am not sure about its meaning ... I can only say that pernicious means something negative.

I am not sure whether I really know what ostensible means. The context here indicates something relative to 'excuse' because all the reasons listed here are actually excuses. But I feel that it is very difficult to identify the exact meaning. I think I need more information.

I did not think much about the words' usage. I spent all my effort inferring the words' meaning, and did not bother to think how the words can be used because it was not required, and I thought ... I don't have relative information about word usage.

4.2 Connectionist Perspective

Another issue that can help explain why dictionary-induced vocabulary learning promotes better word learning than inferencing comes from the cognitive science perspective of connectionism. From this perspective, the complexity of connections that make up the networks of the target words are beneficial to learning (Moonen et al. 2006), and the rich information of dictionary entries for target words can offer complexity of connections when multiple aspects of

knowledge are consulted, whereas the insufficiency of information availability of lexical inferencing is inferior in this respect.

A network consisting of connections (features) among units, in the literature on connectionist models of second language acquisition, refers to the concept of knowledge (Ellis 2003). The term feature is defined in the connectionism literature as a specific connection among units or, in cognitive terminology, a specific pattern of neural activation over a number of neurons (Westhoff 2004, Moonen et al. 2006, Moonen et al. 2014). Every word can be regarded as a union of various features that determine it in a unique way (Zareva and Wolter 2012). And networks involving numerous and varied types of features are more likely to become activated (Moonen et al. 2006). In other words, networks that consist of a greater variety of features concerning target words, such as morphological, sensory or pragmatic, are more likely to be activated than those made up of fewer features. Moreover, the number and variety of features of target words, which reflect the complexity of connections that make up the networks of target words, play a significant role in promoting the learning of target words (Moonen et al. 2006, Moonen et al. 2014).

As the mental actions of dictionary consultation involve more, and more varied, feature categories of the target words than those of inferencing, the networks of the target words constructed by the subjects doing reading comprehension and dictionary consultation consist of more different types of features and are thus more easily activated than the networks constructed by the subjects doing reading comprehension and inferencing. Dictionary consultation thus led to better learning of the target words than inferencing. The following interview transcript of a subject doing reading comprehension and dictionary consultation, for example, demonstrates clearly that the availability of different types of features about a target word (pronunciation, spelling and meaning) provided the subject with the ability to identify the connections among different information units. These complex connections also worked to stimulate the activation of networks of this target word:

I am able to give the meaning of apprehensive because when I consulted its dictionary entry, I noticed that its pronunciation and the pronunciation of its first three letters 'a-p-p' sounded similar to the word 'haipa (worried)' in Chinese, and the word haipa (worried) is given as part of the definition for apprehensive. Thus, I thought about 'haipa (worried)' the moment I pronounced the word 'apprehensive'.

Many pieces of evidence like this can be identified from the interview transcripts of subjects doing reading comprehension and dictionary consultation (see Zou 2012, for details). Furthermore, the availability of different types of features about the target words through dictionary consultation not only gave subjects the ability to identify the connections among different information units, but also assisted in their making of such connections easy to activate. The following interview transcript of another subject illustrates that a network of

rich connections was easy to activate:

I could give the meaning of 'ostensible' because I remembered clearly the phrase 'ostensible reasons'. The reading text used this phrase, and it is also an example phrase given by the dictionary entry. The example sentence is about 'ostensible reason', too. Um, I've encountered it three times, so it is easy for me to recall it.

It is therefore clear that more complex connections were involved in the networks of target words constructed by the subjects who did reading comprehension and dictionary consultation, and that these networks had a greater chance of being activated than those constructed by the subjects who did reading comprehension and inferencing, resulting in a higher degree of task effectiveness of the former than the latter.

In conclusion, as the mental actions that are induced by reading comprehension and dictionary consultation involve more, and more varied, feature categories of the target words than those that are induced by reading comprehension and inferencing, the networks of the target words constructed by the subjects who did reading comprehension and dictionary consultation consist of more different types of features and are thus more easily activated than the networks constructed by the subjects who did reading comprehension and inferencing. Subjects who consulted the dictionary, therefore, had better learning of the target words than those who inferred the meanings. Given such differences, I suggest a change of the involvement load allocated to these two approaches to search. Search induced by dictionary consultation is strong as it entails the processing of multiple categories of word knowledge and involves great complexity of connections that make up the network of target words. Search induced by inferencing is moderate as it induces a lesser degree of elaboration and involves a more limited complexity of connections.

5. Implications

The results of this study demonstrate that the provision of a number of various aspects of knowledge about a target word is very facilitative for word learning. This is inferred from the finding that the reading comprehension and dictionary consultation task was significantly more effective than the reading comprehension and inferencing task because the former provided subjects with more knowledge aspects of the target words. It was the provision of various knowledge aspects that gave subjects a greater degree of elaboration for the target words, and it was noted that rich information offered better chances for subjects to build up networks of target words as a greater complexity of connections was available.

Providing learners with various aspects of word knowledge is crucial to word learning. It follows that language learners should be given access to various knowledge aspects of target words so as to assist them in processing those words from diverse dimensions and making use of more complex connections

to build up networks regarding them. Self-learners are advised to attempt to obtain a variety of aspects of word knowledge they wish to learn rather than concentrating only on a sole aspect such as meaning. Access to a variety of information for individual words is typically provided in learner's dictionaries, so learners are suggested to make more use of dictionaries in their self-learning. To better promote this, dictionary training is also essential, as realization of the benefits of dictionary use and proficient dictionary skills can help users fully utilize the rich information provided in dictionaries.

Teachers are advised to provide learners with various aspects of knowledge about target words when inferencing exercises are done. The very nature of inferencing exercises means that incorrect meanings can be arrived at by a student, and that even if the correct meaning is inferred, a student may not be certain that it is indeed correct. Classroom feedback for inferencing exercises is typically constrained to the context in which the word is encountered in the exercise, leaving aside further knowledge aspects beyond that specific context, such as pronunciation, typical collocations, or how it can be used in other contexts. Thus, it is suggested that when inferencing exercises are done in class, teachers should also provide various kinds of information about the target words to students, rather than only providing information relevant to the context of the exercise.

6. Conclusion

The present research compares the effectiveness of two word-focused tasks: reading comprehension and dictionary consultation, and reading comprehension and inferencing. The results show that dictionary-induced vocabulary learning was significantly more effective than inferencing, based on which the involvement load hypothesis is suggested to be further developed in terms of its allocation of involvement load to different approaches to search.

Limitations of this research include the scope of study, the limited number of approaches to search investigated and the possible influence of experimental conditions. The two tasks studied here are also different to some extent and the text context may be not conducive to inferring the meanings of the target words. Moreover, the completion time of the two tasks are different, around 36 minutes for the reading comprehension and dictionary consultation and almost 29 minutes for the reading comprehension and inferencing. However, the influence of time on task effectiveness is uncertain as it is not the focus of the current research. Also, some previous studies showed negative correlation between time and word learning (e.g. Erten and Tekin 2008, Nakata 2008), some found positive correlation (e.g. Folse 2006, Keating 2008), yet others believed that time had no effect on task effectiveness (e.g. Chen 2002, Hill and Laufer 2003). It is suggested that future work take these limitations into account so as to improve the generalizability of results.

Acknowledgements

I would like to thank Dr. Alice Chan, Dr. James Lambert and Dr. Haoran Xie for their invaluable help. Special gratitude also goes to the Editor, Danie Prinsloo, and the reviewers of Lexikos for their insightful suggestions for improvement. This article is based on part of my Ph.D. dissertation, which was submitted to City University of Hong Kong in 2012.

References

- Abraham, Lee.** 2008. Computer-mediated Glosses in Second Language Reading Comprehension and Vocabulary Learning: A Meta-analysis. *Computer Assisted Language Learning* 21(3): 199-226.
- Carpenter, Shana, Riebana Sachs, Beth Martin, Kristian Schmidt and Ruxandra Looft.** 2012. Learning New Vocabulary in German: The Effects of Inferring Word Meanings, Type of Feedback, and Time of Test. *Psychonomic Bulletin & Review* 19(1): 81-86.
- Chan, Alice Yin Wa.** 2011. The Use of a Monolingual Dictionary for Meaning Determination by Advanced Cantonese ESL Learners in Hong Kong. *Applied Linguistics* first published online November 8, 2011 doi:10.1093/applin/amr038. (*Applied Linguistics* 33(2): 115-140.)
- Chan, Alice Yin Wa.** 2012. Cantonese ESL Learners' Use of Grammatical Information in a Monolingual Dictionary for Determining the Correct Use of a Target Word. *International Journal of Lexicography* 25(1): 68-94.
- Chen, Hongxia.** 2002. *Investigating the Effects of L1 and L2 Glosses on Foreign Language Reading Comprehension and Vocabulary Retention.* Paper presented at the annual meeting of the Computer-Assisted Language Instruction Consortium, Davis, CA.
- 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.
- Chen, Yuzhen.** 2011. Dictionary Use and Vocabulary Learning in the Context of Reading. *International Journal of Lexicography Int J Lexicography* first published online December 2, 2011 doi:10.1093/ijl/ecr031 (*International Journal of Lexicography* 25(2): 216-247.)
- Chiu, Li-Ling and Gi-Zen Liu.** 2013. Effects of Printed, Pocket Electronic, and Online Dictionaries on High School Students' English Vocabulary Retention. *The Asia-Pacific Education Researcher* 22(4): 619-634.
- Coffey, Stephen.** 2011. A New Pedagogical Dictionary of English Collocations. *International Journal of Lexicography* 24(3): 328-341.
- Craik, Fergus I.M.** 2002. Levels of Processing: Past, Present ... and Future? *Memory* 10(5-6): 305-318.
- Craik, Fergus I.M. and Robert S. Lockhart.** 1972. Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior* 11(6): 671-684.
- Craik, Fergus I.M. and Endel Tulving.** 1975. Depth of Processing and the Retention of Words in Episodic Memory. *Journal of Experimental Psychology: General* 104(3): 268-294.
- De Ridder, Isabelle.** 2002. Visible or Invisible Links: Does the Highlighting of Hyperlinks Affect Incidental Vocabulary Learning, Text Comprehension, and the Reading Process? *Language Learning and Technology* 6(1): 123-146.

- Dziemianko, Anna.** 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.
- Eckerth, Johannes and Parveneh Tavakoli.** 2012. The Effects of Word Exposure Frequency and Elaboration of Word Processing on Incidental L2 Vocabulary Acquisition through Reading. *Language Teaching Research* 16(2): 227-252.
- Erten, İsmail Hakkı and Mustafa Tekin.** 2008. Effects on Vocabulary Acquisition of Presenting New Words in Semantic Sets versus Semantically Unrelated Sets. *System* 36(3): 407-422.
- Ellis, Nick.** 2003. Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. Doughty, C. and M.H. Long (Eds.). 2003. *Handbook of Second Language Acquisition*: 33-68. Oxford: Blackwell.
- Folse, Keith.** 2006. The Effect of Type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly* 40(2): 273-293.
- Hill, Monica and Batia Laufer.** 2003. Type of Task, Time-on-task and Electronic Dictionaries in Incidental Vocabulary Acquisition. *International Review of Applied Linguistics* 41(2): 87-106.
- Hu, Hsueh-chao Marcella and Hossein Nassaji.** 2012. Ease of Inferencing, Learner Inferential Strategies, and their Relationship with the Retention of Word Meanings Inferred from Context. *Canadian Modern Language Review* 68(1): 54-77.
- Huang, Lo-li and Chih-cheng Lin.** 2014. Three Approaches to Glossing and their Effects on Vocabulary Learning. *System* 44(1): 127-136.
- Hulstijn, Jan.** 1992. Retention of Inferred and Given Word Meanings: Experiments in Incidental Vocabulary Learning. Arnaud, P. and H. Béjoint. 1992. *Vocabulary and Applied Linguistics*: 113-125. London: Macmillan.
- Hulstijn, Jan and Batia Laufer.** 2001. Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning* 51(3): 539-558.
- Hulstijn, Jan, Merel Hollander and Tine Greidanus.** 1996. Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words. *The Modern Language Journal* 80(3): 327-339.
- Keating, Gregory.** 2008. Task Effectiveness and Word Learning in a Second Language: The Involvement Load Hypothesis on Trial. *Language Teaching Research* 12(3): 365-386.
- Knight, Susan.** 1994. Dictionary Use while Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. *Modern Language Journal* 78(3): 285-299.
- Koyama, Toshiko and Osamu Takeuchi.** 2004. How Look Up Frequency Affects EFL Learning: An Empirical Study on the Use of Handheld-electronic Dictionaries. Meng Chan, W., K. Nyet Chin, P. Martin-Lau and T. Suthiwan (Eds). 2004. *Proceedings of the CLaSIC 2004 Conference: Current Perspectives and Future Directions in Foreign Language Teaching and Learning, 1-3 December 2004, Centre for Language Studies, National University of Singapore*: 1018-1024. Singapore: Centre for Language Studies, National University of Singapore.
- Laufer, Batia.** 2011. The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. *International Journal of Lexicography* 24(1): 29-49.
- Laufer, Batia and Jan Hulstijn.** 2001. Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-induced Involvement. *Applied Linguistics* 22: 1-26.
- Lew, Robert.** 2012. How Can We Make Electronic Dictionaries More Effective? Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 343-361. Oxford: Oxford University Press.

- Liu, Tzu-Chien, Melissa Hui-Mei Fan and Fred Paas.** 2014. Effects of Digital Dictionary Format on Incidental Acquisition of Spelling Knowledge and Cognitive Load during Second Language Learning: Click-on vs. Key-in Dictionaries. *Computers & Education* 70: 9-20.
- Liu, Tzu-Chien and Po-Han Lin.** 2011. What Comes with Technological Convenience? Exploring the Behaviors and Performances of Learning with Computer-mediated Dictionaries. *Computers in Human Behavior* 27: 373-383.
- Lupescu, Stuart and Richard Day.** 1993. Reading, Dictionaries, and Vocabulary Learning. *Language Learning* 43(2): 263-287.
- Mestres-Missé, Anna, Thomas Münte and Antoni Rodríguez-Fornells.** 2014. Mapping Concrete and Abstract Meanings to New Words Using Verbal Contexts. *Second Language Research* 30(2): 191-223.
- Mondria, Jan-Arjen.** 2003. The Effects of Inferring, Verifying, and Memorizing on the Retention of L2 Word Meanings. *Studies in Second Language Acquisition* 25(04): 473-499.
- Moonen, Machteld, Rick de Graaff and Gerard Westhoff.** 2006. Focused Tasks, Mental Actions and Second Language Learning: Cognitive and Connectionist Accounts of Task Effectiveness. *International Journal of Applied Linguistics* 15(2): 35-53.
- Moonen, Machteld, Rick de Graaff, Gerard Westhoff and Mieke Brekelmans.** 2014. The Multi-feature Hypothesis: Connectionist Guidelines for L2 Task Design. *Language Teaching Research* 18(4): 474-496.
- Nakata, Tatsuya.** 2008. English Vocabulary Learning with Word Lists, Word Cards and Computers: Implications from Cognitive Psychology Research for Optimal Spaced Learning. *ReCALL* 20(01): 3-20.
- Nassaji, Hossein.** 2004. The Relationship between Depth of Vocabulary Knowledge and L2 Learners' Lexical Inferencing Strategy Use and Success. *The Canadian Modern Language Review* 61(1): 107-134.
- Nassaji, Hossein and Hsueh-chao Marcella Hu.** 2012. The Relationship between Task-induced Involvement Load and Learning New Words from Context. *International Review of Applied Linguistics in Language Teaching* 50(1): 69-86.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, Paul.** 2015. Principles Guiding Vocabulary Learning through Extensive Reading. *Reading in a Foreign Language* 27(1): 136-145.
- Paribakht, Sima and Marjorie Bingham Wesche.** 1993. Reading Comprehension and Second Language Development in a Comprehension-based ESL Program. *TESL Canada Journal* 11(1): 9-29.
- Prichard, Caleb.** 2008. Evaluating L2 Readers' Vocabulary Strategies and Dictionary Use. *Reading in Foreign Language* 20(2): 216-231.
- Pulido, Diana.** 2009. How Involved are American L2 Learners of Spanish in Lexical Input Processing Tasks During Reading. *Studies in Second Language Acquisition* 31(01): 31-58.
- Sadoski, Mark, Ernest Goetz and Maximo Rodriguez.** 2000. Engaging Texts: Effects of Concrete-ness on Comprehensibility, Interest, and Recall in Four Text Types. *Journal of Educational Psychology* 92(1): 85-95.
- Salsbury, Tom, Scott Crossley and Danielle McNamara.** 2011. Psycholinguistic Word Information in Second Language Oral Discourse. *Second Language Research* 27(3): 343-360.

- Webb, Stuart and Anna Chang.** 2015. How Does Prior Word Knowledge Affect Vocabulary Learning Progress in an Extensive Reading Program. *Studies in Second Language Acquisition* 37(04): 651-675.
- Westhoff, Gerard.** 2004. The Art of Playing a Pinball Machine. Characteristics of Effective SLA-tasks. *Babylonia* 12(3): 58-62.
- Zareva, Alla and Brent Wolter.** 2012. The 'Promise' of Three Methods of Word Association Analysis to L2 Lexical Research. *Second Language Research* 28(1): 41-67.
- Zou, Di.** 2012. *A Study of the Components of the Involvement Load Hypothesis: How Involvement Load Should be Allocated to "Search" and "Evaluation"*. Unpublished Ph.D. Dissertation. Hong Kong: City University of Hong Kong.
- Zou, Di.** 2016. Vocabulary Acquisition through Cloze Exercises, Sentence-writing and Composition-writing: Extending the Evaluation Component of the Involvement Load Hypothesis. *Language Teaching Research* 1362168816652418, first published on June 8, 2016 doi:10.1177/1362168816652418.

Appendix 1: Materials for Task 2

Dealing with Procrastination

Procrastination refers to the act of delaying the work you should do to a later time. It is wasting time when you have some work to do, but choose not to do it early.

Most procrastinators do not feel that they are doing this on purpose. Instead, they feel that they really tried to do the work. But they could not start because there are too many things out of their control. So they have a long list of **ostensible** reasons, such as “I did not have time”, “I had to attend a wedding” or “I had other important things to do”. However, these surface reasons are not true.

When procrastination becomes a habit, it is **pernicious**. If you procrastinate, you may often find yourself not having enough time to do a satisfying work. This can make other people unhappy and get a bad impression of you. Habitual procrastination can even damage your friendships. As you always **renege** on your promises to complete work on time, your friends may no longer trust you. Thus it is very necessary to know the causes of procrastination and learn to deal with it.

Procrastination is often caused by a real or imagined fear or worry. For instance, you might delay preparing for an oral presentation, because you are **apprehensive** that you will not be able to remember the entire speech. You may be so worried about doing a bad job that you decide not to work on it until the last minute.

Being a perfectionist is a main **trait** that causes fear and anxiety. When you imagine yourself making an English presentation, are you comparing yourself to great speakers? Or are you picturing that others **taunt** you about your accent? Rather than worry yourself with these thoughts, think of specific ways to improve the performance, may help to lessen performance anxiety.

“Lack of motivation” may also cause procrastination. If you are forced to teach a subject you are not interested in, you may find yourself wasting time instead of being **assiduous**. Even if you know that the subject can help you get a good job, you will not work hard. It is not easy to think carefully about something you have no interest in.

Feelings of **lassitude** can cause procrastination, too. This often happens when you keep on pushing yourself very hard without getting any rest. If so, you may experience a state of tiredness and feel unable to focus on any work. Learning to balance your time can be helpful in preventing this.

Sometimes you put off doing something because you do not know how to do it. For example, if you start doing a job that requires collecting data and creating graphics, having the right skills is **indispensable**. Knowing how to do a task before you begin it is very important. Sometimes it is difficult to ask for help and sometimes it is even harder to realize that you need help. Being able to **divulge** personal limitations and ask for help is a skill we need to learn.

1. **ostensible** means:

2. **pernicious** means:

3. **renege** means:

4. **apprehensive** means:

5. **trait** means:

6. **taunt** means:

7. **assiduous** means:

8. **lassitude** means:

9. **indispensable** means:

10. **divulge** means:

Considering the Merits, and Some Demerits, of the *Longman English–Japanese Dictionary*

Defa Ren, *College of Foreign Languages and Cultures, Xiamen University,
Xiamen, P.R. China (chinarendefa@163.com)*

Abstract: After comparing English monolingual and bilingual learner's dictionaries, the author discusses the *Longman English–Japanese Dictionary* (LEJD), compiled on the basis of three large-scale corpora. It impresses especially by its user-friendliness, seventeen aspects being listed which accounts for this. The parts comprising the macrostructure of LEJD are then enumerated, giving an overview of the useful information contained in the front, middle and back matter. The microstructure excels especially by the clear and simple definitions, the abundant and appropriate illustrative examples and the colourful and impressive pictorial illustrations. After considering these merits, the author mentions some demerits which will hopefully be reduced or eliminated in revised future editions.

