

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

Lexikos 27

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

Lexikos 27

Redakteur

Editor

H. Steve Ndinga-Koumba-Binza

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 27:2017

AFRILEX SERIES 27:2017



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 © 2017 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2017

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 978-0-9946528-3-6
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/>

Lexikos is available online at <http://lexikos.journals.ac.za/>

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. D. Nkomo (RSA en Zimbabwe/RSA and Zimbabwe)
Prof. T.J. Otlogetswe (Botswana)
Prof. A.N. Otto (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	x
Foreword	xii
<i>H. Steve Ndinga-Koumba-Binza</i>	
'n Woord van AFRILEX	xiv
A Few Words from AFRILEX	xv
<i>Herman Beyer</i>	
Redaksionele doelstellings	xvi
Editorial Objectives	xvii

Artikels / Articles

The Interpretive Function: To Be or Not to Be, That is the Question	1
<i>Heidi Agerbo</i>	
Prolingua se bydrae tot terminologieontwikkeling in Afrikaans	16
<i>Mariëtta Alberts</i>	
Guessing Verb–Adverb Collocations: Arab EFL Learners' Use of Electronic Dictionaries	50
<i>Safi Eldeen Alzi'abi</i>	
A Usability Evaluation of the Prototype <i>Afrikaanse idiomewoordeboek</i>	78
<i>Liezl H. Ball and Theo J.D. Bothma</i>	
Polyseme Selection, Lemma Selection and Article Selection	107
<i>Henning Bergenholtz and Rufus H. Gouws</i>	
Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated?	132
<i>Monika Biesaga</i>	

A Computational Approach to Zulu Verb Morphology within the Context of Lexical Semantics <i>Sonja E. Bosch and Laurette Pretorius</i>	152
The Effectiveness of Using a Bilingualized Dictionary for Determining Noun Countability and Article Selection <i>Alice Yin Wa Chan</i>	183
Deutsches Wortgut in rumänischen Pflanzennamen. Ein Bericht aus der rumänischen Wörterbuchpraxis <i>Radu Drăgulescu</i>	214
Visualisation of Collocational Preferences for Near-Synonym Discrimination <i>Mariusz Piotr Kamiński</i>	237
Intellectualization through Terminology Development <i>Langa Khumalo</i>	252
"Minimum Input, Maximum Output, Indeed!" Teaching Collocations Through Collocation Dictionary Skills Development <i>Sugene Kim</i>	265
Multimodal Exemplification: The Expansion of Meaning in Electronic Dictionaries <i>Xiqin Liu</i>	287
Koalas, Kiwis and Kangaroos: The Challenges of Creating an Online Australian Cultural Dictionary for Learners of English as an Additional Language <i>Julia Miller, Deny A. Kwary and Ardian W. Setiawan</i>	310
The Dictionary in Examinations at a South African University: A Linguistic or a Pedagogic Intervention? <i>Dion Nkomo</i>	346
Planning a Dictionary for Mother Tongue Education: A Conceptual Framework for Gabonese Languages <i>Blanche Nyangone Assam</i>	378

Direct User Guidance in e-Dictionaries for Text Production and Text Reception — The Verbal Relative in Sepedi as a Case Study <i>D.J. Prinsloo, Theo J.D. Bothma, Ulrich Heid and Daniel J. Prinsloo</i>	403
Afrikataalleksikografie: Gister, vandag en môre <i>D.J. Prinsloo en Elsabé Taljard</i>	427
Überlegungen zu einem neuen zweisprachigen phraseologischen Wörterbuch Deutsch–Rumänisch <i>Doris Sava</i>	457
Lost and Found: The Value of a Little Known Bilingual Dictionary Towards the Intellectualization of Ndaou <i>Emmanuel Sithole and Dion Nkomo</i>	478
L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography <i>Sven Tarp, Kasper Fisker and Peter Sepstrup</i>	494
Approaches to the Treatment of Zero Equivalence in a Bilingual Dictionary <i>Marjeta Vrbinc and Alenka Vrbinc</i>	522
Colors in French, American and British Dictionaries <i>Krista Williams</i>	540
Revisiting Lemma Lists in Swahili Dictionaries <i>Beata Wójtowicz</i>	561
Representing the Cultural Dimension of Meaning in Learner's Dictionaries — From the Perspective of Chinese EFL Learners in L2 Reception <i>Mei Xue</i>	578
Leksikonota / Lexiconote	
Etymology in the <i>Polish Academy of Sciences Great Dictionary of Polish</i> <i>Przemysław Dębowiak, Bogumił Ostrowski and Jadwiga Waniakowa</i>	597

Resensies / Reviews

Philip Durkin: *The Oxford Handbook of Lexicography* 609
Li Dehui

Die Taalkommissie van die Suid-Afrikaanse Akademie 614
vir Wetenskap en Kuns: *Afrikaanse Woordelys en Spelreëls*
Marius Swart

Publikasieaankondigings / Publication Announcements 620

Voorskrifte aan Skrywers 622
Instructions to Authors 623

Voorwoord

Hierdie 27ste nommer van *Lexikos* begin die derde siklus van ons stelsel van 'n span van roterende redakteurs wat ná dr. Johan du Plessis se aftrede in 2010 in werking gestel is. Gedurende die eerste siklus, met prof. Elsabé Taljard as redakteur van *Lexikos* 21, prof. Danie Prinsloo as redakteur van *Lexikos* 22 en dr. Johan du Plessis (wat vir prof. Rufus Gouws ingestaan het) as redakteur van *Lexikos* 23, was daar geen verandering in die kwaliteit of gereelde verskyning van ons joernaal nie. Die tweede siklus, met prof. Rufus Gouws as redakteur van *Lexikos* 24, prof. Elsabé Taljard as redakteur van *Lexikos* 25 en prof. Danie Prinsloo as redakteur van *Lexikos* 26, het bewys dat ons nuwe redakteurstelsel van *Lexikos* volhoubaar is.

Die redakteurstaak hou my nederig aangesien ek prof. Rufus Gouws in die redakteurspan vervang het. Dis nie 'n maklike taak om sy skoene vol te staan nie. Die enigste manier waarop ek dit geslaagd kon doen, is met die hulp van proff. Elsabé Taljard en Danie Prinsloo. Daar is reeds met die inwerkstelling van ons nuwe stelsel besluit dat die redakteurspan gesamentlik verantwoordelik sal wees vir elke nommer, en een lid van die span jaarliks as eindredakteur aangewys sal word. Gedurende die voorbereiding van verlede jaar se nommer onder leiding van prof. Danie Prinsloo, het ek 'n tipe opleidingsfase ondergaan terwyl ek as mederedakteur gewerk het.

Proff. Elsabé Taljard en Danie Prinsloo was ook albei baie hulpvaardig en ondersteunend in die totstandkoming van hierdie *Lexikos*-nommer. Ek wil hulle hartlik bedank dat hulle my taak, wat vir my 'n volgehoue leerervaring was, genotvol en hanteerbaar gemaak het. Ek wil ook prof. Rufus Gouws bedank, wat as die naaste lid van die Adviesraad maklik bereik kon word. Sy insig was van groot waarde. Eweneens is ek dankbaar teenoor dr. Johan du Plessis, wat ek ook van tyd tot tyd geraadpleeg het.

Hierdie *Lexikos*-nommer is in pas met die tradisie van ons joernaal, m.a.w. dit bevat uiteenlopende bydraes van oor die wêreld heen rakende verskillende leksikografiese aspekte — teoreties sowel as prakties — van 'n wye verskeidenheid tale. Dit is 'n bewys daarvan dat ons joernaal gevestig is as een van die toonaangewende internasionale joernale in die veld. Dit is dus geen verrassing dat die impakfaktor van hierdie joernaal hoër is as die gemiddeld onder joernale in die geesteswetenskappe wat buite die Westerse wêreld gepubliseer word nie. Namens die redakteurspan wil ek alle outeurs en keurders van artikels bedank vir hul volgehoue bydraes om die hoë standaard en sukses van *Lexikos* te handhaaf.