Keywords: ENGLISH LEARNER'S DICTIONARIES, LANGUAGE CORPORA, USER-FRIENDLINESS, DEFINITIONS, ILLUSTRATIVE EXAMPLES, PICTORIAL ILLUSTRATIONS

Opsomming: 'n Beskouing van die voordele, en sommige van die nadele, van die *Longman English–Japanese Dictionary*. Ná die vergelyking van Engelse eentalige en tweetalige aanleerderswoordeboeke word die *Longman English–Japanese Dictionary* (LEJD), wat saamgestel is op basis van drie grootskaalse korpora, bespreek. Dit beïndruk die outeur, veral weens die gebruikersvriendelikheid daarvan. Sewentien aspekte hiervan word gelys. Die dele waaruit die makrostruktuur van die LEJD bestaan, word daarna genoem wat 'n oorsig gee van die bruikbare inligting vervat in die voor-, middel- en agterwerk. Die mikrostruktuur is uitnemend, veral weens die duidelike en eenvoudige definisies, die talryke en toepaslike illustratiewe voorbeelde en die kleurvolle en indrukwekkende illustrasies. Ná 'n beskouing van hierdie voordele, noem die outeur ook sommige nadele wat hopelik in hersiene toekomstige uitgawes verminder of uitgeskakel kan word.

Sleutelwoorde: ENGELSE AANLEERDESWOORDEBOEKE, TAALKORPORA, GEBRUIKERSVRIENDELIKHEID, DEFINISIES, ILLUSTRATIEWE VOORBEELDE, ILLUSTRASIES

The *Longman English–Japanese Dictionary* (ロングマン英和辞典, henceforth LEJD), co-edited by the British and Japanese linguists Geoffrey Leech and Yoshihiko Ikegami and first published in 2007 by Pearson Education, is the first genuine bilingual learner's dictionary exclusively intended for Japanese English learners. This dictionary has created many innovative breakthroughs, revolutionized the making of English learner's dictionaries, showcased the latest

developments in bilingual learner's dictionaries and embodied high achievements in British English learner's dictionaries in an endeavour to broaden the horizon of English learner's dictionaries. Only when LEJD is placed in a historical context can its significance be fully understood. Therefore, it is necessary to first take a retrospective look at the history of English learner's dictionaries.

With the rapid improvement of language processing technologies using computers and the establishment of the position of English as an international language, dictionary compilers and publishers of English-speaking countries came to realize that the special needs of non-native English learners had long been neglected and that the traditional way of making English dictionaries ought to be fundamentally changed to cater to the special needs of these non-native English learners. This change in the concept of dictionary making finally gave birth to English learner's dictionaries. Since the mid-twentieth century, this serious defect of monolingual English dictionaries has received more and more attention and many English dictionary compilers and publishers have made efforts to address it by compiling more and more excellent and innovative learner's dictionaries.

British dictionary compilers and publishers are pioneers in the compilation and publication of monolingual English learner's dictionaries capable of satisfying the special needs of non-native English learners in respect of English grammar, syntax, collocation, usage and cultural background. Later American dictionary compilers and publishers joined their British counterparts in compiling and publishing high-quality monolingual English learner's dictionaries. They took pains to make various innovations in compiling monolingual English learner's dictionaries, resulting in the production of excellent dictionaries enabling non-native English learners to acquire a better command of English. Generally speaking, British learner's dictionaries are more innovative than their American counterparts. Since the advent of the first generation of monolingual English learner's dictionaries represented by *The New Method English Dictionary* (West and Endicott 1935), *A Grammar of English Words* (Palmer 1938) and *Idiomatic and Syntactic English Dictionary* (Hornby, Gatenby and Wakefield 1942), monolingual English learner's dictionaries have evolved until the appearance of the British "Big Five" Learner's Dictionaries, namely, *Cambridge Advanced Learner's Dictionary* (McIntosh 2013), *Collins COBUILD Advanced Learner's Dictionary* (Collins COBUILD 2014), *Longman Dictionary of Contemporary English* (Pearson Education 2015), *Macmillan English Dictionary for Advanced Learners* (Rundell 2007) and *Oxford Advanced Learner's Dictionary* (Deuter et al. 2015), and the first American learner's dictionary, namely, *Merriam-Webster's Advanced Learner's English Dictionary* (Perrault 2008). As a matter of fact, British and American learner's dictionaries have been learning from each other to improve themselves so as to better cater to the ever-changing varied needs of their target users, helping them to gain a good English knowledge. For many English learners, especially those non-native English learners not living in English-speaking countries, English learner's dictionaries have become indispensable

reference books. For the time being, the compilation and publication of various monolingual English learner's dictionaries have overtaken the momentum of the compilation and publication of other types of English dictionaries. Why have monolingual English learner's dictionaries nowadays become so successful and popular? The underlying reason is that with the rapid development of economic globalization and advancement of science and technology, English has now become the most widely used language in the world. Under the strong influence of the success of monolingual English learner's dictionaries, monolingual learner's dictionaries of other languages are now being compiled or have been published.

In spite of the fact that monolingual English learner's dictionaries have resulted from ordinary monolingual English dictionaries, the former are very different from the latter in quite a few aspects:

- (1) Monolingual English learner's dictionaries are usually compiled on the basis of large-scale corpora and the headwords included in them are mainly core words and less frequently used words, while the headwords included in ordinary monolingual English dictionaries are common and frequently used words.
- (2) In monolingual English learner's dictionaries, the words used to define headwords are usually limited to about two or three thousand basic words to make it easier for users to understand the definitions of headwords, the ways in which headwords are defined are more diversified (sometimes a sentence may be used to define a headword), high-frequency headwords are usually marked to remind users that these words are of special importance, so that more attention should be paid to them and a greater effort be made to have a good command of them, and the arrangement of the different senses of a polysemous headword is usually presented according to the frequency of the use of each sense or the logical order of these senses, while in ordinary monolingual English dictionaries, the words used to define headwords are not limited to a small number, the ways in which headwords are defined are traditional and less diversified, high-frequency headwords are usually not marked and the arrangement of the different senses of a polysemous headword is usually given according to their historical or logical order.
- (3) Monolingual English learner's dictionaries contain more examples illustrating in which contexts and in what ways these headwords should be used than ordinary monolingual English dictionaries.
- (4) Monolingual English learner's dictionaries include more collocations than ordinary monolingual English dictionaries.
- (5) Monolingual English learner's dictionaries contain far more detailed labelling on the parts of speech of the included headwords than ordinary monolingual English dictionaries.

- (6) Monolingual English learner's dictionaries contain abundant and detailed pragmatic information about headwords that can help users to speak and write English properly and correctly, while ordinary monolingual English dictionaries usually do not pay much attention to the pragmatic information of headwords.
- (7) Monolingual English learner's dictionaries usually include a large number of instructive illustrations to help users acquire a better command of the meanings of the headwords, while ordinary monolingual English dictionaries traditionally contain only a very small number of illustrations or no illustrations at all.

As an English learner's dictionary, LEJD has all the above-mentioned conspicuous features shared by all learner's dictionaries. Nevertheless, LEJD is Longman's latest and most innovative dictionary deserving special attention and careful study, because it has made a number of significant breakthroughs:

- (1) It is the first genuine bilingual English–Japanese learner's dictionary intended exclusively to help Japanese English learners to acquire a good command of English. Although there are many other bilingual English–Japanese dictionaries in the domestic Japanese market, they can hardly be called genuine bilingual English–Japanese dictionaries.
- (2) It has been compiled on the basis of three large-scale corpora, the first one being the 330-million-word Longman Corpus Network, an English corpus previously developed by the Longman Company, the second being the Longman Learner's Corpus, an English corpus of the various typical errors likely to be made by non-native Japanese English learners newly created by the Longman Company, and the third being the 50-million-word Contemporary Japanese Corpus, a Japanese corpus newly developed by the Longman Company to verify whether the Japanese equivalents of the English headwords, illustrative examples and collocations in this English–Japanese dictionary are genuine idiomatic expressions of contemporary Japanese.
- (3) The greatest strength of LEJD is that the naturalness of its English, especially the naturalness of the illustrative English examples, can be guaranteed, because of the following two measures:
 - (a) All the illustrative English examples included in LEJD have been carefully selected from a huge corpus of British and American English by the editorial staff of the Longman Company who have long been renowned for compiling excellent original and innovative monolingual English dictionaries. Although there are many other so-called "corpus-based English–Japanese dictionaries" in the domestic Japanese market, the naturalness of the English in those dictionaries has not been guaranteed, because their illustrative examples are

selected from various English corpora and their English has not been confirmed as genuine natural English by native English speakers, especially well-trained and -qualified native English dictionary compilers. It is often the case that an illustrative English example may appear in a corpus, but turns out not to be natural English. Although the various corpora have become more and more powerful, it must be borne in mind that every corpus has its own limitations. LEJD has avoided these limitations by having the illustrative English examples confirmed as natural English by well-trained and -qualified compilers who are native English speakers.

- (b) The English used in this dictionary is Standard English and the basis for its compilation is the Longman Corpus Network, a huge and reliable corpus containing both British and American spoken and written English.
- (4) The most marked breakthrough made by LEJD is its user-friendliness. When compared with LEJD, well-known English–Japanese dictionaries such as *Collins COBUILD English–Japanese Advanced Dictionary of American English* (Collins COBUILD 2008), *Genius English–Japanese Dictionary* (Minamide 2014), *Kenkyusha’s English–Japanese Dictionary for the General Reader* (Takahashi 2012), *Kenkyusha’s New English–Japanese Dictionary* (Takebayashi 2002), *Shogakukan Random House English–Japanese Dictionary* (Konishi 1993) and *The Wisdom English–Japanese Dictionary* (Inoue and Akano 2012), will pale by comparison as far as user-friendliness is concerned. The user-friendliness of LEJD lies in the following aspects:
- (a) On the basis of the statistics of the Longman Corpus Network, the 3,000 most frequently used words in spoken English and the 3,000 most frequently used words in written English presented as head-words in LEJD are all printed in red to make them conspicuous to dictionary users; the 1,000 most frequently used words in spoken English and the 1,000 most frequently used words in written English are marked in red "S1" and in red "W1" respectively; the 2,000 most frequently used words in spoken English and the 2,000 most frequently used words in written English are marked in red "S2" and in red "W2" respectively; and the 3,000 most frequently used words in spoken English and the 3,000 most frequently used words in written English are marked in red "S3" and in red "W3" respectively.
 - (b) According to the classification of JACET (The Japan Association of College English Teachers), the 1000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000 and 8,000 most frequently used English words used as head-words in LEJD are marked in red "J1", "J2", "J3", "J4", "J5", "J6", "J7" and "J8" respectively to remind Japanese dictionary users that these words fall into the category of the basic vocabulary that they should strive to

- master to complete their school education effectively at different learning stages.
- (c) The Japanese equivalents of the 3,000 most frequently used words in spoken and the 3,000 most frequently used words in written English given as headwords in LEJD are all printed in red to make them conspicuous to users.
 - (d) The different senses of each polysemous English headword are all arranged in order of frequency so as to enable users to obtain access to their most needed information as quickly as possible, with higher frequency senses being presented before lower frequency senses.
 - (e) The nuances of English headwords, and the occasions on which these headwords should be used are also correctly reflected in the Japanese equivalents of these English headwords.
 - (f) On occasions when two or more Japanese equivalents of an English headword are provided, LEJD clearly indicates which Japanese equivalent should be used in which context so that its users can acquire a better command of the usage of each English headword.
 - (g) All illustrations are printed in full colour to reduce eye strain when its users look up words.
 - (h) All English collocations are printed in bold type to make them salient.
 - (i) All English sentence patterns are printed in bold type without using complicated and confusing signs.
 - (j) Exclamation marks are widely used to call the attention of users to spellings and pronunciations that are prone to be misspelled or mispronounced.
 - (k) All the mistakes frequently made by Japanese English learners are first analyzed in detail by using the Longman Runners Corpus and then clearly pointed out.
 - (l) When two or several English words or expressions may be used to describe a certain situation in life, study or work, LEJD always points out which English word or expression is the more or the most idiomatic so that its users are enabled to speak or write English correctly and properly.
 - (m) Many important grammatical items in English are provided with a linguistic differentiation column to help explain them clearly and plainly to users so as to reduce mistakes that are highly likely to be made by ordinary Japanese English learners.
 - (n) Language notes are frequently provided on the etymology of English headwords, the nuances of words listed in thesauri, and words for

which it is difficult to find Japanese equivalents, explaining the British and American cultural backgrounds to broaden the horizon of Japanese English learners and enable them to become proficient in English.

- (o) Both British and American English headwords are included in LEJD, which is more convenient for dictionary users to acquire a good command of both British and American English so that they need not look up these words in more than one dictionary.
- (p) As headwords, irregular verbs in both British and American English are followed by their past tense and past participle forms to assist users to apply these verbs correctly.
- (q) In the first two parts of the middle matter of LEJD, a communication guide with regard to oral communication and essay vocabulary is provided to enhance users' ability to communicate in spoken or written form more effectively and smoothly.

In the "oral communication" part, the guide tells users how to agree or disagree with somebody, how to request or invite somebody to do something, how to decline something to somebody, how to make suggestions, how to accept something, how to apologize to somebody or to accept somebody's apology, how to ask for permission, how to thank somebody or to respond to somebody's thanks, how to express one's own opinions or to ask for somebody's opinions, how to express understanding or no understanding of something, how to make proposals, how to confirm or deny something, how to make or answer phone calls, how to express one's wishes or hopes, how to express one's convictions or congratulations, how to praise somebody or to respond to somebody's praises, how to ask for road information from somebody or to give road information to somebody, how to introduce oneself or somebody else to others or how to reply to an introduction, how to greet somebody or how to reply to somebody's greetings, and how to bid farewell to somebody.

In the "essay vocabulary" part, users are provided with many useful high-frequency words closely related to the following fields: (1) art and culture; (2) books/literature; (3) business; (4) computers/the Internet; (5) describing people; (6) education; (7) exercise; (8) movies/films; (9) global issues; (10) government/politics; (11) health care and medicine; (12) hobbies and things one likes to do; (13) jobs and work; (14) music; (15) newspapers and magazines; (16) science and technology; (17) television and radio; (18) the environment; (19) vacation /holiday. Having a good command of the meanings and usage of high-frequency words from the above-mentioned fields will improve the essay writing ability of users as these words will provide them with what is essential to express their ideas freely and clearly. Without mastering a certain amount of vocabulary, no one can write a good essay.

In the last part of the middle matter of LEJD, some frequently used con-

junctions or conjunctive phrases with illustrative examples and detailed language notes on how to use them correctly and properly in spoken and written English are provided, which is very helpful to English learners, especially non-native English learners who in normal circumstances do not know the nuances of English words and often make numerous mistakes that seem amusingly odd and completely unintelligible to native English learners.

The macrostructure of LEJD comprises the following parts: (1) preface; (2) ways on how to use the dictionary; (3) ways on how to use the communication guide; (4) introduction; (5) *Longman English–Japanese Dictionary* — corpora and collocations; (6) the *Longman Contemporary Japanese Corpus* and the various new measures taken in the dictionary; (7) acknowledgements; (8) list of the editors, proofreaders and relevant cooperators of the dictionary; (9) signs and abbreviations used in the dictionary; (10) pronunciation directions; (11) list of the pronunciation signs; (12) the main body of the dictionary; (13) appendices to the dictionary. The first eleven parts form the front matter and the last part composes the back matter of the dictionary. In addition, the "communication guide" and "conjunctive phrases" constitute the middle matter. Obviously, LEJD contains plenty of front, middle and back matter, which is helpful and convenient to users who want to take full and efficient advantage of the dictionary to improve their English.

Regarding the microstructure of LEJD, three points are especially worth mentioning, namely, the clear and simple definitions of the headwords, the abundant and appropriate illustrative examples and the colourful and impressive pictorial illustrations.

The quality of the definitions in a dictionary determines to a very large extent the quality of the whole dictionary. Therefore, compilers should attach great importance to the quality of the definitions. The compilers of dictionaries have in fact for a very long time been vying with each other to contrive various novel ideas to improve the quality of dictionaries so as to better satisfy the needs of users. Generally speaking, the compilers of LEJD have done a very good job in defining its headwords in a clear and simple way so that its users can readily understand the meanings. In addition, the Japanese equivalents of all English headwords are genuine Japanese, based on the *Longman Contemporary Japanese Corpus*. Perhaps the boldest and most innovative move taken by the compilers of LEJD was to present all the main senses of each polysemous headword in a red semantic frame right under the headword, followed by all its detailed senses as well as some illustrative examples. This method of arranging the senses of each headword is very effective in saving the time and energy of users by first giving them a general picture of the rough meanings of each polysemous headword and then guide them to the meaning they are looking for as quickly and accurately as possible.

Illustrative examples have now become a must for all learner's dictionaries. The number and quality of the examples determine to a large degree the usefulness and popularity of a dictionary. Illustrative examples have been

playing a unique and irreplaceable role in helping users to acquire a good command of the meanings and usages of the consulted words by providing users with the typical contexts of a headword and showing users how a headword should be employed in these typical contexts. A proper illustrative example is, in many cases, far better than a prolix note on the usage of an English headword. All the illustrative examples in this dictionary have been carefully selected from large-scale corpora and thus they are authoritative and reliable, which ensures the usefulness and popularity of LEJD.

The abundant and colourful pictorial illustrations in LEJD are particularly impressive and praiseworthy. Traditionally British dictionaries contain almost no illustrations, in sharp contrast to their American counterparts that traditionally incorporate many illustrations. However, since the advent of learner's dictionaries, great changes to this British tradition have taken place and a new tradition of learner's dictionaries has gradually taken form. LEJD has continued the tradition of British learner's dictionaries by including a large number of illustrations. Illustrations in dictionaries play a pivotal role, being very effective in helping users better to understand and grasp culture-loaded or culture-specific words, which are widespread in any culture, stemming from different geographical and social environments, different histories, customs and traditions, different ways of thinking and other factors. In many cases, a good illustration can make an English headword much better comprehensible than an elaborate definition, for it gives users a vivid intuitional image of the consulted word.

Despite the above-mentioned praiseworthy features, LEJD shows the following inadequacies:

- (1) It has no corresponding electronic version in CD-ROM or EPWING format that can be used on ordinary computers or cell phones, rendering it virtually impossible for its users to consult new words on ordinary computers or cell phones while other contemporary mainstream English–Japanese dictionaries like *Collins COBUILD English–Japanese Advanced Dictionary of American English*, *Genius English–Japanese Dictionary*, *Kenkyusha's English–Japanese Dictionary for the General Reader*, *Kenkyusha's New English–Japanese Dictionary*, *Shogakukan Random House English–Japanese Dictionary*, and *The Wisdom English–Japanese Dictionary* all have their corresponding electronic versions so as to provide their users with more options, and making it convenient for them to look up new words.
- (2) The colour contrast of the illustrations needs to be improved, because, although full-colour illustrations can leave a vivid impression on its users, the contrast of the colours of the illustrations is not sharp enough and may cause eye strain if users look at them long or often.
- (3) The font is not large enough to be seen clearly by users with poor eyesight.

- (4) Although it includes 102,000 headwords and phrases, 83,000 natural illustrative examples, 63,000 collocations, 2,000 error or usage notes for Japanese English learners, unique thesaurus notes as well as numerous notes about grammar, formality, sound/pronunciation, etc., it still needs the inclusion of more of these to better cater to the ever-changing needs of its users. It often proves that its editors' standards for selecting headwords, phrases, illustrative examples and collocations are not completely objective, not ruling out the possibility that sometimes the editors think that there is no need to include certain words, phrases, illustrative examples, collocations or notes, which users in fact may consider necessary.

On the whole, however, LEJD is the best English–Japanese bilingual dictionary published so far and the best representative of British learner's dictionaries, because it has revolutionized English–Japanese bilingual dictionaries and is representative of the highest achievements of British learner's dictionaries regarding the following aspects: (1) the naturalness of the English; (2) the genuineness of the Japanese; (3) the user-friendliness of the design; (4) the careful consideration of details; (5) the clarity and simplicity of the definitions; (6) the abundance and appropriateness of its illustrative examples; (7) the colourfulness and impressiveness of its illustrations; (8) the skilful and accomplished way of combining dictionary making with three large-scale corpora and advanced corpus technologies; and (9) the practicality for language learning. This epoch-making dictionary is the first and sole bilingual dictionary entirely based on English and Japanese corpora and designed to help users obtain access to their needed information as quickly and conveniently as possible for learning natural English and genuine Japanese.

Although there are still certain inadequacies in LEJD, these by no means outweigh its merits. Up till now this revolutionary dictionary is the best English–Japanese dictionary available. Most certainly its inadequacies will be gradually reduced or eliminated in revised future editions.

References

- Collins COBUILD (Ed.).** 2008. *Collins COBUILD English/Japanese Advanced Dictionary of American English*. Boston: Heinle ELT.
- Collins COBUILD (Ed.).** 2014. *Collins COBUILD Advanced Learner's Dictionary*. Eighth Edition. Glasgow: HarperCollins Publishers.
- Deuter, M. et al. (Eds.).** 2015. *Oxford Advanced Learner's Dictionary*. Ninth Edition. Oxford: Oxford University Press.
- Hornby, A.S., E.V. Gatenby and A.H. Wakefield (Eds.).** 1942. *Idiomatic and Syntactic English Dictionary*. Tokyo: Kaitakusha.
- Inoue, Nagayuki and Ichiro Akano (Eds.).** 2012. *The Wisdom English–Japanese Dictionary*. Third Edition. Tokyo: Sanseido.