Dit is slegs met ervaring wat 'n taak suksesvol voltooi kan word. In hierdie rol het mes. Tanja Hartevelde en Hermien van der Westhuizen by die Buro van die Woordeboek van die Afrikaanse Taal (WAT) weereens 'n baie belangrike rol

gespeel in die totstandkoming van *Lexikos* 27. Hulle het gehelp met die uitleg van die artikels, het finale kontrolering van taal en uitleg gedoen, ens. Me. Tanja Harteveld is ook die resensieredakteur van ons joernaal, 'n rol wat sy suksesvol vertolk het in die voorbereiding van hierdie nommer. Die redakteurspan is opreg dankbaar vir die uitstekende werk wat deur mes. Tanja Harteveld en Hermien van der Westhuizen gedoen is.

Die redakteurs van *Lexikos* verwelkom altyd kommentaar en kritiek wanneer dit tot die verbetering van die kwaliteit en goeie reputasie van ons joernaal kan lei. Dit moet, trouens, nie net die publisering wees wat van belang is nie, maar hoofsaaklik die maak van 'n impak op ons veld en/of ander taaldissiplines. Mag hierdie nommer in dié doel slaag.

Kontak ons deur 'n e-pos te stuur aan lexikos@sun.ac.za.

Hugues Steve Ndinga-Koumba-Binza
Redakteur

Foreword

This 27th issue of *Lexikos* starts the third cycle of our system of a team of rotating editors that was put in place following the retirement of Dr Johan du Plessis in 2010. There was no change in the quality of or the regularity with which our journal was published during the first cycle with Prof. Elsabé Taljard as editor of *Lexikos* 21, Prof. Danie Prinsloo as editor of *Lexikos* 22, and Dr Johan du Plessis (who kindly deputized for Prof. Rufus Gouws) as editor of *Lexikos* 23. The second cycle with Prof. Gouws as editor of *Lexikos* 24, Prof. Taljard as editor of *Lexikos* 25, and Prof. Prinsloo as editor of *Lexikos* 26 proved that our new system for the editorship of *Lexikos* is indeed sustainable.

It was a humbling experience to replace Prof. Gouws on the editorial team, who left big shoes to fill. Thanks to the help of Profs Taljard and Prinsloo, I was able to do so successfully. Upon initiating our new system, it was decided that the team would take joint responsibility for each issue, with one member of the team being appointed as final editor annually. During the preparation of last year's volume under the editorship of Prof. Prinsloo, I went through a form of training phase working as co-editor.

Profs Taljard and Prinsloo were also very helpful and supportive in compiling the current issue. I wish to thank them sincerely for making my task, which was a continuous learning experience, enjoyable and manageable. I would also like to thank Prof. Gouws, who was, as the closest Advisory Board member, easily reachable. His insights were of great value. Equally, I am grateful to Dr Johan du Plessis, whom I also consulted from time to time.

This issue of *Lexikos* keeps with the tradition of our journal, i.e. it contains diverse contributions from all over the world on a variety of lexicographic issues — both theoretical and practical — of different languages. This shows that our journal has established itself as one of the leading international journals in its field. It is therefore not surprising that the impact factor of the journal is higher than average among journals in humanities published outside the Western world. On behalf of the team of editors, I wish to thank all of our authors and reviewers for their continued contribution to maintaining the high quality standards and success of *Lexikos*.

The successful completion of a task such as this relies heavily on experience. In this regard, Mss Tanja Harteveld and Hermien van der Westhuizen at the Bureau of the Woordeboek van die Afrikaanse Taal (WAT) once again played a very important part in compiling this volume of *Lexikos*. They assisted in typesetting the articles, making final language and structural checks, etc. Ms Harteveld is also the review editor of our journal — a role that she successfully assumed in the preparation of this issue. The team of editors is truly grateful for the excellent work done by Mss Harteveld and Van der Westhuizen.

The editors of *Lexikos* are always open to comments and criticism regarding the improvement of the quality and good reputation of our journal. After all, it should not just be about publishing, but mainly about making an impact in our field and/or other language disciplines. May this issue contribute to fulfilling this purpose!

You are welcome to contact us by sending an email to lexikos@sun.ac.za.

Hugues Steve Ndinga-Koumba-Binza
Editor

'n Woord van AFRILEX

Die African Association for Lexicography (AFRILEX) is dankbaar en besonder trots daarop om 'n internasionaal gevestigde en hoog aangeskrewe vaktydskrif soos *Lexikos* as sy mondstuk te hê. Sedert *Lexikos* in 2011 danksy die inisiatief van die toenmalige AFRILEX-president prof. Gilles-Maurice de Schryver die publikasie-model van Goue Oop Toegang terugwerkend tot nommer 1 volg, het dit ook toenemend die voorkeurpublikasie van internasionale navorsers van buite Afrika geword, soos wat die inhoudsopgawes van die afgelepe klompie nommers duidelik toon. Dit is ook verblydend dat AFRILEX-lede se navorsing steeds 'n beduidende deel van elke nommer van die tydskrif uitmaak. Hierdeur word die vertroue in en akademiese status van *Lexikos* in die metaleksikografie en verwante dissiplines bevestig.

Met die uitgewer van *Lexikos* kan AFRILEX kwalik meer in sy skik wees. Die Buro van die WAT het met die publikasie van die eerste nommer in 1991 — vier jaar vóór die totstandkoming van AFRILEX — die leiding geneem in die metaleksikografiese gesprek in Suider-Afrika. AFRILEX was dus in die gelukkige posisie om te kon assosieer met 'n reeds gevestigde vaktydskrif as sy mondstuk. Onder die hoofredakteurskap van dr. Willem Botha het die Buro van die WAT gesorg dat *Lexikos* sy digitale tuiste as 'n Goue-Oop-Toegang-vaktydskrif gevind het by <http://lexikos.journals.ac.za/>. Die nougesette en onberispelike versorging van elke nommer van *Lexikos* het 'n wesenskenmerk van die tydskrif geword, danksy die standaard wat die eerste redakteurs, mnr. Pieter Harteveld (1991–1995), dr. Dirk van Schalkwyk (1996) en dr. J.C.M.D. du Plessis (1997–2010), almal verbonde aan die Buro van die WAT, daargestel het. Ná dr. Du Plessis se uittrede het die simbiotiese verhouding tussen die Buro van die WAT en AFRILEX verdiep deurdat drie AFRILEX-lede wat nie aan die Buro verbonde is nie op 'n jaarliks roterende basis die redakteurskap van *Lexikos* behartig. Die redakteur van hierdie nommer is dr. Hugues Steve Ndinga-Koumba-Binza, bygestaan deur me. Tanja Harteveld van die Buro van die WAT as resensieredakteur, met tegniese ondersteuning deur me. Hermien van der Westhuizen.

Dit is my voorreg om namens AFRILEX die redaksionele span, die Buro van die WAT en bydraende outeurs van harte te bedank vir nommer 27 van *Lexikos*.