- Konishi, Tomoshichi (Ed.).** 1993. *Shogakukan Random House English–Japanese Dictionary*. Second Edition. Tokyo: Shogakukan.
- Leech, Geoffrey and Yoshihiko Ikegami (Eds.).** 2007. *Longman English–Japanese Dictionary*. Tokyo: Pearson Education.
- McIntosh, C. (Ed.).** 2013. *Cambridge Advanced Learner’s Dictionary*. Fourth Edition. Cambridge: Cambridge University Press.
- Minamide, Kosei (Ed.).** 2014. *Genius English–Japanese Dictionary*. Fifth Edition. Tokyo: Taishukan.
- Palmer, H.E. (Ed.).** 1938. *A Grammar of English Words*. London: Longmans, Green & Co.
- Pearson Education (Ed.).** 2015. *Longman Dictionary of Contemporary English*. Sixth Edition. Harlow: Pearson Education ESL.
- Perrault, S.J. (Ed.).** 2008. *Merriam-Webster’s Advanced Learner’s English Dictionary*. Springfield, Mass.: Merriam-Webster.
- Rundell, M. (Ed.).** 2007. *Macmillan English Dictionary for Advanced Learners*. Second Edition. Oxford: Macmillan Education.
- Takahashi, Sakutarō (Ed.).** 2012. *Kenkyusha’s English–Japanese Dictionary for the General Reader*. Third Edition. Tokyo: Kenkyusha.
- Takebayashi, Shigeru (Ed.).** 2002. *Kenkyusha’s New English–Japanese Dictionary*. Sixth Edition. Tokyo: Kenkyusha.
- West, M. and J. Endicott (Eds.).** 1935. *The New Method English Dictionary*. London: Longmans, Green & Co.

Learner Features in a New Corpus-based Swahili Dictionary

Beata Wójtowicz, *Department of African Languages and Cultures,
University of Warsaw, Warsaw, Poland (b.wojtowicz@uw.edu.pl)*

Abstract: As far as traditionally published Swahili language dictionaries are concerned, throughout the long history of Swahili lexicography, most new dictionaries were based on their predecessors. Thus far the only innovative traditionally printed corpus-based dictionary has been published by Finnish scholars (Abdulla et al. 2002). This dictionary takes advantage of the Helsinki Corpus of Swahili (HCS 2004) and was an inspiration for the new Swahili–Polish dictionary project described in this paper. The Swahili–Polish language pair, which without doubt can be called *less resourced*, does not have many lexicographical antecedents on which a new dictionary could be based. Therefore, the new dictionary relies on data from the Swahili corpus.

In this paper, we present a new, corpus-based Swahili–Polish dictionary that has been published online and printed. The paper introduces the resources used to build the dictionary, its learner-oriented features and grammatical assumptions, with a focus on the idea of the visualisation of derivational hierarchies.

Keywords: SWAHILI, POLISH, BILINGUAL LEXICOGRAPHY, DICTIONARY PROJECT, CORPUS-BASED LEXICOGRAPHY

Opsomming: Aanleederskenmerke in 'n Nuwe Korpusgebaseerde Swahili-woordeboek. Dwarsdeur die lang geskiedenis van Swahilileksikografie is die meeste nuwe woordeboeke, veral tradisioneel gepubliseerde Swahili taalwoordeboeke, op hulle voorgangers gebaseer. Tot dusver is die enigste innoverende tradisioneel gedrukte korpusgebaseerde woordeboek gepubliseer deur Finse vakkundiges (Abdulla et al. 2002). Hierdie woordeboek trek voordeel uit die Helsinki Korpus van Swahili" (HCS 2004) en was 'n inspirasie vir die nuwe Swahili–Poolse woordeboekprojek wat in hierdie artikel beskryf word. Die Swahili–Poolse taalpaar, wat sonder twyfel beskryf kan word as *gebrekkig aan hulpbronne*, het nie baie leksikografiese voorgangers waarop 'n nuwe woordeboek gebaseer kan word nie. Daarom maak die nuwe woordeboek staat op data uit die Swahilikorpus.

In hierdie artikel lê ons 'n nuwe, korpusgebaseerde Swahili–Poolse woordeboek voor wat aanlyn gepubliseer sowel as gedruk is. Die artikel stel die bronne bekend wat gebruik is om die woordeboek op te bou, sy aanleedersgerigte kenmerke en taalkundige veronderstellings, met 'n fokus op die begrip van visualisering van afleidingshiërargieë.

Sleutelwoorde: SWAHILI, POOLS, TWEETALIGE LEKSIKOGRAFIE, WOORDEBOEK-PROJEK, KORPUSGEBASEERDE LEKSIKOGRAFIE

1. Introduction

Swahili is one of Africa's major languages. It is the most widely-used language of Sub-Saharan Africa with tens of millions of people using it as a *lingua franca* throughout the East African region. It is the national and official language of Kenya and Tanzania (alongside English). Due to its status, the number of speakers and the interest of scholars, numerous Swahili dictionaries have been compiled all over the world, with electronic (online) dictionaries joining the market in the 1990s.

Both at European and American universities, African languages are recognized as less-commonly taught languages. This often goes hand in hand with a scarcity of lexicographic and teaching materials. This is also true for Swahili in non-English speaking parts of the world. Although Swahili has been taught alongside Hausa at the University of Warsaw in Poland since the 1960s, only a few teaching resources have been compiled during this time.

Even though numerous bilingual Swahili dictionaries exist, the only publication of this kind targeting Polish was the small Swahili–Polish and Polish–Swahili dictionary by Stopa and Garlicki (1966), which has been out of stock for a number of years. While students can manage with various English-language textbooks available for studying Swahili, they find it difficult without access to a bilingual dictionary in their native language, especially during the first years of their studies. Until recently, Polish students were forced to use non-Polish dictionaries that are not easily available for purchase. Therefore, they have often taken advantage of various resources that are available on the Internet. Two electronic Swahili–English dictionaries were especially popular and widely-used: for years this included the *Internet Living Swahili Dictionary*¹ — the largest such dictionary and on-line community-based initiative that has now evolved into the *Global Online Living Dictionary*, as well as the *TshwaneDJe Swahili–English Dictionary*² (Hillewaert et al. 2012) — the first, and so far the only, corpus-driven electronic Swahili online dictionary. Although the *Internet Living Swahili Dictionary* contained over 50 thousand entries and has been the leading dictionary for years, it suffered as a result of the community-based mode of its creation. It tended to list many minor meanings of the entry in an unsorted manner among the more important (frequent) ones, which beginner learners often found misleading. The *TshwaneDJe Swahili–English Dictionary* is recognised by Polish students as the best dictionary of Swahili at present. It is the first corpus-driven dictionary of Swahili with a new approach to the lemmatisation of headwords (cf. De Schryver et al. 2006). The content is based on web-based corpus data (as the authors themselves describe it, *a balanced and representative Swahili corpus of around fifteen million running words* (De Schryver et al. 2006: 70).

As far as traditionally published dictionaries are concerned, throughout the long history of Swahili lexicography, most new dictionaries were based on their predecessors; thus far the only innovative corpus-based dictionary has

been published by Finnish scholars (Abdulla et al. 2002). It was based on data from the Helsinki Corpus of Swahili, the only annotated and publicly available, with free access for researchers, electronic corpus of the language. This dictionary was a source of inspiration for the project described in this paper.

2. The new Swahili–Polish dictionary project

Over the last several years, we have observed a rising interest in African studies at the University of Warsaw, especially in the field of Swahili language studies.

Given the growing interest in learning Swahili, the fact that Swahili dictionaries are not easily available for purchase in Poland, and that the only Swahili–Polish dictionary is now simply out-dated, the compilation of a new Swahili–Polish (and Polish–Swahili) dictionary became a necessity. Due to the shortage of educational materials, a new project involving the creation of a dictionary was proposed by the Department of African Languages and Cultures, and the idea was then approved by the Polish Ministry of Science.

The project aimed to create a new dictionary that would be developed in accordance with recent lexicographic practices. Therefore, the macrostructure as well as the description of entries are based on corpus data from the Helsinki Corpus of Swahili (HCS 2004). The electronic version of the dictionary³, which is discussed in this article, is primary to the printed version that has also already been published (Wójtowicz 2013). The focus within this article is on the Swahili–Polish direction, as the reverse-language part in the electronic version of the dictionary is delivered in the form of a structured index — a standard in this type of resource.

The dictionary is aimed primarily at Polish students of Swahili. The other target groups include tourists visiting East Africa and anyone interested in the Swahili language and culture. The needs of these three groups partially overlap since the culture-oriented programme of African studies ensures that first-year students acquire vocabulary needed in day-to-day interaction with native speakers of Swahili.

While there are no other Polish resources the students can use, we aimed at creating a dictionary that could also serve as a language learning aid and supplement other foreign teaching materials used in language classes, as well as help students with homework. Furthermore, we also regard the dictionary as a source of cultural knowledge.

The project website, apart from being a dictionary interface, provides sections on *how to use the dictionary*, *the language* — a general introduction to the language and its grammar, a *Swahili phrasebook* — a list of popular and useful phrases, and *Swahili proverbs* that present over 60 different Swahili sayings and proverbs translated into Polish.

The following sections briefly outline the main concepts of the dictionary, as well as the data and software that were used to build it.

3. The Helsinki Corpus of Swahili — the source of dictionary data

The skeleton of the dictionary was derived from the Helsinki Corpus of Swahili (HCS). The HCS is the only large and annotated corpus of Standard Swahili available to the linguistic community. Although it is neither a representative nor balanced corpus, it seems to contain appropriate data considering our dictionary target users. It consists of over 12 million words taken from numerous literary books and current news sources, i.e. texts that an average student of Swahili comes across in his day-to-day learning process. Literary works are read in class, while news items written in Swahili are the most popular and easily accessible texts on the Internet. The size of the corpus may be estimated as small and insufficient as compared to the big corpora available for many European languages. Nevertheless, the HCS is one of the biggest corpora for African or other lesser-resourced languages, and the findings of Prinsloo (2015) prove that a corpus of even a small size should supply enough data to compile a good dictionary for our target group.

The annotation layer of the corpus was provided by SALAMA (Swahili Language Manager, cf. Hurskeinen 2008), an environment for the computational processing of the Swahili language. SALAMA includes a comprehensive language analyser of Swahili text, including morphological analysis, morphological and semantic disambiguation, syntactic analysis, and a bilingual translation module from Swahili to English, making it possible to compile a corpus-based dictionary. The SALAMA Dictionary Compiler (SALAMA-DC), a by-product of SALAMA, is a comprehensive system for producing dictionary entries from any word-form in Swahili. It produces entries with appropriate linguistic information, single-word headwords, multiword headwords, various types of cross-references, and a selection of usage examples in context. Furthermore, the example texts are translated into English and attached to each entry. The whole process from raw text to dictionary entries takes place without manual editing in between, which speeds up the compilation of the dictionary greatly. The output is provided in text format, all as a single file. The dictionary compiled in such an automated process needs manual editing and correction; however, the work needed during this step is only a fraction of the work carried out in manual dictionary compilation.

The dictionary data has three main types of information:

- headword including relevant information (within square brackets)
[kiasi-N-7/8] {quantity, amount, measure} AR 3384
- cross-references (with double square bracket), *taz.* stands for 'cf.'
[kiasi-N-7/8]] {amount} taz. [kadri-N-9/10] [kima-N-7/8]
[kiasi-N-7/8]] {quantity} taz. [idadi-N-9/10]
- examples in context (with triple square bracket)
[kiasi-N-7/8]]] Alifungua bomba, akakinga kiasi [kiasi-N-7/8] na kisha akanywa hali makofi (Opened the pipe, protected the quantity and then drank it is not the slaps)

Since the primary data included in the dictionary has been derived from a tagged corpus, all entries were accompanied by basic linguistic information, depending on the word category, right from the beginning. In this primary data, each entry contains POS information, and furthermore, nouns are described by their class number and animacy categorization, while verbs are described by the type of derivation and references to their base/root in the case of derivatives, or to derivatives in the case of roots. Additionally, all entries are accompanied by their English equivalents. The step of translation into Polish has been further facilitated by the automatic concatenation⁴ of the Swahili–English dictionary with the English–Polish dictionary by Piotrowski and Saloni (1992).

Apart from the grammatical information and English glosses, the raw data contains additional information on verb features, frequency counts, the etymology of borrowings, and references to synonyms. What follows are three typical dictionary entries produced by SALAMA-DC. An automatic concatenation with the English–Polish dictionary has been carried out but no further editing was done. In this dictionary compilation process, all verbal extensions have been given the status of separate entries, as in (2).

(1) Noun:

[ardhi-N-9/10] {land, soil (gleba, ziemia)} AR 2247
[ardhi-N-9/10]] {soil} taz. [chafua-V] [chafulia-V] [chafuliwa-V] [dongo-N-5/6]
[udongo-N-11]

(2) Verb:

[apa-V] [apa] {swear (przysięgać, przeklinać, kląć), take_an_oath} SV 131
[apa-V]] {swear} taz. [apia-V] [apisha-V] [apishwa-V]
[apia-V] [apa] {swear (przysięgać, przeklinać, kląć), take_an_oath} SVO APPL 9
[apia-V]] {swear} taz. [apa-V] [apisha-V] [apishwa-V]

(3) Adjective:

[bandia-ADJ] A-UNINFL {artificial (sztuczny), spurious (pozorny, fałszywy),
counterfeit} AR 153

The status of separate entries was also given to idioms, or more generally, multi-word expressions.

(4)

[fanya_kazi] V SVO IDIOM-V {work} 2584
[unga_mkono] V SVO IDIOM-V {support, be of the same opinion} 951
[makao_kuu] N 6SG MWE {headquarters} 893
[ona_raha-V] SVO IDIOM-V {rejoice (radować się)} 15
[uchaguzi_kuu] N 11/10 MWE {general elections} 791

A great advantage of the SALAMA-DC system is the ability to include into the dictionary multiword examples, isolate them, and produce appropriate translations. Most entries are accompanied by several examples, and each of them is

translated into English, as in example (5). Example sentences were cut on both sides of the keyword, preferably in clause boundaries, but if such boundaries were too far, the sentence was cut after a certain number of words. If the headword has more than one interpretation (different part of speech or noun class), each interpretation has its own examples separately, as in (6). Translations are only given to provide some clue of the meaning, and in cases where translations need to be retained, they should undergo extensive editing. Since long examples were cut, the translation sometimes suffers as a result of incomplete sentences, but it undoubtedly sheds light on what the sentence is about. The English translation is a word-to-word gloss of Swahili text.

(5)

[piga_chafya-V] SVO IDIOM-V {sneeze} 31
[piga_chafya-V]] Hata hivyo mwenzake hakuweza kupiga [piga_chafya-V]
chafya wala kujitingisha. (However the countryman did not be able to sneeze
nor to encircle self.)
[piga_chafya-V]] Pia kutema ovyo na kupiga [piga_chafya-V] chafya karibu na
mtoto ni hatari (Also to cut carelessly and to sneeze near the child is the danger)

(6)

[mpaka-N-3/4] {border, boundary, frontier} 1390
[mpaka-N-3/4]] {border} taz. [pakana-V]
[mpaka-N-3/4]] Ila bila msaada unaovuka mipaka [mpaka-N-3/4] ya mwanga
ungeweza kupotea njia (Except without the assistance whom it crosses the
borders of the light it would be able to be lost the way)
[mpaka-PREP] {until, till} 2140
[mpaka-PREP]] {till} taz. [hadi-PREP]
[mpaka-PREP]] {until} taz. [hadi-PREP] [hata-ADV] [lama-ADV] [mpaka-CONJ]
[mpaka-PREP]] Aliamua kuwa asitelemke Makambako mpaka [mpaka-PREP]
baadaye hivyo alipitiliza moja kwamoja (He decided he should not descend
Makambako until afterwards in this way surpassed one kwamoja)
[mpaka-PREP]] Aliendelea kuzunguka huko jikoni, [mpaka-PREP] na mara
akaiangalia saa: (He continued to go around there kitchen in, and immedi-
ately looked at it the hour:)

The lemma-sign list that was used to build a dictionary consisted of over 10 thousand entries. Most of them, app. 50% were nouns, over 20% were verbs, 10% adjectives, and 6% adverbs plus other parts of speech and multiword expressions. This list was compared with vocabulary from students' books like *Colloquial Swahili* (McGrath and Marten 2003) and *Tusome Kiswahili* (Muaka and Muaka 2006). All vocabulary from the first book was present on the list and around 50 entries were identified as missing from the other book. These were mainly multiword expressions, like *kitinda mimba* 'last born child', *baba wa kambo* 'step father', *chama cha siasa* 'political party', *mwandishi wa habari* 'journalist' and were added to the dictionary. Furthermore closed sets like days of the

week, months, and pronouns were verified and names of countries and continents added. Additional vocabulary was also verified and supplemented by students who worked on chosen sets they found useful in their studies, like animals, musical instruments, or means of transport.

The first phase of dictionary editing was cleaning up messy entries with the help of regular expressions. Several entries were merged together with no space in between, several had wrong examples attached, as the last example in (7) — for *toka* instead of *tohara*. Some entries were wrongly identified, had no POS information or senses attached. But the main work was devoted to identifying wrong translations from English to Polish. In example (8) *akili* is well translated into English as 'intelligence, intellect' but further into Polish as *inteligencja* 'intelligence' and *wywiad* that means in Polish both 'intelligence service' and 'interview', but none of them is an appropriate translation of the Swahili word *akili* and only the first translation into Polish, *inteligencja*, should make it into the dictionary. In (9) English 'interest' is translated into Polish as 'interests, hobby' and the right interpretation of 'bank interest' is completely missing.

(7)

[tohara-N-9/10] {circumcision (obrzezanie), purity (czystość), cleanliness (schludność)} AR 76

[tohara-N-9/10]] {cleanliness} taz. [unadhifu-N-11] [usafi-N-11]

[tohara-N-9/10]]] <ALA> akasema kampeni dhidi ya tohara [tohara-N-9/10] kwa wanawake mbali na kuhitaji (Said give him/her against the circumcision with the women apart from to need)

[tohara-N-9/10]]] <ALA> wazazi kuharakisha kuwapeleka kwenye tohara

[toka-MWE]]] <ART> jumla ya sh milioni 10,826,740 [toka-MWE] katika wilaya zake na kwamba (The total of sh million in his/her/its districts and that)

(8)

[akili-N-9/10] {intelligence (inteligencja, wywiad), intellect (intelekt, inteligencja), nous} AR 1614

(9)

[riba-N-9/10] {interest (zainteresowanie, hobby, interes), usury (lichwa)} AR 146

Already identified senses were further researched in the corpus and completed when needed. For example, based on more corpus examples, *ghorofa* 'floor' got a second sense of 'seat, office'. Also noun class membership was verified and supplemented when needed. *Dawa* 'medicine' was initially described as belonging to class 5/6 but based on corpus examples class 9/10 membership was also added.

4. FieldWorks Language Explorer

Originally, one of the aims of the project was the encoding of the dictionary in XML conformant with TEI P5 Guidelines (TEI Consortium 2015). The virtues of

such a choice lie in its extensibility, interchangeability, manageability, ease of maintenance, and the possibility of creating virtually any kind of output.

To enable easy editing, the data has been converted to the LIFT format for the time being and imported into the FieldWorks Language Explorer (FLEX), free software available from SIL⁵. FieldWorks Language Explorer consists of tools that help manage linguistic and cultural data. It is a powerful tool aimed at helping field linguists to perform many common language documentation and analysis tasks, among others, the dictionary development process. To mention but a few of its features, the lexical data can be formatted within FLEX for previewing before publication, in the print-oriented or electronic dictionary view, the order and formatting of fields can be changed, and it supports root-based or stem-based arrangements of entries. New custom fields can be added, existing ones omitted, reordered, formatted, and example sentences selected, so different publications can be derived from one database. With the possibility to deselect each entry, sense, and example sentence we can for instance create anything from a pocket edition to a full dictionary. It supports an extended system of cross-references, illustrations can be attached and data exported among others to XML. It also offers semantic domain-based, built-in descriptions of the lexicon, which could be the possible next step for the development of the dictionary under presentation. The main drawback of the system is the lack of a website publication module. It relies on the plug-in *Pathway*⁶ to create dictionary documents in various formats, and *Webonary*⁷ to publish the dictionary online. The data needs to be exported in XHTML format and then imported by *Webonary* that supports the creation of the dictionary website.

5. The concept of the new Swahili–Polish dictionary

Given the linguistic orientation of African studies, this general student-oriented dictionary contains extended grammatical information, above and beyond the scope of what is minimally necessary. It is assumed that even though the primary function of this Swahili–Polish dictionary is to provide help in the process of text reception, it may also provide grammatical information that might be helpful in text production.

The dictionary is not strictly translational but rather of a descriptive-translational type. As the cultural differences are substantial and the contacts between the two cultures slight, we aimed at providing culturally-bound entries with broad explanations and possibly pictures in the future. For this reason the list of entries also includes geographical names.

The dictionary is regarded as a work in progress and by closely monitoring user queries, new entries are constantly being added while the existing ones are updated. The dictionary may be searched in both directions with nearly 6100 Swahili entries and over 7000 entries in the searchable Polish index. The searches are carried out on headwords and plural forms of headwords for Swahili nouns. If a user looks up *wapishi* 'cooks', he will be directed to the entry for *mpishi* 'cook'.

5.1 The macrostructure and the treatment of derivatives

The selection of headwords for the Swahili–Polish dictionary (ultimately 10,000 entries, published incrementally) has been made primarily on the basis of a frequency list derived from the corpus, and further expanded with words from various Swahili textbooks and other material used in language classes. The dictionary headword list also includes geographical names and grammatical morphemes. Since it is not possible to translate the morphemes into Polish, their function in Swahili is explained.

While working with a Bantu language we had to address problems not experienced by lexicographers working with European languages. These problems are connected primarily to two issues: the form of headwords and the presentation of the numerous derivatives of a single root.

As a Bantu language, Swahili is an agglutinative language, which means that morphemes are juxtaposed to form words. Within the Swahili lexicographic tradition, the accepted lemmatisation strategy is to list nouns in their full forms with class prefixes, whereas the prefixes of verbs, numerals, pronouns and inflected adjectives are ignored and the stems alone are listed (see Kiango 2000 for a thorough discussion of this issue, and De Schryver et al. 2006 for a novel approach of lemmatising full orthographic words in addition to stems). Given the dictionary culture of our students, we have decided to follow chosen solutions of the lexicographic tradition, used within nearly all Swahili dictionaries, in regard to the process of lemmatization. Therefore, we ignore prefixes and list the stems alone for verbs, numerals, and inflected adjectives. On the other hand, in response to beginner learners' needs, we reject this tradition in some other cases and list pronouns in their full forms (with the stems also included as separate entries).

Another important issue with respect to macrostructure is related to the handling of derivatives, sometimes referred to as the "lumping vs. splitting" debate (cf. Bański and Wójtowicz 2011). Derivation in Swahili is very robust and typically creates dozens of complex lexemes from a single root since Bantu derivational word-families can be extremely numerous, especially those based on verbal roots — for example, De Schryver and Prinsloo (2001: 225ff) estimate that there are over 140 regular derivatives of the root *reka* 'buy, purchase' in Sepedi.

To meet the expectations of the users who find it difficult to cope with root-based dictionaries of Bantu languages, we follow the so-called splitting approach as the default method: rather than lumping all related lexemes in a single entry headed by the root form, we place derivatives of verbs in separate entries, thereby breaking the semantic and lexical connections between the individual derivatives and their respective bases. In order to maintain a system whereby derivatives have the status of headwords, while simultaneously the derivational and semantic relationships between forms are preserved, the dictionary uses an extended mechanism of cross-entry references. Also in the

process of implementing the structure in the form of XML, we realised that what we should do is treat the lumping vs. splitting debate not as a deep issue concerning the structuring of the data, but rather as a surface issue on how to present the data to the user. Therefore, emphasis is placed on the presentation and visualisation of derivational families, as this is regarded as an educational feature useful for the development of our students' linguistic skills.

Traditional cross-entry references, especially among word-families, offer a one-sided view of derivational relationships (derivative → root). Introducing run-on entries offers a view from the opposite side (root → derivatives). Typically, however, word-families feature more than two generations of words, and quite often the link between the ends of the chain (root ↔ complex derivative) is either unclear to the average speaker or at least not as important as the relationship between the immediately related lexemes. Therefore, as an addition to traditional cross-entry references, we introduce a visualisation of the derivational families showing explicitly how the lexemes relate to each other. In the derivational tree, the derivational bases point to the next level of the derivational hierarchy only, and derivatives point to their derivational bases, which crucially need not be the same as their roots. The user is presented with information on the root or its verbal derivatives within the entry and the whole derivational family tree in the form of an information graph accompanying the entry.

The screenshot shows a web interface for a Swahili-Polish dictionary. At the top, it says 'Słownik suahili-polski'. There are navigation links: 'O słowniku', 'O języku', 'Dodatki', and 'Kontakt'. A search bar contains 'angua' and has buttons for 'Szukaj', 'Suahili', 'Polski', and 'Gniazdowy układ haseł'. Below the search bar, it says 'Wyszukiwano: angua' and 'Liczba znalezionych haseł: 1'. The main content area shows the entry for '-angua' with a list of Polish equivalents: 1. upuszczać, 2. przewracać, 3. zrzucać, zrywać, 4. znosić jajka. A Swahili example sentence is provided: 'Rafiki yangu, twende tukaangu maembe ili tute. Przyjacielu, chodźmy zerwać mango, żeby je zjeść.' Below the list, there are arrows pointing to '-angushwa -anguka -angusha -angukia'. To the right, a derivational tree is shown, starting with '-angua' at the top, branching down to '-anguka', '-angukia', 'anguko', '-angusha', and '-angushwa'.

Making users aware of the structure of the hierarchy in one case reinforces their knowledge of the possible derivational patterns that can be productively applied in other cases, i.e. to the creation of new forms or to the analysis of newly encountered words, which need not be present in the dictionary due to their low text frequency.

To overcome the lump or split debate and exploit the possibilities offered by the new media, it is also possible to lump or split entries of verbal deriva-

tives. The user may choose if he wants to see nested entries (all derivatives presented under their roots) or all derivatives as separate entries (no other derivatives are then attached but the information about the root is still provided).

As for homonyms, they are determined on the basis of their morphological features, i.e. as different parts of speech.

5.2 The microstructure

The structure of an entry includes the following elements: the headword, variant/variants of the headword; grammatical information dependent on the POS and the properties of the individual lexical item; equivalents, definitions⁸; examples of use; idioms that the headword is part of; collocations; synonyms; etymological information.

Each dictionary entry is described with the basic grammatical information given in most of the dictionaries, such as the part of speech, noun class number, type of verbal derivation and the root — if applicable.

-lisha *v causative* [< -la] feed

Additionally, the dictionary also marks different grammatical features that the beginner student may find helpful, like the animal/human distinction for nouns from non-human classes that aims at drawing the user's attention to the agreement structure. To facilitate interpreting, subsequent parts of dictionary entries are presented using different colours, e.g. grammatical information is provided in blue and examples in green.

The main additional learner-oriented features of the dictionary are the following:

- explicitly marked class prefixes on nouns and pronouns and full plural forms of nouns:

m | pishi (*pl wapishi*) *n* 1/2 cook

- animacy of nouns from non-human classes (the distinction animate/human is introduced):

n | dege (*pl ndege*) *n* 9/10 *animate* bird

- irregular verbal forms (imperative or retaining of the augment *-ku-*):

-ja *v ku (imp. njoo)* come

- highlighting of the most frequently used words (the entry is red and labelled with a red dot):

-pata ● *v* get

In order to present the information more explicitly, pop-up windows and links are exploited. Moving a cursor over numbers in noun entries evokes a pop-up window with information that these represent the class membership of a given noun. Additionally, the same numbers are also hyperlinked to the agreement chart, where agreement patterns for animate/human subjects from non-human classes have been added. Pop-up windows also inform the users about the relation of the augment *-ku-* with different tenses or why some entries are printed in red.

m | pishi (pl *wapishi*) n 1/2 cook⁹



Klasa rzeczownika

-ja v ku (imp. njoo) come¹⁰



Czasownik zachowuje morfem ku w następujących formach: *ninakuja*, *nimekuja*, *nillikuja*, *nitakuja*, *ningekuja*.

A cross-reference system is also exploited to link synonymous entries, like *ndege* ↔ *eroplani* 'airplane'. Subsequent senses are listed based on their frequency and specialised senses are described with field labels. Real examples extracted from the corpus accompany many entries, and where needed useful phrases and multiword expressions have also been added.

6. Conclusion

The new Swahili–Polish dictionary presented in this article aims at solving the issue of a shortage of Swahili-language learning materials in the Polish market. Created with students of Swahili as the main target group, it provides a variety of learner-oriented features that will possibly simplify the task of acquiring the language and understanding Bantu morphology. The introduction of hierarchical derivational trees reinforces their knowledge of derivational patterns and provides an additional access route to the lexicon.

To keep the dictionary up-to-date, its macrostructure was based on the corpus-driven data from the Helsinki Corpus of Swahili. Most of the features of this data have been preserved in the final version of the dictionary. Within this paper, the primary dictionary data has been presented, while also discussing the main dictionary aims and assumptions already addressed.

Endnotes

1. The dictionary used to be available at <http://kamusi.org>.
2. <http://africanlanguages.com/swahili>. Accessed on 06/05/2016.

3. <http://kamusi.pl>. Accessed on 06/05/2016.
4. With the invaluable assistance of Prof. Arvi Hurskainen, to whom I am very thankful.
5. <http://fieldworks.sil.org/>. Accessed on 06/05/2016.
6. <http://pathway.sil.org/>. Accessed on 06/05/2016.
7. <http://www.webonary.org/>. Accessed on 06/05/2016.
8. When additional explanation is needed, for example of family relations not present in Polish as older brother, or different names for sisters/brothers of our parents.
9. Text in the box: 'noun class'.
10. Text in the box: 'the verb retains the morpheme *ku* in the following forms *ninakuja, nimekuja, nilikuja, nitakuja, ningekuja*'.

References

- Abdulla, A., R. Halme, L. Harjula and M. Pesari-Pajunen (Eds.)**. 2002. *Swahili–Suomi–Swahili-sanakirja*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Bański, P. and B. Wójtowicz**. 2011. New XML-encoded Swahili–Polish Dictionary: Micro- and Macrostructure. Goźdz-Roszkowski, S. (Ed.). 2011. *Explorations across Languages and Corpora. PALC 2009*: 497-514. Frankfurt a. Main: Peter Lang.
- 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., D. Joffe, P. Joffe and S. Hillewaert**. 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67–83.
- (HCS) Helsinki Corpus of Swahili**. 2004. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC — Scientific Computing Ltd.
- Hillewaert, S., P. Joffe and G.-M. de Schryver**. 2012. *Kamusi ya Kiswahili–Kiingereza Katika Mtaandao / Online Swahili–English Dictionary*. Available from: <http://africanlanguages.com/swahili> [01/09/2016].
- Hurskainen, A.** 2008. SALAMA Dictionary Compiler - A Method for Corpus-Based Dictionary Compilation. *Technical Reports in Language Technology*, Report No 2, 2008. Available from: <http://www.njas.helsinki.fi/salama/salama-dictionary-compiler.pdf> [01/09/2016].
- Kiango, J.G.** 2000. *Bantu Lexicography: A Critical Survey of the Principles and Process of Constructing Dictionary Entries*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Muaka, L. and A. Muaka**. 2006. *Tusome Kiswahili. Let's Read Swahili: Intermediate Level*. Madison Wisconsin: NALRC Press.
- McGrath, D. and L. Marten**. 2003. *Colloquial Swahili: The Complete Course for Beginners*. London: Routledge.
- Piotrowski, T. and Z. Saloni**. 1992. *Nowy słownik angielsko–polski polsko–angielski* [New English–Polish, Polish–English Dictionary]. Warszawa: Editions Spotkania. Electronic version available from: http://clip.ipipan.waw.pl/Nowy_slownik_angielsko-polski [01.09.2016].

- Prinsloo, D.J.** 2015. Corpus-based Lexicography for Lesser-resourced Languages — Maximizing the Limited Corpus. *Lexikos* 25: 285–300.
- Stopa, R. and B. Garlicki.** 1966. *Mały słownik suahilijsko-polski i polsko-suahilijski* [A small Swahili-Polish and Polish-Swahili Dictionary]. Warszawa: Wiedza Powszechna.
- TEI Consortium (Eds.)**. 2015. *TEIP5: Guidelines for Electronic Text Encoding and Interchange*. [2.9.1] Available from: <http://www.tei-c.org/Guidelines/P5/> [10/11/2015].
- Wójtowicz, B.** 2013. *Słownik suahili-polski*. [Swahili-Polish Dictionary]. Warszawa: Elipsa.

The Corpora of China English: Implications for an EFL Dictionary for Chinese Learners of English

Lixin Xia, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, Guangzhou, P.R. China (200211103@oamail.gdufs.edu.cn)*

Yun Xia, *Department of International Studies, Nanfang College of Sun Yat-Sen University, Guangzhou, P.R. China (344192304@qq.com)*

Yihua Zhang, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, Guangzhou, P.R. China (bilex@mail.gdufs.edu.cn)*

and

Hilary Nesi, *School of Arts and Humanities, Coventry University, Coventry, United Kingdom (aa3861@coventry.ac.uk)*

Abstract: The localization of the English language in China has brought about a distinctive English variety which has come to be known as China English. Recently, several corpora of China English have been or are being built; these will help us to identify the established linguistic features of this variety, and should greatly facilitate the compilation of an English dictionary for Chinese learners of English who operate in Chinese cultural contexts and need to refer to China-specific concepts and phenomena. This paper briefly introduces China English corpora in terms of their principles, components and current status, and explores their potential application to lexicographical projects in terms of lemma selection and inclusion, definition extraction, and glossing and labeling systems.

Keywords: CHINA ENGLISH, CORPUS, DEFINITIONS, ENGLISH-CHINESE LEARNER'S DICTIONARY, ENGLISH VARIETIES, GLOSSES, HEADWORD SELECTION, LABELS, LEXICOGRAPHY, LINGUISTIC FEATURES OF CHINA ENGLISH, LANGUAGE CONTACT, WORLD ENGLISHES

Opsomming: Die korpora van China-Engels: Implikasies vir 'n EVT-woordeboek vir Chinese aanleerders van Engels. Die lokaliserings van die Engelse taal in China het 'n eiesoortige Engelse variëteit tot stand gebring wat as China-Engels bekend geword het. Onlangs is of word verskeie korpora van China-Engels saamgestel; dit sal ons help om die gevestigde taalkundige kenmerke van hierdie variëteit te identifiseer, en dit behoort die samestelling van 'n Engelse woordeboek vir Chinese aanleerders van Engels wat binne die Chinese kulturele kontekste optree en wat moet verwys na konsepte en verskynsels spesifiek tot China grootliks te vergemaklik. Hierdie artikel stel kortliks China-Engelse korpora bekend volgens hul beginsels, komponente en huidige status, en ondersoek hul moontlike toepassing op leksikografiese projekte

rakende lemmaseleksie en -insluiting, die onttrekking van definisies en stelsels vir glossering en etikettering.

Sleutelwoorde: CHINA-ENGELS, KORPUS, DEFINISIES, ENGELS-CHINESE AANLEERDERSWOORDEBOEK, ENGELSE VARIËTEITE, GLOSSE, LEMMASELEKSIE, ETIKETTE, LEKSIKOGRAFIE, TAALKUNDIGE KENMERKE VAN CHINA ENGELS, TAALKONTAK, VARIËTEITE VAN WÊRELD-ENGELS

1. Introduction

Ge (1980) seems to have been responsible for coining the term "China English". He did not define it precisely, and initially scholars debated whether it existed as a phenomenon distinct from 'Chinese English' or 'Chinglish'. Even now, China English has not been universally acknowledged as a variety, but it is becoming a topic of growing interest to scholars in China and internationally, and it is generally believed to present its own phonological, lexical, syntactic, pragmatic, and discourse features (Cannon 1988, Jiang 2002, Du and Jiang 2001, Xu 2008, Xu 2010b, Bolton 2006, Kirkpatrick 2010, Yu and Wen 2010).

Ge (1980) gave examples of China-specific phenomena and concepts which he believed to be representative of China English, such as *Four Books (Si Shu)*, *May Fourth Movement (Wusi Yundong)*, *ideological remoulding (sixiang gaizao)*, and *four modernizations (sige xiandaihua)*. His ideas were expanded by Li (1993: 19) who went on to describe China English as a variety of English which has Standard English at its core, but is used to express China-specific referents and concepts in the context of Chinese culture. The extent to which it is influenced by the Chinese language remained unclear.

According to Xu (2006, 2008, 2010a, 2010b), China English should be defined as:

a developing variety of English, which is subject to ongoing codification and normalization processes. It is based largely on the two major varieties of English, namely British and American English. It is characterized by the transfer of Chinese linguistic and cultural norms at varying levels of language, and it is used primarily by Chinese for intra- and international communication. (Xu 2010a: 1)

At the phonological level, some patterns in China English are predictable, and are handed down from one generation to another. For example, Chinese users are confused by the stress rule in English (Pan 2005: 49) as Chinese is a tone language. Jiang (2002: 12) notes that for the verb + adverbial pattern, the stress is generally on the verb, so speakers of China English say *SIT down* and *STAND up*, rather than *sit DOWN* or *stand UP*. To an English native speaker the former might sound more like an order, and they would generally use the latter forms except when emphasizing the command. He and Li (2009: 72) report the following phonological features of China English: replacement of /θ/ with /s/

and /ð/ with /z/, insertion of final [ə], avoidance of weak forms for function words, and general lack of voiced fricatives.

It is common for lexical items to enter the vocabulary of Standard English from Chinese; for example Chinese expressions accounted for 20% of the 2,000 new words and phrases added to English in 2005, according to the Global Language Monitor (Radtke 2007). Most of these are accepted into English because of the need to express key features of the physical and social environment in China peculiar to Chinese culture. However we can imagine that such expressions are established within China English some time before becoming fully accepted into Standard English.

The following non-standard forms, identified by Kirkpatrick (2010) in the Asian Corpus of English, are widely attested in China English.

1. The non-marking of the third person singular with "-s"
2. Interchangeability of the relative pronouns, "who" and "which"
3. Flexible use of definite and indefinite articles
4. Extended use of "general" or common verbs
5. Treating uncountable nouns as plural
6. Use of a uniform question tag
7. Use of demonstrative "this" with both singular and plural nouns
8. Use of prepositions in different contexts (Kirkpatrick 2010: 8)

Xu (2010a: 69) identifies additional syntactic features which are typical of China English:

1. Adjacent default tense (So, I have to sign a contract before I came here for the study.)
2. Null-subject/object utterances (A: When you're at home, you talk in Fujian dialect with your parents? B: Sometimes yes, sometimes__use Putonghua.)
3. Co-occurrence of connective pairs (Although it is not as big as Beijing, but I like it.)
4. Subject pronoun copying (Some of my college classmates they like to dress up very much.)
5. Yes-no responses (A: You do not want to make a living by playing guitar on the street. B: Yes. Of course not.)
6. Topic-comment (I think this society, the people get more and more practical.)
7. Unmarked OSV (Probably some other kind of jobs I also want to try.)
8. Inversion in subordinate finite wh-clauses (I really don't know what is international English.)

But some of Xu's features listed above are at the discourse level, such as yes-no response and co-occurrence of connective pairs.

At the discourse and pragmatic level, Chinese characteristics occur in China English in almost every aspect. According to Kirkpatrick (2002), the

frame-main sequence in Chinese is expected to be transferred into China English. As one might expect, China English is greatly influenced by sociopragmatic considerations relating to political status and political life, law and social order, power and hierarchy (Xu, 2010 a). He also reports some pragmatic features of China English, for example, through the analysis of culturally unique expressions in written texts.

2. The Teaching of China English in China

English is widely taught in China from primary schools to universities and colleges; a national language survey conducted at the turn of the century found that 390.16 million Chinese people had learnt English at some stage in their education (Wei and Su 2012).

Currently the pedagogical model for English in China adheres to native speaker norms, although this is an unattainable goal because, as Medgyes (1992: 342) puts it, "non-native speakers can never achieve a native speaker's competence". Moreover, in the age of globalization the goal should be "to use English successfully in lingua franca or multilingual contexts, rather than to acquire standard forms and nativelike proficiency (Kirkpatrick 2011: 221)".

It is for this reason that researchers and educators such as Hu (2007), Pan (2005), Wu (2014) and He and Li (2009) argue that China English should be taught at schools alongside British or American English. The obvious advantage for this new pedagogical model of English is that the inclusion of selected linguistic and pragmatic features of China English will provide learners with more opportunities to express ideas specific to Chinese culture. The latent advantage is that it helps learners of English in China shake off the fetters of 'Standard English', and develop communicative competence in the context of cross-cultural communication between users of different Englishes.

3. Corpora of China English

China English corpora are usually made up of texts written or translated by Chinese authors. The following sections provide a brief review of each of the major corpora of this type.

3.1 The China English Corpus

In 2009, Li Wenzhong initiated the construction of the China English Corpus (Sun and Lv 2010). This corpus consists of written texts, mostly taken from newspapers, magazines, academic journals, dissertations, literary writings, government documents and web pages. To keep the original form of the data, non-textual data such as photos and graphs were also input into the corpus

through links provided in the annotation so the users can retrieve and extract original texts. The data is given linguistic tags.

Before building the China English Corpus, a pilot project was launched to build a corpus of China English News Articles (CENAC). This pilot corpus contained 2,316 texts which are classified into twelve categories: life, politics, economy, law, environment, education, science, sports, communication, advertisement, agriculture, and industry. All the texts came from news articles published in the *China Daily*, *Beijing Weekend*, and *Shanghai Star* in 2002. It has 1,281,498 tokens.

According to Li Wenzhong, the China English Corpus was built by the corpus research and development team of Henan Normal University (Li 2012: 336). However, literature on it is extremely rare except for some papers introducing the initial annotation system and using CENAC to study the word clusters in news articles (Li 2007, Liu 2008, Sun and Lv 2010, Gao 2006). Moreover, the size and status of this China English Corpus are unknown. It is not accessible for public use. Fang (2012: 113) explains that "the corpus is still under construction."

3.2 *China Daily* Corpus of China English

The China Daily Corpus was built by Li Dingshun in 2011 (Li and Liu 2011). All the data came from the *China Daily* and *China Daily Hong Kong Edition*. It has a total of 156,864 texts, 120 million tokens, and 350,000 words. The corpus covers the following 10 domains: politics, economy, culture, life, sports, education, entertainment, environment, technology, and military.

This corpus has been completed, but it is only representative of the *China Daily*; besides, the Hong Kong component of the corpus will represent a rather different cultural environment from that of mainland China.