Herman L. Beyer
President: AFRILEX

A Few Words from AFRILEX

The African Association for Lexicography (AFRILEX) is grateful and particularly proud to have an internationally established and highly regarded journal like *Lexikos* as its mouthpiece. Since 2011, when, thanks to the initiative of the then AFRILEX president Prof. Gilles-Maurice de Schryver, *Lexikos* adopted the Gold Open Access publication model retrospectively to volume 1, the journal has also become the preferred publication for international scholars beyond the borders of Africa, as the tables of contents of the last few volumes clearly demonstrate. It is also heart-warming that the research of AFRILEX members continue to make up a significant part of every volume of the journal. This confirms the trust in and academic stature of *Lexikos* in metalexigraphy and related disciplines.

AFRILEX could not be happier with the publisher of *Lexikos*. With the publication of the first volume in 1991 — four years before the establishment of AFRILEX — the Bureau of the WAT took the lead in metalexigraphic discourse in Southern Africa. Therefore, AFRILEX was in the fortunate position to be able to associate with an already established journal as its mouthpiece. Under Editor-in-Chief Dr Willem Botha the Bureau of the WAT saw to it that *Lexikos* found its digital home as a Gold Open Access journal at <http://lexikos.journals.ac.za/>. The conscientious and impeccable care with every volume of *Lexikos* became a hallmark of the journal, thanks to the standards set by its first editors, Mr Pieter Hartevelde (1991–1996), Dr Dirk van Schalkwyk (1996) and Dr J.C.M.D. du Plessis (1997–2010), all of whom were associated with the Bureau of the WAT. After Dr Du Plessis stepped down, the symbiotic relationship between the Bureau of the WAT and AFRILEX deepened through the annual rotation of the journal's editorship among three AFRILEX members who are not associated with the Bureau. The editor of this volume is Dr Hugues Steve Ndinga-Koumba-Binza, assisted by Ms Tanja Hartevelde of the Bureau of the WAT as reviews editor, with the technical support of Ms Hermien van der Westhuizen.

It is my privilege to, on behalf of AFRILEX, sincerely thank the editorial team, the Bureau of the WAT and contributing authors for volume 27 of *Lexikos*.

Herman L. Beyer
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.

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

The Interpretive Function: To Be or Not to Be, That is the Question

Heidi Agerbo, *Centre for Lexicography,
School of Communication and Culture, Aarhus University,
Aarhus, Denmark (heap@cc.au.dk)*

Abstract: Approximately a decade ago, it was suggested that a new function should be added to the lexicographical function theory: the interpretive function¹. However, hardly any research has been conducted into this function, and though it was only *suggested* that this new function was relevant to incorporate into lexicographical theory, some scholars have since then *assumed* that this function exists², including the author of this contribution. In Agerbo (2016), I present arguments supporting the incorporation of the interpretive function into the function theory and suggest how non-linguistic signs can be treated in specific dictionary articles. However, in the current article, due to the results of recent research, I argue that the interpretive function should not be considered an individual main function. The interpretive function, contrary to some of its definitions, is not connected to acting and therefore the only difference between reception and interpretation is that they work with different types of sign. However, the type of sign is not relevant for a function, or rather, it should not be a criterion for distinguishing between functions. The lemma selection for the communicative, cognitive as well as the operative functions could and should include linguistic as well as non-linguistic signs. Thus, theoretically, there is no reason to identify a fourth dictionary function as suggested by Tarp (2008), and practically, the development of modern technologies has diminished the distance in the treatment of different types of sign, making it easier for lexicographers to lemmatise non-linguistic signs. Concerning the point that non-linguistic signs are also worthy of lexicographical attention, my suggestion from 2016 still stands, the difference in this contribution being that the interpretive function is not considered an individual function.

Keywords: LEXICOGRAPHICAL FUNCTION THEORY, DICTIONARY FUNCTION, INTERPRETIVE FUNCTION, INFORMATION TOOLS, LINGUISTIC SIGN, NON-LINGUISTIC SIGN, ACTING

Opsomming: Die vertolkende funksie: Om te wees of nie te wees: dis die vraag. Omtrent 'n dekade gelede is daar voorgestel dat 'n nuwe funksie by die leksikografiese funksieteorie gevoeg moet word, naamlik die vertolkende funksie¹. Min of geen navorsing is egter oor hierdie funksie gedoen nie, en hoewel daar slegs *voorgestel* is dat hierdie nuwe funksie relevant was en by die leksikografiese teorie ingewerk behoort te word, het sommige navorsers sedertdien *aanvaar* dat hierdie funksie bestaan², insluitende die outeur van hierdie bydrae. In Agerbo (2016) bied ek argumente ter ondersteuning van die opname van die vertolkende funksie in die funksieteorie en doen aan die hand hoe nietaalkundige tekens in spesifieke woordeboekartikels behandel kan word. Weens die bevindinge van onlangse navorsing argumenteer ek egter in die huidige

artikel dat die vertolkende funksie nie beskou behoort te word as 'n aparte hoof funksie nie. Die vertolkende funksie, in teenstelling met sommige definisies, hou nie verband met toneelspel nie en dus is die enigste verskil tussen resepsie en interpretasie dat hulle werk met verskillende soorte tekens. Die soort teken is egter nie relevant vir 'n funksie nie, of eerder, dit behoort nie 'n maatstaf te wees om tussen funksies te onderskei nie. Die lemmaseleksie vir die kommunikatiewe, kognitiewe sowel as die operatiewe funksies kan en behoort sowel taalkundige as nietaalkundige tekens in te sluit. Teoreties is daar dus geen rede om 'n vierde woordeboekfunksie te identifiseer nie, soos voorgestel deur Tarp (2008), en prakties het die ontwikkeling van moderne tegnologie die afstand in die behandeling van verskillende soorte tekens verklein, wat dit vir leksikograwe makliker maak om nietaalkundige tekens te lemmatiseer. Wat betref die punt dat nietaalkundige tekens ook leksikografiese aandag waardig is, bly my voorstel uit 2016 steeds geldig, met die verskil in hierdie bydrae dat die vertolkende funksie nie beskou word as 'n aparte funksie nie.

Slutelwoorde: LEXIKOGRAFIESE FUNKSIETEOORIE, WOORDEBOEFUNKSIE, VERTOLKENDE FUNKSIE, INLIGTINGSHULPMIDDELS, TAALKUNDIGE TEKEN, NIETAALKUNDIGE TEKEN, TONEELSPLE

1. The interpretive function in lexicographical literature and existing dictionaries

Today, function theorists argue that there are four main functions: communicative, cognitive, operative and interpretive. The interpretive function in relation to the function theory was first mentioned briefly in Tarp (2008). In this article, Tarp 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" (2008: 11). Four years later, Bothma and Tarp state that "[t]he *operative* and *interpretive situations* have so far been scarcely studied by lexicography and there are only relatively few lexicographical works that cater for these situations" (2012: 91). Not much has been added to the description of the interpretive function since then despite the fact that it was first introduced in the above-mentioned article by Tarp in 2008, and despite the fact that at a symposium held at the Centre for Lexicography in Aarhus in 2008, it was said that the interpretive (and operative) situations "are starting to gain the attention of lexicographers and are therefore expected to be of interest in the near future" (Fernández and De Alba 2011: 307-308). This lack of attention has also been confirmed in an investigation carried out by this author of lexicographical literature in the period 2008-2016. In addition to this, no dictionary project has yet *intentionally*³ incorporated the interpretive function, though it could be argued that certain dictionaries of symbols can be ascribed such a function.