3.3 The Comparable Corpus of China English

This corpus is being built by Fang Chengyu (Fang et al. 2012) in Hong Kong for the purposes of comparing China English and British English. The corpus is made up of three modules: media English, academic English, and personal microblogs. At present, the first module is complete, and enables comparison of Chinese and British English as used in newspapers, magazines, and web pages, across the following domains: news reports, editorials, life, culture, and business.

The Chinese Media English Corpus has a total of 2,226 texts. Each text has been cut so that only the first 600 words have been included, so the sub-corpus contains 1,125,627 tokens. The British Media English Corpus was compiled in the same way as the Chinese one and has 1,133,643 tokens. The second module representing academic English is currently being built.

This comparable corpus is particularly useful for comparison of the linguistic features of Chinese and British English. By including both Chinese and British English in the corpus, it should reveal aspects of China English that the other corpora cannot. However, the size of the corpus is modest compared with other modern corpora.

3.4 China-Related English Corpus

This is a project led by Zhang Yihua at Guangdong University of Foreign Studies supported by the National Social Science Fund of China. The project began in 2013.

The corpus is made up of two sub-corpora: China-related English by Chinese native speakers and China-related English by English native speakers. For the first of these, data was collected from English media sources in China, such as the *China Daily*, *Beijing Review*, *China Today*, *Global Times*, and from novels and stories written in English by Chinese people or translated from Chinese into English by Chinese people. For the second sub-corpus, data was collected from news and reports about China and Chinese culture in British and American media, such as the *BBC*, *Reuters*, *VOA*, *CNN*, *ABC* and the *New York Times*, and from books and stories on China and Chinese culture written and published in the UK and the USA, and from works written by Chinese but translated into English and published in the UK and the USA.

At present, more than 150-million tokens have been collected and processed, and a pilot version is available for public use (<http://202.116.192.97/#>). The corpus is scheduled to be completed by 2017.

This corpus is by far the largest and most comprehensive one of this type in terms of its size and genres it covers. It may be the only one of this type that is designed and constructed with lexicography as its primary aim. To the designers of this corpus, China-related English means English on China and Chinese culture no matter whether it is written by Chinese people or English native speakers. From the components of the corpus, it can be inferred that it enables comparison of China English and 'standard' British or American English at different levels. However, one may argue that it is difficult to identify 'English native speakers'. Just because they write in English in Britain or the USA doesn't necessarily mean they are native speakers of English.

All the corpora mentioned in this section are made up of written texts only; and none of them contains spoken data. It is generally agreed that varietal differences are much more evident in speech than in writing. Therefore, corpora of this type should be made up of both written and spoken data. Moreover, the issue of representativeness should be tackled carefully as China is so big.

4. Corpora of China English and Lexicography

Dictionary production can be seen as a four-stage process: deciding the target

user, collecting linguistic data, analyzing linguistic data, and synthesizing dictionary text (Rundell 2010: 367). At the first stage, it is important to note that dictionary users may be composed of speakers of different Englishes in cross-cultural communication. The globalization of English has caused "lexicographers to think about the adequacy of their products for a diverse and complex world of English-users (Benson 1997: 126)". To meet the needs of the users, an English dictionary should record and describe the actual use of world Englishes rather than solely that of British or American English.

As a result, some regional English dictionaries or dictionaries of a local English variety have been made. *The Macquarie Dictionary* (1981) was the first comprehensive dictionary of Australian English (Delbridge and Butler 1999), and the most innovative aspect of its third edition (1997) is that it includes words from Southeast Asian English "which have never been in a dictionary before" (Butler 1997: 103). The *Dictionary of Caribbean English Usage* (1996) employs a specially-designed labelling system to denote the social or grammatical register of Caribbean English (Allsopp 1996). *Cosmo Dictionary of Indian English* (Yule et al. 2006) provides a ready reference for words and phrases of Indian English not included in other English dictionaries. *A Dictionary of South African Indian English* (Mesthrie 2010) records the use of Indian English in South Africa, and it contains around 1700 entries. These dictionaries are all innovative because they included non-western varieties of English.

It can be predicted that more and more national English dictionaries of local varieties will be joining *the Macquarie Dictionary* in the foreseeable future. It might not be the right time to compile a dictionary of China English. However, the EFL dictionary specifically designed for Chinese learners of English should take the lead in including and describing words and use of China English in order to meet their needs in expressing concepts and phenomena specific to China and Chinese culture. Furthermore, the dictionary should provide information on linguistic and cultural features of China English systematically in its definitions, glosses, labels, etc.

4.1 Criteria of Lemma Selection and Inclusion

"There is general consensus among scholars and researchers of 'China English' that the defining feature of 'China English' is its lexicon (He and Li 2009: 72)." That being the case, it would be natural for a dictionary user to expect a substantial amount of Chinese words in an English dictionary. However, it is not the case that there has been substantial, or even adequate inclusion and description of them. In the process of writing a dictionary, any item from an English variety in Kachru's expanded circle (1985), like China English, has to fight hard in order to enter the English dictionary. One may argue that it is because of their low frequency in a corpus that they are not given the right to enter a dictionary as a headword since all major English dictionaries are based on corpus data. The inclusion criteria themselves seem to be scientific. But if

the corpora on which these English dictionaries are made do not include texts from China, problems arise. The outcome would be the loss of certain words of Chinese origin from dictionaries. For example, the author extracted some of the texts from the first sub-corpus of the China-related English Corpus for the purpose of this study. It has about 2 million tokens, and will be used as an acting corpus of China English (hereafter ACCE) for this study. A search of the English word *gaokao* from Chinese in the ACCE got 57 concordance hits. But there are no hits of the word in the BNC which has 112,289,776 tokens (data searched on 4 July, 2016).

These items are also excluded because of long-standing lexicographic traditions. In the fields of linguistics and lexicography, British English and American English are looked up to as the standard. But many varieties of English especially those from the expanded circle have not been accorded the status of a legitimate variety of English. They are regarded as non-standard Englishes. Although contemporary lexicography aims to describe not prescribe, a dictionary has an inherent nature of standardization to the extent that once any of a linguistic item is included in a dictionary, the dictionary raises "it to the status of the norm (Svensén 2009: 67)". Therefore, it's safer to exclude these 'non-standard' items from a dictionary.

Although the mainstream English learner's dictionaries such as the 'Big Five' claim to be specifically designed for foreign learners, few lemmas from Chinese are included in them. There are only 13 words borrowed from Chinese in the *Oxford Advanced Learner's Dictionary* (8th edition, 2010, CD-ROM version) (*cheongsam, mahjong, samfu, taipan, wok, wonton, yang, yin, fengshui, dim sum, kung fu, foo yong, and t'ai chi ch'uan*), although it claims to include as many as 184,500 headwords. That is to say, words of Chinese origin only account for 0.007% of all the headwords. Benson (1997: 133) notes that these learners' dictionaries not only contain fewer references to China than their larger counterparts such as the OED, but also that these references form a much smaller proportion of the total number of entries — 0.12% for OED 2, and 0.08 for OALD4. The OED online version (<http://www.oed.com>) now has a total number of 600,000 entries. Through the advanced search, one can get 251 entries of Chinese origin (data retrieved on 3 July, 2016), which makes up 0.041% of the total entries. Nearly 20 years passed from Benson's work, but the situation remains unchanged.

Dictionary writers are extremely cautious about admitting new entries to their dictionaries. They have more reasons not to include an alien word from Chinese since it is in the expanded circle of the English family (Kachru 1985). However, Chinese English learners are brought up in the context of China, and they have to express something typical of China and Chinese culture. If a dictionary doesn't include them, they will not know how to express these culture-specific ideas. That may account for the incompetence of Chinese learners when they want to express some specific Chinese referents and concepts.

It therefore seems a good idea to create an EFL dictionary for Chinese learners of English, including more entries relating to China-specific concepts. A search in the China-Related English Corpus can generate hundreds and

thousands of hits for words and expressions with Chinese characteristics. We find that most of the single words are in their Pinyin form, such as *laojiao* and *gaokao*. However the multi-word expressions are mostly in their anglicized form, such as *educated youth*, *four virtues*. Since these expressions are heavily culture-bound, it is almost impossible to guess their meanings from the meaning of their components. Therefore, both types of lemmas need to be included and explained in the EFL dictionary for Chinese users. In summary, the criteria of inclusion should take into account the corpus evidence from corpora of China English for a word's frequency and use.

4.2 Defining Styles

The major contribution of corpora of China English to lexicography is that they enable us to reexamine the ways in which lemmas from Chinese are defined in English dictionaries. In general, these lemmas are not accurately defined in them. One can even say that they are often vague, and sometimes even misleading. For example, a native speaker of Chinese may fail to figure out what the headwords refer to in the real world by reading the definitions of the following two entries: *cheongsam*, and *samfu* in OALD8. The lemma *cheongsam* is defined as "a straight, tightly fitting silk dress with a high neck and short sleeves and an opening at the bottom on each side, worn by women from China and Indonesia". Native speakers of Mandarin Chinese have a different picture of the *cheongsam*, as shown in Figure 1. In mainland China it is understood to be a long and loose-fitting piece of clothing which covers the body and reaches the ground; formerly in China *cheongsams* were especially worn by educated men, and were a sign of rank because poor, illiterate people couldn't afford them.

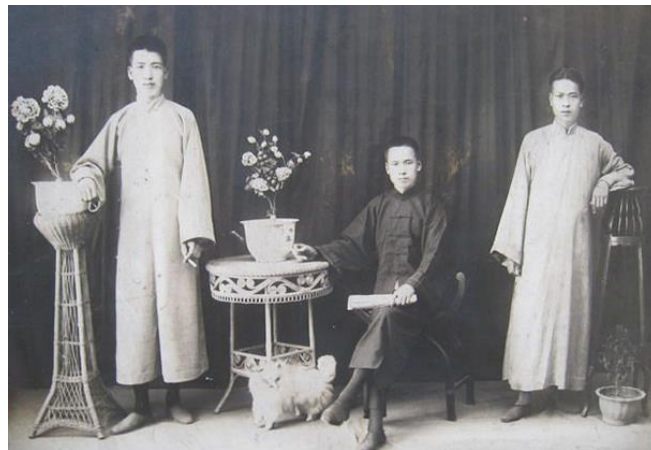


Figure 1: An illustration of *cheongsam*

The definition in OALD8 is completely different, and refers to an item of clothing that is known as a *qipao* in Mandarin Chinese (see Figure 2), and is translated as such in the bilingualised version of OALD8. The translator might have spotted the mismatch between the headword and the definition, but the error has not been corrected.



Figure 2: An illustration of *qipao*

The word form *cheongsam* is Cantonese, and the dress originated from Shanghai in the 1920s, as an exclusively traditional gown for women. For Cantonese speakers in the areas of Guangzhou, the word *cheongsam* usually constitutes a jacket with long sleeves, not necessarily a long dress covering the whole body. In Hong Kong, the word *cheongsam* refers to a dress for both women and men. It seems that OALD8 adopted the signifier of the concept in Cantonese and the signified object from Shanghai areas. No regional uses were shown in the definition (Xia and Zhai 2016). That's the reason why speakers of Mandarin Chinese would feel puzzled by the definition in OALD8.

The headword *samfu* is defined as "a light suit consisting of a jacket with a high collar, and loose trousers/pants, worn by Chinese women" in OALD8. A speaker of Mandarin Chinese might not be able to figure out what a *samfu* is by reading the definition. Unfortunately, this entry has not been included in the bilingualised version of OALD8. Maybe the translator found the problem, and dropped the entry from the original dictionary deliberately.

As mentioned above, these words and expressions from Chinese only account for an extremely small proportion of the entries in the English dictionary. So it is not a big problem for Chinese English learners even if they are not defined accurately. They have bigger problems with common English words and expressions. For example, the first sense of the lemma *house* in OALD8 reads:

house¹ *noun* ... 1 [C] a building for people to live in, usually for one family:
The family live in a two-storey house...

In Britain and America, a house is often a detached, semi-detached or terraced building with two or three levels for one family to live in. The definition is concise and good for them, but not for Chinese English learners. Here is a concordance line from ACCE of China-related English Corpus:

House prices in Beijing are absolutely ridiculous. When a young couple purchases a house, parents and grandparents from both sides need to help out. The collective effort of four families is required to support a young family's decision to buy a house. Three generations of savings are thus exhausted in buying a single house.

House prices are too high for a young couple to buy even a small home in cities in China. Their parents and grandparents have to help them buy one with their savings. They are buying an apartment in a high building for the young couple, not a two-storey house. They can't afford an apartment, let alone a big house. Therefore a house in China English as shown in the above example is a set of rooms in a high building for people to live in or for rent, something like an apartment or a flat as shown in Fig. 3.



Figure 3: An illustration of *house* for people in the city

For older Chinese people, especially those from the countryside, a house is a small building with a tiled roof, on one level as described below in the concordance line from ACCE, although more and more people in the rural areas are building multi-storey houses now.

Traditionally, a typical rural residence is a three-room, one-storey house with a spacious courtyard where trees and flowers can be planted.



Figure 4: An illustration of *house* for older people in the countryside

The closest equivalent to a house is probably *bieshu* as shown in Fig. 5. A *bieshu* in China refers to a detached or semidetached building with two or three levels for one family, especially a rich one.



Figure 5: An illustration of *bieshu*

More concordance lines of 'house' can be found in ACCE, some of which are listed below:

1. a bird's-eye view of past and future: squat tile-roofed **houses** built around courtyards
2. ain. Now he has bought a **house** for the couple and hopes that one day they can wed.
3. My wife and I bought a modest **house** in suburban Shanghai a few years ago, and on
4. in the capital. "I'm getting ready to buy a **house** in Beijing, a big house," Mo said,
5. prospective husbands are vetted about whether they own a **house**, and preferably also
6. is already pregnant again and he is building a new **house** on the site of his old house
7. ts who hover over their children, "doing their work for them and **buying** them houses.
8. is very poor, his house is very shabby, but he always does good things for the vil-lager

9. We can't afford to rent an apartment if we want to buy our own **house**," Liu said.
From
10. takes a reporter to see the former village leader's home — a three-story **house** whose
11. Wang Hong sat inside a small stone-and-brick **house** with one of China's aging
musical
12. in front of his traditional brick house a few yards away, Kong Qingyu, 77, a
farmer, spo
13. He revisited his clan village and his **house**, now abandoned, and met an elderly
man

From the contexts it may be inferred that the word 'house' in concordance lines 1, 6, 8, 10, 11, 12 and 13 refers to a building in the countryside. Most of the houses are single-storey buildings except in line 10 where the reference is to a three-storey house. The owner of the house is the village leader. However, the word 'house' in concordance lines 2, 3, 4, 5, 7 and 9 probably refers to an apartment or flat in a high building in the city, not a two or three-storey house. Although it is not clearly indicated in the sentences, one can infer it through the contexts and background knowledge. As house prices are ridiculously high in China, most citizens cannot afford a house, especially in big cities like Shanghai (concordance line 3) and Beijing (concordance line 4). Most urban residents live in an apartment or a flat.

From the above analysis, we can see that the definition of the English word *house* will probably form different images in the mind of Chinese users. Therefore, to help Chinese learners better understand the meaning and use of the English lemmas, they should be defined in a way that accords with Chinese learners' cognitive schema. The most important thing is to tell the dictionary users the difference between China English and British English. So the definition of *house* could be like this:

house *noun* [C] **1** a building for people to live in, usually for one family **SYN.**
villa **2** In China a house is a small building that has a tiled roof and only
one level in the countryside, or a set of rooms in a high building for people
to live in or to rent in the town **SYN. apartment, flat, bungalow, home ...**

Compared with the definition in OALD8, a new sense of the headword *house* is given, depicting clearly the referent in the real world in China English. In other words, the English word *house* in China English most probably denotes two kinds of buildings: a single-storey building and a multi-storey high building.

By reading the definitions, Chinese English learners would recognize the difference between the extensions of the word *house* in 'standard' English and China English. The word *house* in China English has a wider extension than that in 'standard' English. Thus, when they use the word *house* in different regional contexts, they would note its regional use. For example, when they mention a house in China, they could use a specific word such as *apartment, flat, home, villa, bungalow*, etc. to denote it instead of a more general word *house*.

The inclusion and definition of English lemmas with Chinese characteristics will undoubtedly enrich the English language since more English speakers will become familiar with them and use them in their communications.

4.3 The Glosses

Apart from lemma inclusion and definitions, glosses can be used to remind dictionary users of the characteristics of China English. China English has its particular linguistic features. For various reasons, some headwords have more than one spelling. However, of the 13 headwords of Chinese origin, OALD8 only gives a variant for *t'ai chi ch'uan* as below.

t'ai chi ch'uan ... (also **t'ai chi**) *noun* [U] (*from Chinese*) a Chinese system of exercises consisting of sets of very slow controlled movements ...

T'ai chi is often used as the shortened form of *t'ai chi ch'uan* both in the West and in China (Xia and Zhai 2016). However, this gloss is problematic since the word *t'ai chi* is also a philosophical term in ancient China, meaning the quintessence of the universe, or the Supreme Ultimate. The word *t'ai chi ch'uan* refers to a Chinese martial art and a system of calisthenics.

Data from corpora of China English also shows that the shortened form is used almost exclusively to refer to the sports activities. No concordance lines of *t'ai chi* are found to refer to the philosophical thinking. The reason for this might be that the sports activity is more popular among people than the philosophical thinking.

1. after participating in 12 weeks of **t'ai chi** training. The study's lead a
2. esized that the dancelike style of **t'ai chi** might be an appropriate way
3. for 12 weeks for those learning **t'ai chi**. A control group, whose mem
4. y, those who participated in the **t'ai chi** exercises reported improvem
5. e attractive to them." A Beijing **t'ai chi** master agreed to conduct the
6. than half of them continued the **t'ai chi** class after the study was over
7. year-old mother-in-law studied **t'ai chi** in China for two weeks and
8. rditional Chinese calligraphy, **t'ai chi**, and martial art to better und
9. classes were taught by several **t'ai chi** masters included Zeng Haixia
10. where many still practise their **t'ai chi** in the misty early mornings, w
11. n joggers pant by still, and the **t'ai chi** practitioners silently exercise a
12. groups of pensioners practising **t'ai chi**, Chinese children can be seen p
13. likely candidate to be practising **t'ai chi** in a glass-fronted shop in Man
14. dicine. She also recommended **t'ai chi** as a means of relaxation. "Every

In this situation, it is advisable to add a note to the entry in an EFL dictionary for Chinese learners of English, telling dictionary users the original meanings of the variant *t'ai chi*. The entry could be rewritten as below:

t'ai chi ch'uan ... (also t'ai chi) *noun* [U] (*from Chinese*) a Chinese system of exercises consisting of sets of very slow controlled movements ...

Note: The word form *t'ai chi* as a philosophical term means the quintessence of the universe, but it is used more frequently as an informal form of *t'ai chi ch'uan*.

With the help of a corpus of China English, lexicographers can get all the variants of a lemma and their frequency. For example, through data retrieval in ACCE, we get all the spellings of *kung fu* (961), *kungfu* (99), *kongfu* (8), and *gong fu* (8). Based on the word frequency, *kung fu* should be listed as the headword, *kungfu* as the variant, by way of a gloss as shown below:

kung fu ... (also kungfu) *noun* [U] (*from Chinese*) a Chinese sport people do to fight and stay healthy. Kung fu is a martial arts. ...

In the above entry, *kungfu* is listed as a variant of the headword *kung fu* since it is less frequently used, and the word *also* in the gloss indicates this. The word forms of *kongfu* and *gong fu* are not listed as a variant because of low frequency.

At the grammatical level, typical characteristics of China English can be indicated by glosses. Because of negative transfer from their first language, speakers of China English would naturally use the collocation structure of their mother tongue instead of a 'correct' English one. Yu and Wen (2010) made a comparative study of the collocation patterns of evaluative adjectives and nouns in China English and British English with data from CENAC (China English News Articles Corpus) and the news part of BNC. They found that the collocation patterns of adjectives and nouns in China English are greatly influenced by Chinese. For example, the word *great* co-occurs with *change(s)*, *progress*, *contribution* and *achievement(s)* more frequently in China English than in British English with a standardized frequency of 24:0.5, 11:0.5, 10:0.4 and 10:1 respectively. They argued that this collocation pattern was transferred directly from Chinese. Therefore, this collocation pattern should be treated and represented in an EFL dictionary for Chinese learners of English.

Semantic prosody also needs to be considered in an EFL dictionary. Sinclair (1991: 121) noted that "many uses of words and phrases show a tendency to occur in a certain semantic environment. For example the word *happen* is associated with unpleasant things — accidents and the like". A search of *happen* in BNC found that the following words co-occur with it: *accident*, *incident*, *attack*, *crash*. However, during the use of English in China over the years, it has acquired a unique flavor of its own. As a result, some words and phrases have a different semantic preference from their original use in 'standard' English. For example, the word 'happen' was found to be together frequently with positive words in ACCE. Here are some examples:

1. Though we are seeing signs that a **pick-up** might happen in the US economy, it certainly seems that China is not going to return to the heady days of 20% export growth a year.

2. You know, if **lots of lucky things** happen, we can get out relatively painlessly. But I think the probability of that is very, very low.
3. There needs to be **wide-spread application of these best practices** across the board in all new building and construction projects. I am looking forward to seeing that happen.
4. But over time, and with the help of business associations that helped firms develop their own capabilities, these firms evolved, became more sophisticated, more profitable, invested in modern production and human resources management techniques ... and eventually, all of these abuses went away. I expect **the same** can happen in China.
5. Frank Sha, a senior consultant with Zou Marketing, a sports branding company based in Shanghai, says that demand is strong among Chinese consumers for big events like Saturday's soccer match. What is lacking is the people to make **such events** happen regularly.
6. It hopes to launch the rover-releasing moon probe in about two years. Chinese experts believe **a moon landing** will happen in 2025 at the earliest.

The words in boldface in the above examples are generally associated with pleasant things, such as a pick-up in economy, lucky things, big events and a moon landing. It, therefore, can be concluded that the word *happen* tend to co-occur with words with both positive and negative associations in China English.

4.4 Labelling System

Due to the cultural differences between Chinese and Western culture, the same concept may have different connotations. These differences can be indicated in the EFL dictionary by means of a label or a usage note. But labels used for the English lemmas of Chinese origin in OALD8 tend to be imprecise and arbitrary.

For example, the Chinese equivalent of "individualism" has a negative connotation, meaning being selfish or self-centered. But Westerners value individualism positively. Here are some concordance lines from ACCE.

1. As many get richer, loyalty to the party and the authority (itself a product of propaganda) has been gradually replaced by the **individualism**, materialism and the pursuit of wealth.
2. Our traditional values of communalism and social harmony have been replaced by selfishness and **individualism**.
3. Always with foreign devils the same obsession with the self. Bourgeois **individualism**.
4. This world view has emerged not only with the celebration, almost fetishisation, of choice and **individualism**, but also from the collapse of vigorous working-class institutions that might have kept the wider culture more honest.
5. Money worship, hedonism and extreme **individualism** are spreading among some members of society. Looking for profit and forgetting righteousness, a lack of integrity, a loss of moral norms are occurring.

All the uses of *individualism* in the above concordance lines have negative connotations. *Individualism* in China English is described as materialistic in 1, selfish in 2, bourgeois in 3, fetishistic in 4, and hedonistic in 5. The dictionary could tell its users about the difference as shown below:

individualism *noun* [U] 1 the belief that the freedom of individual people is more important than the needs of society or the government: *Capitalism encourages competition and individualism. ...*

Note: In China, *individualism* is considered to be completely contrary to the traditional value of collectivism. However, it is encouraged in Western culture.

In the above entry, a note is added to explain the connotational meaning of individualism in China. In addition, the illustrative example demonstrates clearly the core value of individualism in Western culture.

The headword *taipan* in OALD8 is defined as "a foreign person who is in charge of a business in China" without any labels or notes. However, it is an informal term used during the 19th and the beginning of the 20th century, and is becoming archaic. Therefore, a currency label *old-fashioned* should be added to guard the users against misusing it.

taipan *noun* [C] (*old-fashioned*) a foreign person who is in charge of a business in China. ...

Such a label would mean that the word is no longer used, but its counterparts in the real world exist, and we use a different name (*laoban*) for them. A similar label (*arch.*) can be added in the entry of *cheongsam* to indicate that the word form is still used, but some of its counterparts in the real world no longer exist.

Other subject field and register labels can be added in the entries for headwords of Chinese origin. For example, "*sports*" can be added in the entry for *t'ai chi ch'uan*, "*philosophy*" in the entry for *yang* and *yin*.

5. Conclusion

To sum up, dictionaries of the future would become more regionalized with the globalization of the English language and the construction of corpora of world Englishes. The first drive is the theory of world Englishes. The existing EFL dictionaries for Chinese learners of English are Anglo-American centered no matter whether they are made in China or abroad. British or American English has long been regarded as the only 'standard' in China, while China English is seen as belonging on the periphery. The time has come, however, for long-standing beliefs and lexicographical practices to be questioned and changed. Dolezal (2006: 702) argues that "the lexicography of world Englishes creates a challenge to redefine, change, or reaffirm commonly held linguistic attitudes and ideas".

Under the theoretical framework of world Englishes, lexicographers should take varieties of English into account in their dictionaries. In the case of EFL dictionaries especially made for Chinese learners of English, it is advisable to include and describe words and use of China English in order to meet users' needs. It is generally acknowledged that English in China has distinctive linguistic and cultural features. The description of these in the dictionary can help its users adequately express ideas specific to Chinese culture.

Another drive is the availability of corpora of China English. Computer technology and corpora used in lexicography are generally regarded as an important "revolution in lexicographical resources and practices (Rundell 2010: 367)". Corpora of China English can provide lexicographers with fresh and revealing insights into the possible contents of an EFL dictionary. They enable lexicographers to analyze data on China English and make lexicographical decisions about lemma selection, defining styles and glosses and labels.

But the amount of China English to include in an EFL dictionary remains a matter for debate. A dictionary is always looked up to as a 'standard' or 'norm' by dictionary users, and they may find it difficult to accept an item from an English variety in a dictionary until they are provided with an adequate description of it by linguists and lexicographers. This therefore calls for linguists and lexicographers to conduct basic research into China English.

Fortunately we have seen concerted efforts of linguists, computer scientists and lexicographers in China. Thanks to the newly-developed corpora, linguists are able to investigate systematically the linguistic features of China English. Backed up by their research findings and corpus evidence, lexicographers are able to describe the linguistic and cultural features of words and phrases from China English adequately in their dictionaries.

5. Acknowledgements

This work was supported by a grant from the Humanities and Social Science Planning and Funding office of Ministry of Education of the People's Republic of China under Grant No. 15YJA740048 (China English or Chinese English: A Study Based on China English Corpus), a grant from the National Philosophy and Social Science Foundation of China (Grant No.12AZD115) and a grant from Guangdong Planning Office of Philosophy and Social Science (Grant No.GD14XWW20). We would also like to express our gratitude to anonymous reviewers for their constructive and insightful comments on earlier version of this paper.

References

- Allsopp, R. 1996. *Dictionary of Caribbean English Usage*. Oxford: Oxford University Press.
- Benson, P. 1997. English Dictionaries in Asia: Asia in English Dictionaries. Bautista, M.L.S. (Eds.). 1997. *English is an Asian Language: The Philippine Context*: 125-140. Sydney: The Macquarie Library.

- Bolton, K.** 2006. *Chinese Englishes: A Sociolinguistic History*. Cambridge: Cambridge University Press.
- Butler, S.** 1997. Corpus of English in Southeast Asia: Implications for a Regional Dictionary. Bautista, M.L.S. (Eds.). 1997. *English is an Asian Language: The Philippine Context*: 103-124. Sydney: The Macquarie Library.
- Cannon, G.** 1988. Chinese Borrowings in English. *American Speech* 63: 3-33.
- Delbridge, A. and S. Butler.** 1999. The Macquarie Dictionary, its History and its Editorial Practices. *Lexikos* 9: 152-171.
- Dolezal, F.** 2006. World Englishes and Lexicography. Kachru, B.B., Y. Kachru and C.L. Nelson (Eds.). 2006. *The Handbook of World Englishes*: 694-708. Malden: Blackwell.
- Du, R. and Y. Jiang.** 2001. Jin ershinian "zhongguo yingyu" shuoping ("China English" in the past 20 Years). *Waiyu jiaoxue yu yanjiu* (Foreign Language Teaching and Researching) 33(1): 37-41.
- Fang, C., F. Le and J. Cao.** 2012. Zhongguo yingyu duibi yuliaoku de sheji, jianli he chutan (The Design, Establishment and Preliminary Study of a Comparable Corpus of China English). *Linguistic Research* 2: 113-127.
- Gao, C.** 2006. Shijie yingyu lilun yu zhongguo yingyu yanjiu zongshu (Literature Review on World Englishes Theory and China English). *Guowai waiyu jiaoxue* (Foreign Language Teaching Abroad) 4: 55-60.
- Ge, C.** 1980. Mantan you han yi ying wenti (Random Thoughts on Some Problems in Chinese-English Translation). *Fanyi Tongxun* (Chinese Translator's Journal) 2: 1-8.
- He, D. and D.C.S. Li.** 2009. Language Attitudes and Linguistic Features in the 'China English' Debate. *World Englishes* 28(1): 70-89.
- Hu, X.** 2007. Quanchihua huanjing xia "global English" de fazhan qushi ji dui zhongguo yingyu jiaoxue de qishi (The Development of "Global English" and its Implications for ELT in China in the Context of Globalization). *Guowai waiyu jiaoxue* (Foreign Language Teaching Abroad) 2: 16-23.
- Jiang, Y.** 2002. China English: Issues, Studies and Features. *Asian Englishes* 5(2): 4-23.
- Kachru, B.B.** 1985. Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle. Quirk, R. and H.G. Widdowson (Eds.). 1985. *English in the World: Teaching and Learning the Language and Literatures*: 11-30. Cambridge: Cambridge University Press.
- Kirkpatrick, A.** 2002. *Englishes in Asia: Communication, Identity, Power and Education*. Melbourne: Language Australia.
- Kirkpatrick, A.** 2010. Researching English as a lingua franca in Asia: The Asian Corpus of English (ACE) Project. *Asian Englishes* 13(1): 4-18.
- Kirkpatrick, A.** 2011. English as an Asian lingua Franca and the Multilingual Model of ELT. *Language Teaching* 44(2): 212-224.
- Li, D. and G. Liu.** 2011. China Daily zhongguo yingyu yuliaoku gaikuang (An Overview of China Daily China English Corpus). *Yingyu guangchang* (English Square) 6(3): 58-59.
- Li, W.** 1993. Zhongguo yingyu yu zhongguoshi yingyu (China English and Chinglish). *Waiyu Jiaoxue Yu Yanjiu* (Foreign Language Teaching and Researching) 4: 18-24.
- Li, W.** 2007. Zhongguo yingyu xinwen baokan zhong de cicu (The Word Clusters in China English News Article Corpus). *Zhongguo Waiyu* (Foreign Languages in China) 4(3): 38-43.
- Li, W.** 2012. Yuliaoku biaoji yu biao zhu: yi zhongguo yingyu yuliaoku wei li (Corpus Markup and Annotation: China English Corpus as an Example). *Waiyu Jiaoxue Yu Yanjiu* (Foreign Language Teaching and Researching) 44(3): 336-345.

- Liu, G.** 2008. Xiandai daxing jidu yuliaoku de biao zhu fangfa: yi CEC zhengfu wenjian zi yuliaoku weili (The Annotation Methods of Modern Computer Corpora: A Case Study on the Annotation of the Government Document Corpus). *Shandong Waiyu Jiaoxue (Shandong Foreign Language Teaching Journal)* 126(5): 53-59.
- Medgyes, P.** 1992. Native or Non-native: Who's Worth More? *ELT Journal* 46(4): 340-349.
- Mesthrie, R.** 2010. *A Dictionary of South African Indian English*. Cape Town: University of Cape Town Press.
- Pan, Z.** 2005. *Linguistic and Cultural Identities in Chinese Varieties of English*. Beijing: Peking University Press.
- Radtke, O.L.** 2007. *Chinglish*. Layton: Gibbs Smith.
- Rong, P.** 1991. Chinese English is an Objective Reality. *Jiefangjun Waiyu Xueyuan Xuebao (Journal of PLA Foreign Languages Institute)* 1: 2-7.
- Rundell, M.** 2010. Taking Corpus Lexicography to the Next Level: Explicit Use of Corpus Data in Dictionaries for Language Learners. Zhang, Y. (Ed.). 2010. *Learner's Lexicography and Second Language Teaching: Proceedings of First International Symposium on Lexicography and L2 Teaching and Learning*: 367-386. Shanghai: Shanghai Foreign Language Education Press.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sun, H. and J. Lv.** 2010. Zhongguo yingyu yuliaoku yu yingyong (Corpus Establishment and Application of China English). *Journal of Henan Normal University (Social Science Edition)* 37(1): 219-221.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Wei, R. and J. Su.** 2012. The Statistics of English in China. *English Today* 28(3): 10-14.
- Wu, J.** 2014. EFL shijiao xia gaoxiao yingyu jiaoshi de yuyan taidu yanjiu (A Study of the Language Attitude of College English Teachers from the Perspective of EFL). *Waiyu Jiaoxue (Foreign Language Education)* 35(2): 55-58.
- Xia, L. and L. Zhai.** 2016. Treatment of Entries with Chinese Characteristics in English Learner's Dictionaries: A Case Study of *Oxford Advanced Learner's Dictionary*. Eighth edition. *Kernerman Dictionary News* 24: 12-15.
- Xu, Z.** 2006. Rectifying "Chinese English". Hashim, A. and N. Hassan (Eds.). 2006. *Varieties of English in Southeast Asia and Beyond*: 283-291. Kuala Lumpur: University of Malaya Press.
- Xu, Z.** 2008. Analysis of Syntactic Features of Chinese English. *Asian Englishes: An International Journal of the Sociolinguistics of English in Asia/Pacific* 11(2): 4-31.
- Xu, Z.** 2010a. *Chinese English: Features and Implications*. Hong Kong: Open University of Hong Kong Press.
- Xu, Z.** 2010b. Chinese English: A Future Power? Kirkpatrick, A. (Eds.). 2010. *The Routledge Handbook of World Englishes*: 282-298. London: Routledge.
- Yu, X. and Q. Wen.** 2010. Zhongguo yingyu baozhang zhong pingjia xing xingrongci dapei de bentuhua tezheng (Collocation Patterns of Evaluative Adjectives of English in Chinese Newspapers). *Waiyu Jiaoxue Yu Yanjiu (Foreign Language Teaching and Research)* 254(5): 23-28.
- Yule, H., A.C. Burnell and G.T. Kurian.** 2006. *Cosmo Dictionary of Indian English*. New Delhi: Cosmo Publications.

Peter Gilliver. *The Making of the Oxford English Dictionary.* 2016, xix + 625 pp. ISBN 978-0-19-928362-0. Oxford: Oxford University Press. Price: £40.

This comprehensive publication dealing with the *Oxford English Dictionary* as a project consists of thirteen chapters. It traces the development of the *OED* from its earliest beginnings through the various stages of its realization up to the present day. Two 'interludes' are placed in between the main chronological narrative. The first, following on chapter 2, examines the lexicographical output of the 'sub-editors' under the editorship of Frederick Furnivall during the 1860s and 1870s, and the second, following on chapter 6, describes the various stages of the process of compiling a Dictionary entry. Interspersed in the text of most chapters, but separate from the main narrative, can be found 'capsules', a term and concept which Gilliver borrowed from Norman Davies's *Europe: A History* (1996). For each of these a particular word forms the basis of a certain topic, such as an aspect of editorial policy, which seemed necessary at a specific point. Copious and informative footnotes are given at the bottom of each page of text. Forty-six figures, inserted in the text, show certain persons and buildings, or present specific entries and objects mentioned or discussed. The most important part of the front matter consists of the Preface, while that of the back matter concerns the extensive Bibliography and the detailed Index.

I

The first two chapters form an introduction to the next seven chapters comprising the greater part of the book which deals with the *Oxford English Dictionary* itself.

Chapter 1 'Beginnings: to 1861' mainly treats the announcement in 1857 of the Philological Society's Council that Herbert Coleridge, Frederick Furnivall and Richard Chevenix Trench had been appointed as a Committee to collect 'unregistered' words in English. With these were meant words which do not appear in two important dictionaries published before the Society's foundation in 1842: *A Dictionary of the English Language* by Henry Todd, the officially revised version of Samuel Johnson's Dictionary published in eleven parts between 1814 and 1818, and the *New Dictionary of the English Language* by Charles Richardson, containing the revised material in Samuel Taylor Coleridge's *Encyclopædia Metropolitana* and published in two volumes from 1835 to 1837. However, later in the year a larger scheme for a completely new English Dictionary was proposed, and early in 1858 two supervisory Committees (a 'Literary and Historical' and an 'Etymological') were formed. Todd's and Richardson's Dictionaries were to be abandoned for purposes of comparison. The scope of the Dictionary was extended back to the middle of the thirteenth century, and for the purpose of collecting quotation evidence, the history of English was divided into three periods. For each of these periods readers were provided with a baseline (or 'basis of comparison') in the form of a wordlist.

The absence of an item of vocabulary from the wordlist could then prompt readers to supply a quotation for it. During 1859 Coleridge, who was offered the editorship established a more detailed editorial policy for the Dictionary. A document was drafted by Coleridge and approved by Furnivall and Trench before it was presented to the Society under the title 'Canones Lexicographici'. Before editorial questions could be tackled, a considerable amount of evidence had still to be collected. However, Coleridge had already been seriously ill and died early in 1861.

Chapter 2 'Furnivall's Dictionary: 1861–1875' describes how Furnivall continued with the work started by Coleridge. After having prepared and published three parts of the 'Basis of Comparison' for the three periods for which materials were to be collected, he appealed for collectors of quotations and for sub-editors of the material to come forward. The first group of co-editors, presented with a circular and an example entry, were set to work on certain letters of the alphabet. A Concise Dictionary was planned, Furnivall hoping that the work of the sub-editors would be of such a quality that only light editing would be necessary before publication. The publisher John Murray agreed to undertake the printing of the Dictionary. For the first years the project advanced on two fronts: sub-editing and collecting more material. By 1866, however, it became clear that the progress had slackened. By 1872 there was not much to report. During 1873 and 1874 Furnivall devoted little time to the Dictionary, because of the demands of other projects. Becoming frustrated with the near stagnation of the project, the Rev. George Wheelwright, who in 1862 signed up to sub-edit the letter F, in 1875 had a pamphlet published by the Clarendon Press, Oxford, with the title 'An Appeal to the English-speaking Public on behalf of A New English Dictionary'. To this Furnivall's name and address were added as a point of contact for those wishing to offer help. As a result the Philological Society's Council discussed the possibility of raising a fund to pay an Editor of the Dictionary for ten years. Although Wheelwright died suddenly in December 1875, he spurred the Society into action again with regard to the Dictionary.

II

Chapters 3–9 form a unity, starting with the idea of resuming the compilation of an English dictionary on historical principles and ending with the reprinting of the Dictionary together with the publication of the first Supplement.

Chapter 3 'Manoeuvres: 1876–1879' describes how the Council of the Philological Society approached James Murray to compile a dictionary. After discussions with Macmillan to print the dictionary had failed, the Clarendon Press was approached. The Delegates of the Press commented on the specimen which Murray compiled for Macmillan. Murray then produced a revised one, the Delegates putting forward their views especially on pronunciation and etymology. Negotiations were opened between the Delegates, the Dictionary Committee of the Philological Society and Murray. Murray estimated that it would take three years before he could begin sending copy to the press, and

thereafter ten years overall. A 7,000-page limit for the Dictionary was envisaged. More than three years after the proposal to approach the Clarendon Press, agreements were at last sent to the Philological Society and to Murray.

Chapter 4 'The road to *Ant*: 1879–1884' describes how the material of the Dictionary was transported from Furnivall's house to the 'Scriptorium' erected by Murray at his house in Mill Hill. Material had also to be recovered from some of the sub-editors who left editing in various stages of completion. The material had furthermore to be sorted. Much reading of texts and recording of words had still to be done. The eighteenth century assigned to American readers twenty years earlier, had scarcely been touched. A renewed appeal for readers was directed to the public. Notwithstanding the incomplete sub-editing of A and the deficiencies in the material, compilation of entries began immediately. Editorial issues were discussed at meetings of the Philological Society. The slow rate at which Murray proceeded through the alphabet soon proved that it would be impossible to complete the Dictionary within ten years. Reference was made to corresponding pages in the 1864 edition of Webster's Dictionary. Although this was a useful yardstick to measure the scope of the entries, condensation was difficult and time-consuming. The Delegates now agreed to a limit of 8,400 pages. After typographical details had been finalized, printing could at last begin in May 1882. Part I was ready by 1883. Meddling by the Delegates in the wording of the Prospectus, and especially of the title-page of and the Preface and Introduction to the Dictionary, nearly caused Murray to resign, but finally, in January 1884, Part I of *A New English Dictionary on Historical Principles* was published.

Chapter 5 'The Dictionary divides: 1884–1887' deals with the appointment of a second editor. After the positive reviews and sales of Part I, the Delegates insisted that constant progress should be made by producing two parts a year. The appointment of a second editor was considered. Henry Bradley, who wrote an informed review of Part I was suggested. Murray now moved to 78 Banbury Road, Oxford where the Scriptorium was re-erected. More assistants were engaged. After Bradley had been appointed, he started with a section of Part III, which was only published in March 1887. The finalizing of Part IV was still done on a collaborative basis by Murray and Bradley. However, because of the slow pace, Bradley was made an independent editor. From the beginning of 1888 the Dictionary therefore also acquired a home in Bradley's house in Bleisho Road, Lavender Hill.

Chapter 6 'Storm and stress: 1888–1897' describes the vicissitudes of the Dictionary until the appearance of Volume III. With the publication of Part IV in March 1888, Volume I was complete. Although *New English Dictionary* remained the official title throughout the history of the first edition, the name *Oxford English Dictionary* now became used in advertising. Murray started making good progress, but Bradley's rate of progress remained slow, also because of the scale of his entries. A Committee was formed to look into the slow progress. Consideration was given to a third editor. It became apparent

that the letter C alone would be contained in Volume II, making it necessary to publish a shorter fascicle extending only to the end of C. This proved so popular with the public that the Delegates decided to continue publishing smaller fascicles. In September 1894 Charles Onions became an assistant. During the first half of 1896 the Delegates tried to force a scale of 6:1 of Webster onto Murray and Bradley, but eventually the crisis was solved when a scale of 8:1 was agreed on. Bradley was requested to move to Oxford, North House being offered to him where he settled in July 1896. As space was limited in the Scriptorium, he was given a room in the Press itself. In June 1897 William Craigie was appointed as third editor. When the last section of D was published in July 1897, Volume III, containing the letters D and E could finally appear. The date 1918 was now considered plausible for the completion of the Dictionary.