Apart from Tarp (2008), other texts commenting on this function often refer to the definition provided in this article, e.g. Bergenholtz, Bothma and Gouws (2015: 2), who write that when concerned with this function "[w]e need

knowledge to be able to understand certain non-linguistic signs and to act accordingly (interpretative knowledge), cf. Tarp (2008)." Here, the authors specify that (1) the signs are non-linguistic, and (2) the user's intention is to find out how to act or react based on what the sign means or communicates. However, not all sources that mention the interpretive function or situation claim that a (re)action is involved, e.g. Tarp (2011: 65), who states that it is a situation where people need information in order "to receive help to interpret and understand non-textual and non-verbal phenomena or symbols." However, on the same page, Tarp (2011) also says that the interpretive situation is "when somebody needs an explanation (*sometimes followed by recommendations to take action*) in order to interpret a sound, a symbol or some similar non-linguistic and non-verbal sign" (emphasis added). This is also mentioned by e.g. Bergenholtz and Bothma (2011: 62), who point out that "[i]n **interpretative situations** the potential user has a need for assistance to interpret a non-language sign of some kind, e.g. a traffic sign. The right interpretation is normally needed to be able to act in the correct way," with the word "normally" indicating that it is not always the case that the situation involves a following action. Thus, the definition of the interpretive function is yet not clear nor agreed upon by function theorists.

Only one lexicographical project has so far described its dictionary as having an interpretive function: Leroyer (2010) and Leroyer and Høy (2013) describe the Oenolex Burgundy wine dictionary, for which they have planned an interpretive function. This dictionary is to be used when the user wants to understand a certain wine label. However, the function has yet not been incorporated into this project because the customer thinks that there are too many labels to work with and thus that it is an impossible or time-consuming task (Leroyer, personal communication). Though not turned into a real project, Leroyer (2009) suggests that for a tourist dictionary, it would be useful to incorporate an interpretive function so that "help [would] be given to interpret signs and signals of recommendation ensuring the positive outcome of tourist experience" (2009: 115), e.g. the star ranking system found in Michelin guides. Also only at a conceptual level, Agerbo (2016) suggests that for a sports dictionary, an interpretive function would be relevant as users may have problems understanding e.g. the hand signals or cards given by the referee to a player in a match and that the user therefore needs an information tool that explains what they mean and why the players act the way they do according to these signals. However, so far, no function-based dictionary project with an interpretive function has been produced and published.

Assuming for now that the interpretive function is an independent lexicographical function (although this will be questioned later), some existing information tools also have an interpretive function though this is not directly stated in the outside matter. A number of the investigated sports tools in Agerbo (2016) contained data — some in the central list and some as part of the outer text — that may be analysed as supporting an interpretive function. In *The Dictionary of Sports* (1949), referee signals are thematically collected in the

back matter. Here, the images are supplemented by text explaining (1) what the referee is doing and (2) what his movement means. In *Webster's Sports Dictionary* (1976), the images are only supplemented by a short meaning explanation of the signal, not a description of the referee's movements. None of these two dictionaries apply cross-references, e.g. from *time-out* in the central list to the back matter or vice versa. In *Sports: The Complete Visual Reference* (2000), a number of flags used in motorsport are shown with explanations of their meanings. Interestingly, for the black flag with the orange disk, the explanation also tells the user how to react when this flag is shown, i.e. it includes an instruction. For archery, a description is given of the colour and sound signals⁴ used in archery to indicate when the archers can shoot. As in the example with the flags, the dictionary user is told what he can see (the colour) and how he should react to these signals. For American football and basketball, we yet again find examples of referee signals in the form of photos supplemented by the general meaning, e.g. "holding" (the verbal sign that is synonymous with the visual sign) and an elaborate text. Examples of such non-linguistic signs can also be found in dictionaries that are not connected to sports. For example, the *Dictionary of Symbols* (1991), which is divided into three sections (the symbol list in which the symbols are divided into thematic groups, the graphic index and the word index) has an interpretive function as its lemmata are symbols, not words made of alphabetical characters. The crowdsourced online *Emoji Dictionary* features a vast number of emojis, i.e. digital images (pictograms), used to express emotions or ideas, which are being incorporated more and more in people's everyday digital communication. A person reading an SMS may come across an emoji that she does not understand and will therefore access the *Emoji dictionary* in order to understand the meaning of this pictogram. The different emojis are structured thematically in broad categories, thus the user will have to search carefully to find the given emoji. In addition, general dictionaries may contain some non-linguistic signs, e.g. the signs %, § and ? (these are non-linguistic signs, which are synonymous with the corresponding linguistic signs, e.g. % is synonymous with the term *percent sign*), which for example are searchable and lemmatised in the online dictionary *Den Danske Netordbog* (En. *The Danish Internet Dictionary*).

It is easy to find dictionaries that apply non-linguistic signs in their explanations of certain lemmata, e.g. sports dictionaries and dictionaries of mathematics. However, these should not be confused with dictionaries that have an interpretive function. In the former case, non-linguistic signs are applied to describe linguistic signs, e.g. as in *The Concise Oxford Dictionary of Mathematics* (2013), whereas in the latter case, non-linguistic signs are explained with linguistic signs. Pictorial dictionaries may also be considered examples of interpretive dictionaries, but these incorporate images, which indeed are non-linguistic signs in the Peircean sense of the word, but most likely not the type of non-linguistic sign commented on by Tarp (2008) since these signs simply represent what they actually portray.

2. Defining the interpretive function

In the investigated literature that comments on the interpretive function, the authors usually provide a very general definition of this function without further elaboration:

Tarp (2008: 11)	to interpret signals and symbols in the surrounding world
Leroyer (2009: 115)	to interpret signs and signals
Bergenholtz and Bothma (2011: 62)	to interpret a non-language sign of some kind, e.g. a traffic sign
Tarp (2011: 65)	to interpret a sound, a symbol or some similar non-linguistic and non-verbal sign
Gallardo (2013: 87)	to interpret non-linguistic signs such as traffic signs, the explanatory tables of some manuals or the graphics common to manuals of economics
Bothma and Tarp (2014: 351)	to interpret and understand a sign, signal, symbol etc.
Fuertes-Olivera and Tarp (2014: 51)	to understand or interpret a specific phenomenon, sign, symbol, text, etc.
Agerbo (2016: 24)	to interpret signals and symbols in the surrounding world

Though these eight definitions all define the same function, they differ in two important points:

Question 1: Do you need to understand the given sign in order to carry out *a certain act*?

Question 2: What *types of sign* are relevant for the interpretive function?

The following two subsections discuss these two points, i.e. (non)-acting and (non)-linguistic signs, in relation to the interpretive function.

2.1 Acting or not acting

Some of the above-mentioned definitions include acting as part of the interpretive function, others mention it as possible, but not mandatory, and some do not include it at all, i.e. there are three different perceptions of this function. If we argue that the user always or sometimes has to act or react, there is a certain overlap with the operative situation. In Bergenholtz, Bothma and Gouws (2015: 9),

the process and results of operative and interpretive situations are described as follows (emphasis added):

- operative situations:
 - *knowledge problem how to act* →
 - information source →
 - act in the world after having got the needed information →
 - meta-reflections
- interpretive situations:
 - *acting problem reacting to a sign or a symbol* →
 - information source →
 - acting or not acting in the world after having obtained the needed information →
 - meta-reflections

The emphasised parts show that the authors argue that operative situations involve a knowledge problem concerning how to act, whereas interpretive situations involve an acting problem on how to react to a sign (it must be assumed that the authors are talking about non-linguistic signs and not signs in general, i.e. linguistic as well as non-linguistic signs). This, however, is not described clearly enough and could instead be formulated in the following way:

- operative situations:
 - *knowledge problem concerning how to act* →
 - information source →
 - understand given problem based on the information (instruction) →
 - act in the world, according to one's skills, after having received the needed information →
 - (meta-reflections)
- interpretive situations:
 - *knowledge problem concerning the meaning of a non-linguistic sign and possibly how to act according to it* →
 - information source →
 - understand given problem based on the information (explanation + possibly instruction) →
 - act or not act in the world after having obtained the needed information →
 - (meta-reflections)