Chapter 7 'And then there were four: 1897–1915' describes the period after the appearance of Volume III. An impression of continuous progress was effected by the issuing of small quarterly fascicles instead of larger, more irregular Parts. William Craigie, having been given a favourable assessment by Bradley after a three months' trial, was engaged for a further three years, the second of which was to be spent working under Murray's direction. At the end of March 1901 Bradley and Craigie took up residence in the Old Ashmolean Building. Both Bradley and Craigie still struggled to keep their entries satisfactorily compressed. Other publications were also undertaken. Work on an Abridged edition of the *OED* by William Little started in 1903. The *Concise Oxford Dictionary of Current English* edited by Henry and Frank Fowler appeared in 1911. After this Frank Fowler started working on the *Pocket Oxford Dictionary*. Volumes IV and V of the *OED* were completed in 1901. Volume VI appeared in September 1908. After Onions had been given a trial period, he started as independent Editor in January 1914. Progress was slowed down by the War, which brought about a reduction in the editorial staff, and the staff in other departments of the Press. Murray's health now began to fail. The last Dictionary entry, which bears his handwriting is that for *twilight*. He died of heart failure on 26 July 1915.

Chapter 8 'After twilight: 1915–1923' deals with the period after Murray's death. The slower progress was caused, not only by the War, but also, as Bradley, Craigie and Onions pointed out, by a rise in the editorial standards and the amount of material to be assessed. The Armistice brought a return of absent staff, so that work could continue as usual. When the last of the manuscript of S was sent to the printers, the Dictionary was complete from A to T. Craigie started work on U, while Onions moved to XYZ. With the *un-words* scale soon became problematic. However, it was found that the deviation in the scope of *un-words* in comparison with the unabridged Webster of 1847 was not so big. Although Murray's final letter T had ended up with a Webster scale of 12 which was gradually accepted as a new realistic figure, the scale of V proved to exceed Webster considerably. By the second half of 1921 three sections of U, a sizable portion of W and Onions's XYZ were complete. When Little, working on the Abridged, died early in 1922, Onions and his staff were committed to

assist with the Abridged. Henry Fowler, busy on the *Pocket*, collected corrections and additions for incorporation in the *Concise*. The Supplement to the main Dictionary now also needed consideration for the future.

Chapter 9 'Limping over the finishing line: 1923–1933' deals with the completion of the main Dictionary and the Supplement. When Henry Bradley died in May 1923, Craigie had to assist with *W*. It now also became apparent that it might take Craigie longer to finish the unpublished portion of *U*. Although he accepted a professorship at the University of Chicago, he was still able to assist with the Dictionary. But progress slowed down considerably. During Craigie's stay in Oxford in the summer of 1926, he succeeded in finishing *W*, the last copy going to the press in July 1927. The final volume of the Dictionary was published on 19 April 1928, completing the first edition of the *OED*. Craigie secured funds for the extraction of material from Oxford for his planned period dictionaries, and also for an assistant professorship for George Watson in Chicago to help with the *Dictionary of American English*. Still trying to edit the Abridged at the same time, Onions and his assistants working on the Supplement soon fell far behind Craigie who was responsible for only the American vocabulary. There was therefore no satisfactory progress on the Supplement and the Abridged. As a reissue of the Dictionary became necessary, an accompanying Supplement was imperative to prevent criticism labelling the Dictionary as dated. James Wyllie was now given a trial on the Supplement, assigning him to prepare material for the letter *L* to be approved by Craigie, who gave Wyllie coaching during his summer visit to Oxford. The recruit Stella Mills was assigned to work on the Abridged, which Onions struggled to keep going alongside the Supplement. Onions failed to complete *A–K* by the end of June 1931. A separate unit within Onions's staff headed by John O'Loughlin was created, Onions taking a supervisory role. Onions and O'Loughlin undertook *S–T* as well as the remaining portion of *G–K*, and Wyllie, supervised by Craigie, the rest of the alphabet, with the final handover of copy on 1 December 1932, making publication in the spring of 1933 possible. The Abridged renamed the *Shorter Oxford Dictionary* was published on 16 February 1933. The Historical Introduction to the Dictionary written for the most part by Craigie was printed in both the Supplement and the reissue of the main Dictionary. This reissue now bore the official title of *Oxford English Dictionary*. The Supplement was also printed in a slightly different form to accompany the first edition. Both forms included, in addition to the pages containing the entries, a Bibliography of quoted sources and a list of 'Spurious Words'. The scheduled date for publication was 14 November 1933, thus concluding the endeavour of English historical lexicography started in 1877.

III

Chapter 10 'Interregnum: 1933–1957' deals with the period before an editor for the new Supplement was appointed. Material collected for the first edition of the *OED* and the Supplement was used for a number of lexicographical under-

takings outside Oxford, some of which benefited more directly from contribution by *OED* lexicographers, and most of which retained connection with the Press for printing and publication, such as the period and regional dictionaries conceived by Craigie. In addition to the *Oxford Latin Dictionary* on which several editors consecutively worked, the various smaller dictionaries such as the *Concise*, *Pocket* and *Little* were revised. The collection of material was continued by a small group of readers, while Wyllie maintained his database of revisions and additions, referred to as the 'O.E.D. Collections'. Separate from these, Craigie accumulated his own collection of such material. After his memorandum of 1942, called 'O.E.D. Supplement 2', in which he envisaged an expansion of the 1933 Supplement, he continued work on his material for the remainder of the War. Onions also collected new words and meanings meriting a place in the *OED*, especially through his work of revising and expanding the two versions of the Addenda to the *Shorter*, which appeared in 1939 and 1944. By the early 1950s it became clear that a new Supplement was necessary because the *OED* and the 1933 Supplement were becoming dated, especially when compared to the new editions of the *Concise*, *Pocket* and *Little*. Searching for a suitable editor for both the revision of the Supplement and a new edition of the *Shorter* now began. Alan Horsman, who was appointed accepted a professorship however, and was replaced by Robert Burchfield. For the remainder of his time in Oxford Housman carried out preliminary work. For a wordlist of items to be included in the new Supplement he, with the assistance of Raymond Goffin, identified different sources that could yield material. It was anticipated that the greatest part of the new vocabulary would probably be technical. At this stage a Supplement of only a single volume was envisaged. When Horsman started to compile his wordlist, it became evident that constraints of space would require the list to be quite selective, especially with regard to technical vocabulary.

IV

Chapters 11–13 describe the developments regarding the *OED* and the smaller OUP dictionaries from the appointment of Robert Burchfield onwards.

Chapter 11 'Learning to swim (again): 1957–1972' deals with the collection of material for the Supplement until the publication of Volume I. The extent of new vocabulary coming to light far exceeded that which could be included in the new Supplement as it was anticipated to be completed in 1967. By the beginning of 1969 Burchfield started drafting entries. The publication of the third edition of *Webster's New International Dictionary* made him aware that the materials collected for the Supplement compared unfavourably with those in Webster, especially with regard to the profusion of new words in science and technology. By September 1963 Burchfield and his assistants had prepared copy corresponding to about 40 pages of the 1933 Supplement. The new entries, as well as the general style of presentation met with a favourable response. During 1964 and 1965 the drafting of entries continued. There was no

longer any question of confining the text to a single volume. Burchfield's editorial resources had now often to be diverted to other projects, including new impressions or editions of other dictionaries. One of these was to prepare, in collaboration with George Friedrichsen, a publishable text of the *Oxford Dictionary of English Etymology* which was left almost completed by Onions at his death. Redeployment in favour of the Supplement, and to the detriment of the *Shorter* took place during 1966 and 1967. A full-scale revision of the *Shorter* was abandoned in favour of a re-setting of the existing text with revised etymologies and Addenda. Philip Gove's offer to put the quotation files of Merriam-Webster (the publishers of Webster₃) at the disposal of the Supplement provided antedatings for about a third of Oxford's material. Albert Boni's suggestion of publishing a *Compact OED*, which appeared in October 1971 proved to be very successful. When the proofs for Volume I were completed, Burchfield and other members of the Supplement were reassigned to the *Shorter* to ensure its publication in 1973. The appointment of John Sykes as editor of both the *Concise* and the *Pocket* in September 1971 meant that he was less available for work on the Supplement. Volume I, containing the letters A–G, finally appeared in October 1971.

Chapter 12 'Second Supplement to Second Edition: 1972–1989' tells how the next three volumes of the Supplement were completed and how these together with the main Dictionary and the 1933 Supplement were all combined for the second edition of the *OED*. Early in 1973 Burchfield could resume work on Volume II, after he had completed the new Addenda for the *Shorter*, which was now ready to be published. The new *Concise*, which appeared in July 1976 preceded the publication in November 1976 of Volume II of the Supplement, containing only the letters H–N. In addition to the English dictionaries, Burchfield became, from July 1976, also responsible for the planning of the Press's growing group of bilingual dictionaries. The misleading title of a talk by Burchfield on Webster₃ caused G. & C. Merriam to withdraw from its collaboration with the *OED* project, as it was believed that the dictionary had been unfavourably reviewed. In the autumn of 1977 the Oxford Dictionary Department moved from 41 Walton Crescent to 37a St Giles'. Schedules for the Supplement had to be considered alongside those for all the other dictionary projects in progress during this time, causing editors often to be moved from one project to the other. John Simpson and Edmund Weiner were promoted to senior positions on the Supplement staff. Richard Palmer, having completed the *Oxford Latin Dictionary* at last, was also transferred to the Supplement. Publication of Volume III of the Supplement, containing O–Scz, took place on 15 July 1982, together with that of the new *Concise*. For the formation of an ongoing database, the New English Word Service (NEWS) was established. At the beginning of 1984 Weiner was moved over to the *New OED* computerization project. 'Phase 1' comprised the integration of material from the Supplement into the main body of the Dictionary. For the creation of an electronic database containing the integrated text of the *OED* and Supplement, from which it was also

planned to produce a printed version, a machine-readable text had to be prepared. The goals of 'Phase 2' included the publication of the Dictionary in electronic form, and the ongoing revision, updating and enhancing of the text. Simpson, now joining the *New OED* project, took over the running of NEWS. Volume IV of the Supplement, comprising Se–Z, was published on 8 May 1986. During 1986 new publications edited by other OUP lexicographers also appeared. Simpson and Weiner now became co-editors of the *New OED*. Many of the changes to this version of the Dictionary, which had begun to be referred to as the second edition were necessary, for example to bring about consistency in the presentation of the headwords and pronunciation and to incorporate additional senses in the main sense sequences of large entries. New entries prepared by the NEWS team were added to the *OED* database. A two-disc version of *OED*₁ was released in late 1987. Computerization of the *OED* now suggested the possibility of accelerating the editing process of the revised *Shorter*. The official publication of the twenty volumes of *OED*₂ occurred in March 1989.

Chapter 13 'Towards *OED*₃: 1989–' deals with the post-*OED*₂ period. *OED*₂ became available in a micrographically printed 'Compact *OED*₂' in 1991, followed by a CD-ROM version in 1992. New vocabulary items researched and documented by the NEWS team were further made available in the '*OED* Additions Series', of which Volumes 1 and 2, co-edited by Simpson and Weiner, appeared in November 1993. A 'guide' to some of the words which became prominent in that decade, the *Oxford Dictionary of New Words*, compiled by the NEWS editor Sara Tulloch, was published in 1991. Apart from the collection of material for the *OED* database, two key publishing goals were reached. The publication of the *Shorter*, the new edition being called the *New Shorter Oxford Dictionary*, took place in September 1993. The planned intermediate dictionary (provisionally titled the *Oxford Dictionary of the English Language* in 1991) finally appeared in August 1998 as the *New Oxford Dictionary of English*, the 'New' being omitted in later editions. In 1993 the entire Dictionary Department moved from 37a St Giles' into offices in the main OUP buildings in Walton Street. During the year detailed investigation of particular points of policy and practice was carried out by the '*OED* Revision' team. Also established was the *OED* Advisory Committee, an external body of linguists and others who could inform on key matters of policy. The most important policy decisions were those relating to the coverage of earlier periods of English, the scope of the etymologies and the inclusion of material. At the end of 1993 John Simpson became Chief Editor of the *OED*, with Edmund Weiner becoming Deputy Chief Editor to set him free to concentrate on specific key aspects of the revision process. While the work of creating the text of *OED*₃ was continuing, the third volume of the *OED* Additions Series, edited by Michael Proffitt, appeared in July 1997. This was the last such volume as plans for *OED Online* were now sufficiently definite that new entries could become available through online publication. The publication model which was decided on would provide a new release of alphabetical ranges of revised entries every quarter. From March

2000 *OED*₃ material was released online, the first version containing the full text of *OED*₂ and the three Additions volumes, together with a range of 1,000 revised entries. After the retirement of John Simpson in October 2013, Michael Proffitt became Chief Editor, with Philip Durkin as second Deputy Chief Editor alongside Edmund Weiner.

V

These summaries of the different chapters give the bare factual outline of the main historical development of the *OED*. In each chapter this development is described in great detail: the difficulties of acquiring competent staff, the intricate negotiations about remuneration, the struggle to let the projects progress satisfactorily ... To lexicographers these are all too familiar. To non-lexicographers it might be revealing to learn what such a project involves.

Interspersed within every chapter are also found the opinions and attitudes of a wide range of people, from the planners to the editors, about what lexicographical information should be presented in such a historical dictionary. These and criticism of the presentation can be garnered from all the chapters. The two 'interludes' and the thirty-two 'capsules' play a supplementary and enhancing role to these.

The first interlude, 'The work of Furnivall's sub-editors', gives an overview of the editing done before the project almost came to a standstill. Because Furnivall did not do any systematic monitoring of the work, there was a great variation of approach among the sub-editors. Some of them seemingly did no sub-editing. The method Furnivall set out did not differ essentially from the methods of the *OED*'s lexicographers, except that the organization of the senses was to be into a 'logical succession'. Although it was called a 'Concise' Dictionary, it becomes clear from the policies of presentation, scale, inclusion and documentation that it was conceptualized far more extensively. The idea was that the entries should have been brought to such a state of completion that they could have been printed immediately. However, a very small proportion of the alphabet was ever completed by Furnivall's sub-editors to a point at which printing would even have been possible.

The second interlude, 'Method: from quotation slip to published entry', describes the procedure followed in the compilation of entries, which remained essentially unchanged during the work on the first edition of the *OED*, and which was basically employed by each Editor's team. The components of this process were the sorting of the quotations, first alphabetically, and then chronologically for each word, the sub-editing and re-sub-editing comprising the identifying of the individual sub-senses and writing definitions, the drafting of entries and the selection of quotations by assistants and the Editor, finally followed by correcting the proofs and reading the corrected proofs again. All the Editors of the first edition of the Dictionary strove to create a stylistically homogeneous text, succeeding remarkably well. Murray seems to have man-

aged, most of the time, to achieve a higher degree of compression in his portions of the text than the other Editors.

The capsules describe decisions on inclusion which the OED lexicographers had to make, but sometimes also reveal their personal peculiarities. Apart from being informative, these capsules often reflect the enjoyment and fascination of working with words and treating their meanings, as can be seen from the following examples.

The capsule *macadamized* explains how the discovery of a quotation of *unmacadamized*, dated earlier than that of *macadamized*, led to a search for an earlier use of *macadamized* too, which led to a re-revision of the word in *OED Online*. The capsule *inquorate* describes how *quorate* was included in Volume III of the Supplement together with *inquorate*, for which there was no evidence when Volume II was compiled, an anomaly which could only have been rectified in the second edition. In the capsule discussing the almost obsolete *agreeance* compared to the far more common and familiar *agreement*, it is shown how the word gained renewed life when it was used by Fred Durst, the lead singer of Limp Biskit, resulting in its revised entry in *OED Online*. The capsule *rime* relates how Murray's preference for the spelling of *rime* over *rhyme*, which is the long established form became problematic when both forms appeared in definitions, therefore making two separate sequences of entries for the two spellings necessary. The capsule *African* tells how Murray at first omitted words derived from geographical and personal names by the addition of a suffix, but when this eventually proved untenable, he came up with a rationale for selective inclusion in cases where they either have other uses or have derivatives which need explanation. Under the capsule *insinuendo* where a value judgement is given in the *OED* Supplement, it is mentioned that evidence of prescriptivism in the *OED* is found since Murray and his fellow editors, a practice continued by Burchfield. The capsule *chav*, a word which only received a complete entry in *OED Online*, explains how it came about that every year two separate words, one each for British and American English, are nominated 'words of the year' by OUP lexicographers.

In conclusion: The book gives an absorbing and vivid account and a detailed and exhaustive presentation of the meticulously researched facts dealing with the development of the *OED* and its related dictionaries. At the same time it is a worthy and meritorious tribute to the hundreds of people, who collaborated in establishing one of the great historical dictionaries.

J.C.M.D. du Plessis
Bureau of the Woordeboek van die Afrikaanse Taal
Stellenbosch
South Africa
(jcmddp@sun.ac.za)

Oskar Reichmann. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen.* Studia Linguistica Germanica 111. 2012, 586 pp. ISBN 978-3-11-028255-9. e-ISBN 978-3-11-028269-6. Berlin/Boston: Walter de Gruyter.

In the introduction to this substantial work, Oskar Reichmann states his three objectives with the publication. Firstly, he wished to describe the scientific (especially linguistic and philological) and cultural-pedagogic activity called "historical lexicography" within its social context. As a result, he wanted to expose the ideas and interests behind dictionary writing that inspired lexicographers to compile dictionaries in their own languages about historical eras, works of authors, linguistic variation, etc. since the early modern period in German, Dutch and English societies which consequently made historical lexicography possible. Secondly, he wished to disclose the presuppositions in both the theory of language and the theory of history through which historical lexicography became a reality. His third objective was to identify the entire specialised apparatus comprising the methods used in lexicography. According to Reichmann, one can describe these three objectives as "ideological", "linguistic-theoretical" and "methodical". The reason for involving the lexicographical works of three languages is to highlight correlations and differences within a certain part of Europe, which can serve as a norm in the cultural sciences and declare such correlations as being normal.

Reichmann uses numerous examples from German, Dutch and English historical lexicography to illustrate his theoretical discussions, even though the examples from German historical lexicography appear more frequently — especially examples from the *Frühneuhochdeutsches Wörterbuch (FWB)*, on which he himself worked intensively for many years. He affords attention to the lexicography of phases in the history of these languages as well the development of the lexicography of *langue*, including for example, author lexicography. He maintains that lexicography traditionally was usually oriented more towards semasiology than towards syntax or morphology, and therefore concentrated more on semantic and pragmatic issues than on grammatical ones. Nouns received more attention than, for example, particles. He admits that his focusing so much on German lexicography may not be interesting to speakers of English for instance, but he did not aim to write a comprehensive and detailed history of German, Dutch and English lexicography. For scholars interested in historical lexicography, however, this publication will nevertheless be of invaluable assistance.

It is impossible to deal in detail with every aspect of this comprehensive study. Therefore, I will simply give a brief overview of some important aspects of the content and mention some issues for those interested in using the book as a reference source.

The first part of the book (Part A) deals with historical lexicography in social and scientific contexts when societies wanted to document their own history in order to understand the present better. Often, this was motivated by

their interest in language and literary texts. Reichmann asserts that lexicographers usually aim to do something meaningful in their societies when they decide to compile a dictionary, and their personal convictions play a role here. Historical lexicography concentrates on putting a historical perspective on the content; it also takes into account proven theoretical guidelines, for example, with regard to language phases. Historical dictionaries, however, are not simply linguistic special-field publications for specialist use only: they are also tools in the process of establishing and strengthening a language-historical textual and literary culture within a specific nation.

Lexicographers encounter several problems during the compilation process. These include the problems of collection, the ordering of the collected data, the treatment of the data, and deciding on the structure of the data, resulting in the problem of data accessibility. In addition, they have to decide *what* needs to be documented, having to make selections from all available data, and having to decide what is more important. Finally, they have to decide which medium of publication they will use, e.g. printed or electronic.

In their decision on an approach, historical lexicographers have a choice between a purely *historical perspective* or a *perspective relevant to present times*. In the first approach, one would typically document an earlier stage of a language (e.g. Old High German) or authors or groups of authors who used earlier stages of a language (e.g. Martin Luther or the Minstrels). One could then describe the text elements diachronically or synchronically. The latter would only concentrate on the specific language phase as if there were no past or future, whereas a diachronic description would take into account different developmental phases in language use. When a historical dictionary approaches the data from a perspective *relevant to present times*, on the other hand, it will contain all available etymological and historical semantic information on the language items. One example of this type of approach is the *Deutsches Wörterbuch (DWB)* by the brothers Grimm. Reichmann (2012: 20) presents an illuminating figure depicting the various approaches.

Reichmann also looks at the production processes of historical dictionaries, which vary according to type. Nowadays, historical lexicography in the German-speaking countries seems to be mainly limited to academic circles, but there are no empirical studies to determine the reception of historical dictionaries. Funding always remains a problem in the realisation of these dictionaries. According to Reichmann, historical lexicography — be it in the old style or electronically produced — needs a new ideology and concomitant new implementations.