These process descriptions show that in the interpretive situation, the person first needs to understand a certain non-linguistic sign, and then he *may* have to act in a certain way according to the piece of information provided by the selected information tool. A question that arises from this is whether such

interpretive situations which involve acting actually consist of one or two connected phases: (1) interpretation and (2) operation. For comparison, in the case of translation, Fuertes-Olivera and Tarp (2014) argue that from a lexicographical perspective, the translation process consists of three main phases and a number of sub-phases that among other things involve understanding, writing and revising a text. They write that "[i]n all these phases and sub-phases, the translators may experience various types of need which require specific types of *lexicographical data* as well as allowance for specific types of *data access*, in order to be satisfied" (2014: 67). In the same way, it could be argued that for interpretation, a user will also experience various types of need that require different types of data. However, there is an important difference between these two situations: the translation process always consists of three main phases, and the different translators may experience different problems in relation to these phases; but the interpretation process will not always be the same because the type of non-linguistic sign and the type of user will co-determine whether or not it is relevant for the user to act. For example, a chauffeur may have to understand a certain traffic signal because he needs to act according to it⁵; his passenger may also want to understand the signal, but he is not in a situation or role where he needs to act according to it. Thus, acting should therefore not be considered a part of the definition of the interpretive function. This means that the process description looks as follows:

- interpretive situations:
 - knowledge problem concerning the meaning of a non-linguistic sign →
 - information source →
 - understand given problem based on the information (explanation) →
 - (meta-reflections)

The relation between acting and not acting may not only be a relevant discussion for interpretation, but also for reception though this is usually not part of the definition of this function. Bergenholtz, Bothma and Gouws (2015: 9) write that the communicative situation (e.g. reception and production) involves "acting in a text". However, this is a different meaning of the word *act*. When someone makes a lookup in an information tool, it could be argued that this will always be followed by some kind of action or acting, whether it is applying the information in the understanding of a text (reception) or when storing information in one's memory (cognition). But if we look at the definition of *acting* in relation to the operative function, Tarp (2008: 185) writes that operation is "where the user needs advice and instructions in order to perform any kind of mental or manual action, e.g. to operate a machine." Thus, the action that the person wants to perform requires that the person should be provided with advice and instructions, e.g. situations where a person wants to bake a cake (action) and uses a cookbook (instructions) in order to do so, or where a person wants to assemble a piece of furniture (action) and consults a user manual (instructions) in order to do so. Acting in a text and storing information in

one's memory are thus not the types of action referred to in the definition of the operative function as these are not actions requiring advice or instructions; a reader does not get information on how to apply an explanation of a word in a given text. Thus, the acting part in the case of communication and cognition is not relevant for the actual purpose of the information search — and neither is it for interpretation.

2.2 Linguistic and non-linguistic signs

In the definitions presented earlier, either reference is made to non-linguistic signs, or the words *signal*, *symbol* and/or *sign* are mentioned. What the authors in general mean when they talk about the interpretive function is that people are confronted with and want to understand signs that are not verbal (related to language) and therefore are not usually treated in dictionaries; for instance, a graph in a maths book, the mating signal of a plant hopper, a red rash on one's body, a certain (cultural) gesture made by a native Spanish speaker to a native Danish speaker, or a referee signal in rugby. As these examples show, there are many different types of sign that could be considered in the discussion of the interpretive function. Inspired by Pierce's typology of signs, a list of these could look as the following⁶ (notice that one sub-sign does not exclude another, e.g. a sign can be expressed with one's body *and* be connected to a certain culture, cf. the second bullet below):

- Linguistic signs (the ones traditionally treated in lexicography)
- Signs invented or produced and transmitted by humans consciously and creatively to convey a message (intentional/motivated)
 - Using one's body to express meaning
 - Using a machine, e.g. a traffic light, to express meaning
 - Official, semi-official or intuitive signs
 - Signs connected to a certain field; signs connected to a certain culture
 - Tables, graphs etc. that require some background information to be understood
- Signs imitating what they represent, e.g. a drawing of a car or a flower
- Signs produced and transmitted by humans subconsciously or genetically (non-intentional/unmotivated) and which cannot always be ascribed a specific meaning
 - Specific meaning: e.g. a reaction or behaviour characteristic of a specific person
 - No specific meaning: e.g. a person making a burp
- Non-human produced signs (including animal sounds) attributed with meaning by humans, i.e. natural signs, e.g. a dark cloud in the sky indicating rain

-
- Non-human produced signs that cannot be ascribed a specific meaning, e.g. the sound when a car drives too close to an item and gets scratched

Many of these signs are typically not incorporated in lexicographical tools, but why should they not be included if a user has an information problem related to one of these signs? However, a much more interesting question in the given context is the following: why should a distinction between linguistic and non-linguistic signs result in different (main) functions; why are non-linguistic signs not relevant for cognitive tools and operative tools, and why should a distinction be made between reception and interpretation only in terms of the sign type? For example, a user may read a text in which a certain sign, e.g. a certain graph, is shown, and he wants to understand what it means — this corresponds to reception, though it involves the understanding of a non-linguistic sign; or a person may wish to know more about a certain non-linguistic sign as part of a school assignment — this corresponds to cognition, though, again, it concerns knowledge about a non-linguistic sign; or a person may want to know how to act according to a certain traffic sign — this corresponds to operation, though, as in the two previous examples, it concerns knowledge about a non-linguistic sign. It therefore seems that a distinction based on linguistic sign versus non-linguistic sign is not decisive for the function and that the interpretive function should be rejected as a main function since the other functions would be able to cover both linguistic and non-linguistic signs.

A potential objection to this rejection of the interpretive function would be that not all types of non-linguistic sign can be connected to one of the three main functions, i.e. communication, cognition or operation, e.g. a dark cloud in the sky (which indicates that it may soon start raining). It is definitely possible to imagine an information tool in which different types of weather signs are explained, e.g. what it means when the temperature suddenly drops or when you can sense a certain smell in the air; thus, these signs are assumingly relevant for reception (communication). Nevertheless, in the traditional understanding of communication, a message (sign) would involve a sender and a receiver. Yet, the determining factor for reception is that a person sees a sign that he wants to understand, i.e. the relevant factors are the receiver and his information need, not the sender. For some signs, we know who the sender is (a woman uses the word "palindrome" in a conversation with a man and he does not know what this word means); for others, we do not know who it is (a person has put up a sign in a park which many people read daily, but without them knowing who the sender is); for a third category, a person is not aware that he is making/sending a sign to someone (a man is yawning unintentionally because he is bored listening to his wife talking, and this is seen by a person sitting next to the couple); and for a fourth category, there is no particular sender (a person gets a rash and wants to understand what it means). In the case of text production and translation, only signs made by human beings are lexicographically relevant, but for reception — and also cognition — all types of sign are (or could potentially) be lexicographically relevant.

3. Future tools: user situations and needs related to search options and signs

Until now, general dictionaries have mainly been concerned with linguistic signs, and it is only some specialized dictionaries that to some extent incorporate non-linguistic signs though the search methods and presentations are primitive. For example, instead of grouping these non-linguistic signs thematically (cf. *Dictionary of Symbols*), it would be more useful for the user if he could, in an e-tool, either search for a non-linguistic sign using an actual photo or by typing in keywords describing the sign he is looking for, in this way making the software search in an explanation field connected to the relevant lemma. Below, the possible search options and user needs connected to specific non-linguistic and their related linguistic lemmata are suggested for future reference works.

3.1 *Yellow card* and (linguistic sign and non-linguistic sign)

- **Reception:** You want to know what the term *yellow card* or the photo of a yellow card means so you search either with the linguistic or the non-linguistic sign in the search field
- **Production:** You want to know how to write the term in the plural so you search with the linguistic sign in the search field
- **Translation:** You want to know what the term means (its equivalent) in a different language so you search with the linguistic sign in the search field
- **Cognition:** You want to know more about yellow cards in sports in general so you search with either the linguistic or the non-linguistic sign in the search field
- **Operation:** You want to know how/when to give someone a yellow card so you search with either the linguistic or the non-linguistic sign in the search field (NB: this is only relevant for referees)

3.2 A specific bird sound (non-linguistic sign)

- **Reception:** You want to know what robins communicate to each other so you insert the non-linguistic sign (your own recording of robins communicating) in the search field (NB: whether this is plausible is questionable)
- **Cognition:** You want to know what kind of bird makes a specific sound so you insert the non-linguistic sign (your own recording of an unknown bird sound) in the search field
- **Cognition:** You want to know as much as you can about robins for a certain purpose so you insert the non-linguistic sign (a recording of a robin singing, either your own or from some other source) in the search field

3.3 The emoji called *face with tears of joy* and 🤩 (linguistic sign and non-linguistic sign)

- **Reception:** You want to know what this emoji means so you search either with the linguistic or the non-linguistic sign in the search field (NB: for the linguistic sign, it is a problem that there is no official name for each emoji)
- **Production:** You want to know how you write a text message using this non-linguistic sign so you search with the non-linguistic sign in the search field; or you want to know how to write the term (the linguistic sign) in the plural (is it *faces with tears of joy* or *face with tears of joys*?)
- **Cognition:** You want to know as much as you can about this emoji so you search either with the linguistic or the non-linguistic sign in the search field

3.4 A specific occurrence of a rash:  (non-linguistic sign)

- **Cognition:** You want to know what caused the rash (and possibly how to get rid of it) so you search with the non-linguistic sign (a photo of the rash) or with keywords connected with Boolean operators (e.g. rash + back + "red dots")
- **Operation:** You want to know how to get rid of the rash so you search either with the photo or (if you know what caused it) the name of the "thing" or phenomenon that caused the rash (e.g. tree nut allergy)

3.5 The traffic sign called the *full stop sign* and  (linguistic sign and non-linguistic sign)

- **Reception:** You want to know what the term *full stop sign* or the photo of this traffic sign means so you search either with the linguistic or the non-linguistic sign in the search field
- **Production:** You want to know how to write the term in the plural so you search with the linguistic sign in the search field
- **Translation:** You want to know what the term means in a different language so you search with the linguistic sign; or you want to know if this traffic sign looks different in a different culture so you search with the non-linguistic sign
- **Cognition:** You want to know more about this traffic sign so you search either with the linguistic or the non-linguistic sign in the search field

Not all of the resulting dictionary articles would occur in a general dictionary. For example, the operative article for *yellow card* could occur in a set of rules or a referee's handbook structured according to lexicographical principles. Bird sounds could occur in a specialized cognitive e-tool for ornithologists, which would relate different sounds to different birds and explain when, for how

long, where, etc. they sing. The dictionary article of a rash could occur in an online medical handbook for specialists or for laymen who want to know what it is a symptom of and how (if possible) they can treat it. Images of traffic signs could occur in general dictionaries as lemmata, but future reference tools might also be integrated into car computers in order to register traffic signs on the road so the computer could inform the driver about the meaning of the given sign.

As these examples show, though some of the future information tools may resemble traditional dictionaries, many others will be innovative and therefore unique in their design and search options. This is not only due to the integration of computer and information technologies in lexicography, but also to the formulation of an independent lexicographical theory based upon dictionary functions, cf. Nielsen (2013: 356), who states that "[w]hat objects people will regard as dictionaries may change, however, owing to a range of factors, including the types of need identified, the media available, and the types of help provided." Hanks (2012: 81) mentions that "[n]o doubt it would be technically straightforward enough to include film clips of typical elephant behaviour, including the sound of elephants trumpeting. Perhaps the technology is not far away by which we shall be able to sit at our computer and touch a simulation of an elephant's skin or smell a bull elephant in musth." Moreover, Lew (2010: 299) writes, "[p]erhaps a radical solution becomes available in the near future in the form of projector glasses, which would display an image on the inside of their lenses for the wearer to view, thus utilizing a large fraction of the human field of vision. In the not-so-near future, 3D holographic projectors may become a possibility⁷, but for the moment this is the stuff of science *fiction* rather than science." Though these suggestions may seem unrealistic at the moment, we, lexicographers, should not be afraid of suggesting such innovative moves as this is the only way to move forward. Kwary (2012), for example, suggests that calculators should be incorporated into a reception e-dictionary of economics, and he also shows that the four components tooltip, speech recognition, auto-summarize and definition-finder could be incorporated into a similar type of dictionary. Of course, as argued by Nielsen (2013: 370), the lexicographical platform should be supported by two pillars: a practical one (i.e. available technical features), and a theoretical one, i.e. needs-adapted data presentation. Thus, it is not simply a matter of using new technologies, but applying these in a way that the user can easily extract the information he needs to solve his information problem.

Some characteristics of future lexicographical information tools

- New types of information needs (e.g. needs for acting according to or understanding non-verbal signs) can be satisfied with these tools
- New types of data (e.g. videos and smells) are incorporated
- New search options (e.g. searches made with sounds and images in addition to words and letters) are available

- New ways of structuring and presenting data in order to ensure a more dynamic approach to the incorporation of data

4. The interpretive function: not to be

A dictionary function should be based on a certain type of information need that a person experiences in a specific type of situation, not on whether the problem involves a linguistic or a non-linguistic sign. The main reason for having called the interpretive function an independent dictionary function has probably been that lexicographers have not been used to lemmatising non-linguistic signs in lexicographical information tools, and therefore it may appear as a different situation when we come across such signs. However, the basic function is the same whether we want to, for example, understand what a linguistic sign or a non-linguistic sign means or how to react to a given type of sign. Therefore, the function theory does not operate with four main functions, but three: communicative, cognitive and operative. The communicative function has been well-researched, which has resulted in a further division into the following four subfunctions: reception, production, translation and revision/marking. Though cognition has been considered a main function together with the communicative function since the first mention of the function theory in the beginning of the 1990s, it is still unclear what it actually covers. We may distinguish between punctual information problems (e.g. when was Napoleon born?) and general information problems (e.g. I want to know as much as is relevant in my situation about the Napoleonic Wars), but further subfunctions have yet not been identified. The operative function was first mentioned in lexicographical literature 10 years ago, but surprisingly, no research has yet been conducted on this function, though see Agerbo (in review). For all of these three main functions, both non-linguistic and linguistic signs could be relevant. Incorporating non-linguistic signs in paper format is problematic since these cannot be incorporated in the same way as linguistic signs; but with the computer technologies that are available today and probably will be available in the near future, it will become much easier to incorporate such non-linguistic lemmata in lexicographical e-tools. It is a matter of using the technical tools that are available and not being afraid of thinking out of the box when trying to satisfy the needs of the intended dictionary users.

Endnotes

1. In some articles, this function is called *interpretative*, in others *interpretive*. I choose to apply the term *interpretive function* in this article.
2. Of course, the word *exists* should not be understood as if there is some kind of objective reality in which this function can be found; it means that it is relevant to construct such a category to work with in lexicographical theory.