After some sections dealing with groups of users and reasons for the use of historical dictionaries (amply illustrated with examples), Reichmann concludes the first part of his book with a discussion of ideologies influencing the making of historical dictionaries. Dictionaries may often serve political, patriotic or nationalist ends, but they may also aim at educating the broader society with regard to their own history, literature, environment and culture. These ideological viewpoints may lead lexicographers to apply value judgements

during the processing of the material. For example, qualitative categories may appear, such as "high literature". Certain word fields may be restricted, as Reichmann (2012: 76, note 102) mentions, where, for example, *A New English Dictionary on Historical Principles (NED)* of 1928 would highly value words from literature but at the same time limited inclusion of words from the scientific, technical and sexual fields. An extreme example of this approach can be found in *Trübners Deutsches Wörterbuch* (Trübner) which displays nationalist and fascist ideas (see also Mückel 2005 for an analysis of the first four volumes of this dictionary which were compiled during the National-Socialist era in Germany). Fortunately, German lexicography is nowadays pluricentric, which means it tries to be neutral and be inclusive of diverging viewpoints.

Part B of *Historische Lexikographie* classifies historical dictionaries into different types (see pages 98-99 for a comprehensive list of types and their features as suggested by Reichmann), whereas Part C describes the corpora and corpus excerption. The latter involves the extraction of citations and the interpretation of the word occurrences in historical texts.

Aspects taken into account here are, amongst others, the size and nature of the corpus — the number of existing texts and the possible regional and field-specific varieties to be excerpted. The method of selecting the lemmas and the information to be included will depend on the type of dictionary envisioned, its size and aims and the medium of publication (whether printed or electronic). In addition, one should distinguish between primary and secondary corpora. The former consist of the actual texts from a particular period, and the latter are scientific works from later periods discussing and interpreting texts or lemmas from the period. Of course, preference is given to primary corpora, and secondary corpora mainly function to verify the data in the primary corpora. Reichmann deplores the fact that none of the historical dictionaries of the older language phases in German, Dutch and English contains any information on the way in which the compilers selected and balanced their corpora, the only exception being the *FWB*.

Reichmann discusses different methods used in corpus excerption. One distinction is between *linear* and *punctual* excerption. The former occurs when the person doing the excerption reads a source from beginning to end and selects potential units for inclusion in the dictionary. It is also possible to do linear excerption by working through the glossary of a text. Usually the person doing excerption here is not the same person who will determine the corpus and/or formulate the dictionary articles. Punctual excerption, on the other hand, is the exact opposite, when the same person doing the excerpts is also the one who has to write the dictionary articles. In this process, the person proceeds from the lexicographical units which need to be included, and then accesses the texts at the exact points where these units occur in the texts — so that the text is not read as a whole. Of course, only someone who is an expert in a particular field can do this type of work.

Only after the successful excerption of the lexical units, can the actual lexi-

cographical process start. Reichmann explains this in Part D, which forms the most comprehensive part of the book. This lexicographical process includes the compilation of hundreds of thousands of dictionary articles accompanied by items that give more or less a paraphrase of their meanings and refer to other information. The types of lexicographical information and the components included in such dictionary articles receive attention against the background of the ideological assumptions, theoretical objectives and specific methods of historical lexicography underlying specific dictionaries. Reichmann adds his critical comments, drawn from his lifelong experience with historical lexicography, especially with the *FWB*.

In Part D of the book, Reichmann devotes many pages to the presentation of a systematic and theoretical foundation for the approach to meaning explanations. He also discusses the value of lexical signs, onomasiological networks, the treatment of phrases and syntax, word formation and word families and finally the citations.

The first aspect with which Reichmann deals is the conventions according to which information is selected for inclusion. At least the lemma sign, the meaning(s) and citations/examples should appear in the dictionary. Additional information could include variations in spelling and form, morphological inflections, syntax and phrases; pragmatic information on its distribution across language regions, historical eras, social strata and groups, as well as text types, similar expressions to which it is related or stands in opposition, etc.

Problems in selecting lemmas (whether one-word or multiword lemmas) include asking oneself whether a word is "worthy" of being included, in the sense of the lemmas having a general or specialised meaning. Other issues include the orthography of the lemmas with their variants and the order in which one would list them. Reichmann (2012: 178-179) shows some examples of markers that are used to help with the identification of lemma forms within the dictionary articles. The famous "finding problem" is also relevant here, because a dictionary user will not be able to find a particular word if its written form is unknown. Here, Reichmann finds approaches from several perspectives, such as:

- (i) a linguistic approach, dealing with the fact that lexicographers want to select, from a series of connotations and variant forms, a form that could serve as a so-called "construct lemma" (problematic concepts here are *inter alia* "basic form", "construct lemma", "citation lemma", etc.);
- (ii) an ideological language-historical approach, which for example gives preference to certain historical periods, such as Middle High German, as the ideal form;
- (iii) a historical-teleological approach, which claims that earlier forms can be found more easily when they are classed with later forms from later periods (which are more familiar to users);
- (iv) a location-oriented approach, which considers one particular region as the most important area and uses the forms of that area;

- (v) an explanatory approach, which believes older forms to be insufficient and need to be "corrected";
- (vi) a practice-oriented approach which, amongst others, deals with the identification of similar lemma signs on the basis of one's prior knowledge;
- (vii) an approach oriented towards the present, trying to link older forms to present-day forms;
- (viii) a frequency-oriented approach, which takes the most frequently used form as lemma; and
- (ix) an approach compliant with convention, building on current lexicographical conventions.

Reichmann also investigates the use of indices in historical dictionaries. In addition, compounds and inflections receive attention in relation to the "finding problem", as well as the etymology of the words. He gives examples of the treatment of etymological items in dictionary articles (2012: 191-210) and lists methods which different historical dictionaries use to deal with them. The Grimm brothers, for example, attempted to include all the meanings of a lexical expression, all the meanings of all the units that could be established in a word family, all historical, vernacular and social variants of each member of the word family, plus all similar units in other, genetically related languages. This led to the formation of a construct called the *Urbegriff* or *original concept*. The *Third Edition of the Oxford English Dictionary* (OED 2000 online), on the other hand, after its long preceding history, now has a sceptical approach towards tracing back words to their roots. In its foreword, it states: "References are no longer made to hypothetical reconstructed Indo-European forms. Instead, etymologies refer to recorded cognates formed from the assumed base" (2012: 197, note 62). This is in stark contrast to early dictionaries, such as that by Samuel Johnson, who even stated his intentions in the dictionary title: *A Dictionary of the English Language: in which the Words are Deduced from their Originals [...]*. Reichmann, however, mentions that Johnson's idea was not feasible, because combining the writings of the "best writers" with the original meanings of the words was not very practical.

The Dutch *Woordenboek der Nederlandsche Taal* (WNT), under influence of the Grimm brothers, gives much attention to etymology, but is on the other hand critical of the Grimms' approach (Reichmann 2012: 206). Reichmann lists some excellent etymological dictionaries, among which the Dutch *Etymologisch Woordenboek van het Nederlands* (*Et. Wb. Ned.* (2003)), of which three volumes have so far appeared, as well as that by J. de Vries (1997) and that by P.A.F. van Veen and N. van der Sijs (1997).

The next section of the book deals with the meaning of the lemma signs. Reichmann discusses this aspect in detail, using amongst others the example of the Early Modern High German word "geschichte" (2012: 255-278). Lexicographers follow certain steps when they want to determine the meaning of lemma signs. The first step is to read all the citations at least once, in order to orientate

themselves in the field and the contexts in which the lemma signs appear. Then follows the interpretation of the citations, which in its turn is followed by sorting them into classes. The question of how a lexicographer systematically determines the meanings, especially if there is only one lexicographer at work, has up to the present not been described in detail. There is, he maintains, no metalexicographical literature in German, Dutch or English which explains how a dictionary article comes into being (2012: 255). Reichmann states that this could become a very impressionistic procedure when done by a single lexicographer. He gives examples of methods that could guarantee objective descriptions of lemma signs. Synonymous expressions found in the citations could be one such method, as are expressions showing a high degree of similarity. In addition, expressions which have been used across different time periods may also be helpful. One of the problems is that historical texts mainly originated as spoken texts, and only gradually found their way into written documents. The lexicographer only has the written texts with which to work — the social and historical-actional contexts have to be considered.

To explain and convey the meaning of the lemma signs systematically, Reichmann distinguishes between different types of meaning explanations, namely, (i) "simple" versus "complex" meaning explanations, (ii) explanations making use of "synonyms" and those using "paraphrases", and finally (iii) "compact" versus "diffuse" meaning explanations. The simple explanation usually consists of only one or a few synonyms, or a short paraphrase. Complex explanations, on the other hand, contain longer series of synonyms and more detailed, syntactically complex word groups (which could be synonyms, but do not have to be). They usually contain paraphrases rather than merely synonyms and mention relationships such as frames to which a particular lemma sign may belong, as well as all kinds of additional descriptions. Such complex explanations also have comprehensive citations.

As for the second type of meaning explanations, describing the meaning of lemma signs by means of synonyms and paraphrases is not exactly the same as the dichotomy of "simple" and "complex". When making use of synonyms, the lexicographer tells the user that a particular synonym can take the place of the lemma in the context. Synonym explanations indicate that the meaning has remained more or less stable throughout the ages. Paraphrases, on the other hand, indicate the possibility, not the reality, that the explanation has the potential to fulfil the interpretative gap. They indicate however that there is no specific synonym or equivalent for the particular lemma sign, and that the lemma sign is used differently at present from its use in historical texts. These two types of explanations occur in combination in all historical dictionaries.

According to Reichmann, one can directly perceive the distinction between "compact" and "diffuse" meaning explanations when looking at their typographical layout. Compact explanations contain letter types in different sizes and blocks of citations, whereas diffuse explanations contain all the article items in one microstructural unit. This distinction is purely descriptive and not

subject to evaluative judgement. However, Reichmann discusses reasons for using compact explanations, and they definitely make a dictionary article visually more accessible. On the negative side, they use more space, which is problematic in printed dictionaries. If they become too long, they become complex, making it tedious to work through all the hierarchical structures.

One issue with which Reichmann deals is that of the governing criteria for approaches to meaning explanations: the measure of abstraction, the delimitation of one meaning from another, the number of explanations and their internal ordering, as well as the networks between them. Here, Reichmann distinguishes five procedures of verification, namely, (i) a field test, (ii) testing opposites, (iii) testing word formation, (iv) testing syntagmas, and (v) testing the value of the signs.

In the field test procedure, the lexicographer draws up a network field of meaning-related expressions, based on the citations. The more citations, the more comprehensive such a field will be. The testing of opposites, on the other hand, rests on the same principle as the field test, but deals with antonyms. In testing word formation, the lexicographer will look at the possibilities for compound words to be formed in relation to the particular lemma sign, and the syntagma test comprises of drawing up lists of syntagmas which cluster around a specific distinguishable meaning (of a particular lemma sign or expression). Testing the value of the signs means that the lexicographer can verify particular meanings, based on the citations.

One can also order single meanings in historical dictionaries by applying various criteria. Reichmann distinguishes between historical criteria, geographical and sociological ones, the criteria of frequency, generic relations, specificity, concreteness, the possibilities for word formation, closeness in content, syntactical structure, etymology, alphabetical order, and practical considerations. He discusses each of these criteria (2012: 341-351), using interesting examples to illustrate his arguments.

When consulting historical dictionaries, users can get help through different designs for or markers within a dictionary article. These include the medium of language (e.g. by listing synonyms), but also explicit and implicit items of separation within the dictionary article, so that readers may find different sections more easily. Reichmann also discusses the use of labels (2012: 369-379), which is not only relevant for marking lemma signs according to historical periods, linguistic and geographical areas, social groups and levels, but also according to text types. Once again, he gives many examples from Dutch, English and German historical dictionaries (2012: 372-379).

An entire section is devoted to onomasiological networks and the way they are treated within the articles of historical dictionaries (2012: 379-404). Reichmann's conclusion is that citations in existing historical dictionaries often do not give clearly discernible onomasiological demarcations for the synonyms in a particular field, which hampers user-friendliness. Nevertheless, even in dictionaries focusing on a semasiological approach, onomasiological informa-

tion is necessary in order to reveal relationships of synonymy to the user. Reichmann singles out the newly published *Historical Thesaurus of the Oxford English Dictionary* (HTOED 2009) as a giant leap towards describing the origin and history of a word as well as "tracing how a *meaning* emerged and came to be expressed in words". The authors built up a taxonomy that distinguishes on the first level between the "external world", the "mental world" and the "social world". On lower hierarchical levels, they then form "categories" and "subcategories" — creating onomasiological links in this way. The precursor to this *Historical Thesaurus* is the *A Thesaurus of Old English* (TOE). In this dictionary, which is "conceptually arranged", there are eighteen large groups, which include the "physical world", with subdivisions such as "earth", which once again has subdivisions such as "surface of the earth", which in their turn is subdivided into categories such as "north", and so forth.

In his discussion on the use of paraphrases to describe the meanings of words, Reichmann is of the opinion that if a lexicographer uses a corpus in the compilation process, it might be difficult to prove that a specific paraphrasal description of a word is indeed acceptable as a meaning explanation. Some problems that arise here include whether paraphrases should receive the same status as the lemmas they describe and be entered as lemmas as well, and the order in which to present such paraphrases within a dictionary article. In spite of his warnings in relation to the use of paraphrases, Reichmann admits that at the beginning of his career he regarded paraphrases as less relevant, but as time went by he came to realise their usefulness (2012: 411, note 242).

Syntagmatic information in historical dictionaries should also be drawn from the corpus, and lexicographers should ask themselves whether they have space to add syntagmatic citations apart from the normal citations which explain the meaning of a word. The particular dictionary function will help to determine the nature and extent of the syntagmatic information.

The section on word formation and the treatment of word families again contain numerous examples from historical dictionaries in German, Dutch and English (2012: 432-471).

In addition, Reichmann presents a detailed discussion of citations and their treatment. Apart from the different requirements for citations, he mentions three possibilities where citations may appear within a dictionary article. It could stand *before* the meaning explanation(s), *after* the meaning explanation(s) or *within* the part of the dictionary article that explains the meaning — in the latter position, it can also take on different forms.

He presents a typology for citations with reference to the problem of selection of the lemmas. These include (i) the text-internal definition type of explanation, (ii) characterisations, (iii) identifications, (iv) frame-oriented attributive citations, (v) expressions indicating antonyms and (vi) expressions indicating hyponyms and hyperonyms. For each of these types, Reichmann offers several examples. The order for listing citations also receives attention. They may be ordered in chronological order according to historical periods, in

a specific geographical order, or in combinations using different principles.

Dictionaries sometimes present comments by the lexicographer, and Reichmann also discusses this phenomenon. Even though one could say that every dictionary article that a lexicographer writes is already a specific comment (or formulation by the lexicographer), the comment referred to here is actually rather a metalexicographical one — it documents linguistic facts *about* the language items in question. These comments can appear in any of the information types. Usually, however, they deal with the lexicographer's assessment of issues such as particularly relevant features to be taken into account, motivation for the selections or choices which may also have been motivated and selected differently, verification of particular selections and formulations, and additional knowledge of lexicographical, semantic, factual or text-historical knowledge stemming from the lexicographer's own experience and field-related expertise. Reichmann pays attention to lexicographers' comments in each type of dictionary item that occurs in the dictionary articles of historical dictionaries, illustrating his discussion with numerous examples.

The last question that Reichmann deals with is whether it is possible to give information on frequency of occurrence. According to him, this is not actually possible in historical lexicography, in the sense of submitting information for each single lemma by means of an indication of quantity or a specific number of occurrences. Absolute numbers can, of course, nowadays be obtained by using modern digital methods, usually in connection with citations. Lexicographers generally make use of additional formulations such as "relatively", "usually", "somewhat", "often", etc., for lemmatic items.

As conclusion to the book, Part E contains a comprehensive reference list of the dictionaries mentioned in the book as well as other works relevant to the topic. Apart from the many historical and other dictionaries he used in his investigation, Reichmann also worked through monographies, volumes and dissertations dealing with the topic of historical dictionaries in general and with specific dictionaries in particular; discussions of cultural, philosophical and historical issues such as memory (e.g. Aleida Assmann 2006) and other relevant aspects pertaining to historical dictionaries and their compilation. A very handy index appears right at the end of the book.

This study is very worthwhile for students of historical lexicography, and for lexicographers who intend to compile historical dictionaries. Reichmann deals conclusively and comprehensively with all aspects, and the reader can learn much from this formidable master of historical lexicography. It is a book that can be recommended without reserve.

References

- Assmann, A. 2006. *Der lange Schatten der Vergangenheit. Erinnerungskultur und Geschichtspolitik*. München: C.H. Beck.
- De Vries, J. 1997. *Nederlands Etymologisch Woordenboek. Met aanvullingen, verbeteringen en woordregis-*

- ters door F. de Tollenaere. Fourth Edition. Leyden/New York/Cologne: E.J. Brill.
- DWB = Grimm, J. and W. Grimm (Eds.). 1854–1971. *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. 16 Volumes. Leipzig: Verlag von S. Hirzel.
- Et. Wb. Ned. = Philippa, M., F. Debrabandere and A. Quak (Eds.). 2003. *Etymologisch woordenboek van het Nederlands*. Amsterdam: Amsterdam University Press. Three Volumes so far.
- FWB = *Frühneuhochdeutsches Wörterbuch*. Eds. Anderson, R.R. (Vol. 1) / Goebel, U. (all volumes) / Lobenstein-Reichmann, A. (Vols. 5, 6, 10–13) / Reichmann, O. (all volumes; for Vols. 4 and 7 in cooperation with the Institut für deutsche Sprache). Vols. 1–3 compiled by Reichmann, O.; Vol. 4 compiled by Schildt, J.; Vol. 5 compiled by Dengler, M., D. Hüpper, O. Pfefferkorn, J.-Macha and H.-J. Solms; Vol. 6 compiled by Schildt, J. and O. Reichmann; Vol. 7 compiled by Lobenstein-Reichmann, A., O. Pfefferkorn and O. Reichmann; Vol. 8 compiled by Winge, V.; Vol. 9 compiled by Lobenstein-Reichmann, A. and O. Reichmann; Vol. 11 compiled by Reichmann, O. Berlin/New York: Walter de Gruyter.
- HTOED = Kay, C. et al. (Eds.). 2009. *Historical Thesaurus of the Oxford English Dictionary with Additional Material from A Thesaurus of Old English*. Vol. 1: Thesaurus. Vol. 2: Index. Oxford: Oxford University Press.
- Johnson, S.** 1755. *A Dictionary of the English Language: In which the Words are Deduced from their Originals, and Illustrated in their Different Significations by Examples from the Best Writers; to which are Prefixed, A History of the Language, and an English Grammar*. London: Printed by W. Strahan, for J. & P. Knapton. [Newer edition New York: AMS Press, 1967].
- Mückel, W.** 2005. "Trübners Deutsches Wörterbuch" (Band 1-4) — ein Wörterbuch aus der Zeit des Nationalsozialismus. *Eine lexikographische Analyse der ersten vier Bände (erschienen 1939-1943)*. Lexicographica Series Maior 125. Tübingen: Niemeyer.
- NED = 1884–1928. *A New English Dictionary on Historical Principles*. 10 Volumes. Oxford: Clarendon Press.
- OED 2000 = Simpson, J.A. (Ed.). 2000. *The Third Edition of the Oxford English Dictionary*. Online at: www.oed.com.
- TOE = Roberts, J. and C. Kay, with L. Grundy (Eds.). 2000. *A Thesaurus of Old English in Two Volumes*. Vol. 1: Introduction and Thesaurus. Vol. 2: Index. Second Impression. Amsterdam/Atlanta: Rodopi.
- Trübner = Götze, A. (Ed.), from 5th Volume Mitzka, W. (Ed.). 1939–1957. *Trübners deutsches Wörterbuch*. 8 Volumes. Berlin: De Gruyter.
- Van Veen, P.A.F. and N. van der Sijs.** 1997. *Etymologisch woordenboek. De herkomst van onze woorden*. Utrecht/Antwerpen: Van Dale Lexicografie.
- WNT = De Vries, M. and L.A. te Winkel (Eds.). 1882–1998. *Woordenboek der Nederlandsche Taal*. 29 Volumes, Supplement 1956. 's-Gravenhage/Leiden: Martinus Nijhoff.

Maria Smit
Independent Lexicographer and Language Practitioner
Stellenbosch
South Africa
(ria.eden@gmail.com)

Publikasieaankondigings / Publication Announcements

Peter Gilliver. *The Making of the Oxford English Dictionary*. 2016, xix + 625 pp. ISBN 978-0-19-928362-0. Oxford: Oxford University Press. Price: £40. (Review in this issue.)

Geoffrey Leech and Yoshihiko Ikegami (Eds.). *Longman English–Japanese Dictionary*. 2007, xxi + 2040 pp. ISBN 978-1405872980. Tokyo: Pearson Education. Price £24.99. (Lexiconote in this issue.)

S. Paas (Editor). *Chichewa–English/English–Chichewa Dictionary*. Fifth edition. 2016, 1158 pp. ISBN 978-0-19-041659-1 (Paperback). Cape Town: Oxford Orbis. Price R171.

Oskar Reichmann. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen*. *Studia Linguistica Germanica* 111. 2012, 586 pp. ISBN 978-3-11-028255-9, e-ISBN 978-3-11-028269-6. Berlin/Boston: Walter de Gruyter. Prys €170. (Review in this issue.)

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 ingeen 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 nuusvystellings 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, Postbus 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 versoek 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 deurlpend in die vorm van boskrite 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**.