3. In this context, the word *intentionally* means that the producer directly writes that the given tool has an interpretive function or that the producer in some other way makes it clear that the purpose of the tool is to help the user understand a non-linguistic sign (this would typically be mentioned in the introduction, i.e. the outside matter).
4. Of course, sounds cannot be directly represented in a paper dictionary, only by description, but this would be possible in an electronic dictionary.
5. Kwary (2012: 32) makes a distinction between real needs and ancillary needs of which the latter may occur during dictionary consultation. This could be connected to the given case in which a person may look up in an information tool to understand the meaning of a certain traffic signal, and based on this, he also wants to know how to act according to the sign.
6. It may be that other types of sign should be added to this list.
7. Nielsen (2013: 368) makes this same suggestion, "[i]t is possible that online dictionaries may present data in three-dimensional form, including holograms."

References

Dictionaries

- Bergenholtz, H. (Ed.).** 2017. *Den Danske Netordbog*. Odense: ordbogen.com.
- Clapham, C. and J. Nicholson.** 2013. *The Concise Oxford Dictionary of Mathematics*. Fifth edition. Oxford: Oxford University Press.
- Copeland, R.** 1976. *Webster's Sports Dictionary*. Springfield, Mass.: G. & C. Merriam.
- Cummings, P. (Ed.).** 1949. *The Dictionary of Sports*. New York: A.S. Barnes & Company.
- Emoji Dictionary* = <https://emojidictionary.emojifoundation.com>.
- Fortin, F.** 2000. *Sports: The Complete Visual Reference*. Willowdale, Ont.: Firefly Books.
- Liungman, C.G.** 1991. *Dictionary of Symbols*. New York: Norton.

Other Literature

- Agerbo, H.** 2016. The Incorporation of Specialised Data in Lexicographical Meaning Explanations: A Discussion Based on Sports and Fitness Terms. *Lexikos* 26: 1-35.
- Agerbo, H.** In review. *Monofunctional and Polyfunctional Information Tools with an Operative Function*.
- Bergenholtz, H. and T.J.D. Bothma.** 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.
- 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.
- Bothma, T.J.D. and S. Tarp.** 2012. Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.
- Bothma, T.J.D. and S. Tarp.** 2014. Why Relevance Theory is Relevant for Lexicography. *Lexicographica* 30: 350-378.
- Fernández, E.S. and B.P.C. de Alba.** 2011. Conclusions: Ten Key Issues in e-Lxicography for the Future. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*: 305-311. London: Continuum.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter.

- Gallardo, Á.R.** 2013. The Function Theory and its Application on Manuals of Economics. *Hermes. Journal of Language and Communication in Business* 50: 83-95.
- Hanks, P.** 2012. Corpus Evidence and Electronic Lexicography. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 57-82. Oxford: Oxford University Press.
- Kwary, D.A.** 2012. Adaptive Hypermedia and User-oriented Data for Online Dictionaries: A Case Study on an English Dictionary of Finance for Indonesian Students. *International Journal of Lexicography* 25(1): 30-49.
- Leroyer, P.** 2009. Balancing the Tools: The Functional Transformation of Lexicographical Tools for Tourists. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century: In Honour of Henning Bergenholtz*: 103-122. Amsterdam/Philadelphia: John Benjamins Publishing.
- Leroyer, P.** 2010. Ej blot til lyst — Konsultation og navigation i leksikografiske informationsværktøjer. *Nordiske studier i leksikografi* 10: 313-328.
- Leroyer, P. and A. Høy.** 2013. Vinsmagningsordbogen Oenolex Bourgogne. En Milepæl i Pragmatisk Fagleksikografi. *Nordiske Studier i Leksikografi* 12: 287-302.
- Tarp, S.** 2008. Revival of a Dusty Old Profession. *Hermes. Journal of Language and Communication in Business* 41: 175-188.
- Tarp, S.** 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 54-70. London/New York: Continuum.

Prolingua se bydrae tot terminologieontwikkeling in Afrikaans

Mariëtta Alberts, *Navorsingseenheid: Taal en Literatuur
in die SA Konteks, Noordwes-Universiteit, Potchefstroomkampus,
Potchefstroom, Suid-Afrika (albertsmarietta@gmail.com)*

Abstract: Prolingua's Contribution to Terminology Development in Afrikaans. The research problem addressed in this article deals with the isolated development of Afrikaans terminology in a multilingual dispensation. The research problem manifests itself in three ways: the need for such terminology; the sustainability thereof; and whether a single institution, such as Prolingua, could maintain such development. Prolingua, an association for English and Afrikaans language practitioners, was founded in 1950. At the time it was known as the *Transvaal Association of Municipal Translators*. As the name indicated, the members were mainly municipal officials. Municipal services included several services that, at the time needed bilingual terms for various areas in the workplace. The country's bilingual language policy then required terms to be translated and created for all municipal departments — suddenly there was a need for Afrikaans terms. The aim of the newly established association was, among other things, to translate, create and standardise terminology and to discuss problems with translation. Later more attention was given to terminology-related problems and the association became more inclusive and inviting for language practitioners from other institutions to join its ranks. This article deals with the need for Afrikaans terminology and gives information on the history of Prolingua and its contribution to terminology development in Afrikaans, i.e. the Prolingua term list and other products and services. The contributions made by external collaborators regarding terminology development also receive attention. Prolingua's cooperation with other institutions is discussed. The article concludes with recommendations for a way forward.

Keywords: CLASSIC ORIGIN, COLLABORATOR, COOPERATION, DATABASE, DIGITISATION, DIGITALISATION, HARMONISATION, LANGUAGE POLICY, LANGUAGE PRACTITIONER, NEOLOGISM, ONLINE TERMINOLOGY PRODUCTS, STANDARDISATION, TECHNICAL DICTIONARY, TERM CREATION, TERM LIST, TERMINOLOGY DEVELOPMENT, TERMINOLOGY, TERMINOLOGY MANAGEMENT SYSTEM

Opsomming: Die navorsingsprobleem wat in hierdie artikel bespreek word, handel oor die geïsoleerde ontwikkeling van Afrikaanse terminologie in 'n meertalige bestel. Die probleem manifesteer homself op drieërlei wyse: die behoefte aan sodanige terminologie; die volhoubaarheid daarvan; en of 'n enkele instansie, soos Prolingua, daartoe in staat is om sodanige ontwikkeling te handhaaf. Prolingua, 'n vakvereniging vir Engelse en Afrikaanse taalpraktisyne, is in 1950 gestig. Dit het toe bekend gestaan as die *Transvaalse Vereniging van Munisipale Vertalers*. Soos die naam aandui, was die lede hoofsaaklik munisipale beamptes. Munisipale dienste het verskeie dienste

ingesluit wat gedurende hierdie tydperk tweetalige terme vir verskeie gebiede in die werksplek benodig het. Die land se tweetalige beleid het vereis dat terme vertaal en geskep moes word vir alle munisipale departemente — skielik was daar 'n behoefte aan Afrikaanse terme. Die doel van die nuutgestigte vereniging was om onder andere te vertaal, te skep en terminologie te standaardiseer en probleme wat met vertaling ondervind is, te bespreek. Later is meer klem op terminologieverwante probleme gelê en die vereniging het meer inklusief en aanloklik vir taalpraktisyne verbonde aan ander instansies geraak om by sy geleedere aan te sluit. Hierdie artikel gee aandag aan die behoefte aan Afrikaanse terminologie en die ontstaan van Prolingua en sy bydrae tot terminologieontwikkeling in Afrikaans, soos die Prolingua-termlys en ander produkte en dienste. Die bydraes deur eksterne medewerkers ten opsigte van terminologieontwikkeling kry ook aandag. Prolingua se samewerking met ander instansies word bespreek. Die artikel word afgesluit met aanbevelings oor die pad vorentoe.

Sleutelwoorde: AANLYN TERMINOLOGIEPRODUKTE, DATABASIS, DIGITISERING, DIGITALISERING, HARMONISERING, KLASSIEKE OORSPRONG, MEDEWERKER, NEOLOGISME, SAMEWERKING, STANDAARDISERING, TAALBELEID, TAALPRAKTISYN, TERMINOLOGIE, TERMINOLOGIEBESTUURSTELSEL, TERMINOLOGIEONTWIKKELING, TERMLYS, TERMSKEPPING, VAKWOORDEBOEK

1. Inleiding

Vaktaal is zowel een zegen als een vloek. Het is een zegen om met nauwkeurig gedefinieerde termen snel en eenduidig informatie uit te wisselen over handelingen en objecten binnen een vakgebied. Maar buiten de kring van vakgenoten wordt vakjargon niet begrepen en evenmin gewaardeerd.

(Burger en De Jong 1997: 169)

Die navorsingsproblematiek wat in hierdie ondersoek onder die loep kom, handel oor die geïsoleerde ontwikkeling van Afrikaanse terminologie in 'n meertalige bestel. Die probleem manifesteer homself op drieërlei wyse:

- die behoefte aan Afrikaanse terminologie wat nie deel is van meertalige termlyste, vakwoordeboeke of termdatabasisse nie;
- die volhoubaarheid van die geïsoleerde ontwikkeling van Afrikaanse terminologie; en
- die moontlikheid dat 'n enkele instansie, soos Prolingua, daartoe in staat is om sodanige terminologieontwikkeling te handhaaf en selfs uit te bou.

Terminologieontwikkeling vind nie in isolasie plaas nie en geskied binne die konteks van die taalbeleid van 'n regering. Suid-Afrika beskik oor voortreflike wetgewing wat die talebestel behoort te beskerm en taalontwikkeling te fasiliteer. Die Grondwet van die Republiek van Suid-Afrika (1996) maak voorsiening vir die ontwikkeling van al elf amptelike tale en die Nasionale Taalbeleidsraamwerk van Maart 2003 gee uiting aan die grondwetlike vereistes vir taalgebruik en -ontwikkeling. Verder gee die wetgewing oor die Gebruik van

die Amptelike Tale (2012) en die voorgestelde regulasies (2013) riglyne oor taalgebruik op nasionale, provinsiale en plaaslike regeringsvlak. Die Suid-Afrikaanse Taalpraktisynsraad (2014) maak onder andere voorsiening daarvoor dat taalpraktisyns oor die nodige vaardighede beskik om in die taalpraktyk werkzaam te wees. Wetgewing vereis dus by implikasie dat daar tans aan meertalige terminologieontwikkeling aandag gegee moet word (Alberts 2017: 150-163).

Terminologie fasiliteer taalontwikkeling, die skep van kennis en die verspreiding van inligting, want terminologie is die medium waardeur inligting versprei en kennis oorgedra word. Wanneer dit by vaktaal kom, moet die gestandaardiseerde terminologieë van verskeie vakgebiede, domeine en werkverwante belangegebiede dus in plek wees vir suksesvolle konseptualisering, om inligting te ontsluit en kennis te orden, vir inligtingoordrag en om kennisverwerking te fasiliteer. Daar is derhalwe 'n voortdurende behoefte aan terminologie (Alberts 1998; Alberts 2003: 128).

As gevolg van die verskeidenheid beroepe wat deur die lede van 'n gemeenskap beoefen word, ontstaan daar 'n aantal beroepstale of tegnolekte wat ook vaktaal genoem word. Elke beroep, wetenskaplike of tegnologiese rigting ontwikkel sy eie vaktaal en vir elke nuwe vakdissipline wat ontwikkel, ontstaan daar 'n nuwe vaktaal met 'n unieke terminologie. Hierdie vaktale se terminologie word gedokumenteer en aan die gemeenskap beskikbaar gestel om sodoende by te dra tot gestandaardiseerde vakkommunikasie (vgl. Alberts 1998; Alberts 2003: 128).

Tegnologiese vooruitgang, nuwe spesialisvelde en die tempo van nuwe ontwikkelings plaas vakkundiges en taalpraktisyns midde-in die wêreldwye kennisontploffing waarbinne effektief gekommunikeer moet word (Luther 2005). Die behoefte aan ondubbelsinnige kommunikasie in die teoretiese en toegepaste velde van menslike aktiwiteit neem gevolglik steeds toe (Mollema en Alberts 2011; Alberts en Mollema 2013). Hierdie wêreldwye tendens beïnvloed die werksomgewing en die manier waarop gemeenskappe daaglik interaktief verkeer en daar moet gevolglik voortdurend kreatief gedink word oor termskepping (Prolingua 2015: 1). 'n Groot behoefte aan meertalige terminologieskepping ontstaan aangesien verskillende sake in die tale van die onderskeie gemeenskappe bedryf word (Mtintsilana en Morris 1988: 109). Taalpraktisyns moet as fasiliteerders dien om vakinligting in verskillende tale aan vakgemeenskappe toeganklik te maak.

Die ontwikkeling van die verskillende inheemse tale het deur verskeie fases plaasgevind. Afrikaans het van 'n ondergeskikte taal ontwikkel tot 'n hoëfunksietaal wat op enige vakgebied of terrein gebruik kan word. Waar Afrikaans tans op 'n handhaaffase moes wees, is die taal egter terug na 'n stryd om oorlewing. Die Afrikatale word ten spyte van hul amptelike taalstatus nie voldoende ontwikkel nie. Engels is steeds die taal van voorkeur vir baie en dit wil voorkom of al die ander amptelike tale met moeite van die hegemonie van Engels sal ontkom (Alberts 2017: 140-150). Dit is belangrik om daarop te let dat 'n taal wat funksioneel is 'n taal met 'n toekoms is: "Hoe meer funksies 'n taal

verloor, hoe vinniger gaan die taal se gebruikswaarde agteruit; hoe kleiner krimp sy taalwêreld; en hoe makliker verskuif sy taalgemeenskap na 'n ander, meer bruikbare taal" (Buys 2014: 4). Sonder (vak)taal kan daar immers nie sprake wees van behoorlike kommunikasie nie.

'n Blik op die historiese ontwikkeling van vaktaal in Suid-Afrika toon dat dit 'n moeisame en langsame proses was (Alberts 2017: 163-170). As daar egter gekyk word na die hoeveelheid vakwoordeboeke en termlyste wat amptelik sedert 1950 geproduseer is, is vaktaalontwikkeling in Suid-Afrika 'n suksesverhaal (NTD 1998). 'n Oudit wat in 2015 in opdrag van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns en Dagbreek Trust uitgevoer is, toon dat daar 362 vakwoordeboektitels was. Hoewel sommige vakwoordeboeke en termlyste uit druk is, blyk dit dat daar aan 'n verskeidenheid dissiplines aandag gegee is, hoewel daar egter ook baie dupliserings van projekte was ten spyte van vele koördinerende liggame (Alberts 2015a; Alberts 2017: 165, 171). Daar word gesê dat spesialiskennis, afhange van die betrokke spesialisveld, elke vyf tot vyftien jaar verdubbel (Luther 2005). Leemtes en toekomstige behoeftes kan bepaal word deur na bestaande vakwoordeboekprodukte te kyk.

Afrikaanse vaktaal is sedert die vyftigerjare aktief ontwikkel en 'n verskeidenheid vakgebiede se terminologieë is sedertdien gedokumenteer en gestandaardiseer. Die getal vakwoordeboekpublikasies met Afrikaans as een van die behandelde tale het van sowat 150 beskikbare titels in 1990 (baie van hulle toe reeds uit druk, maar destyds nog verkrygbaar) tot minder as 20 titels in 2005 gekrimp (Luther 2005).

In Suid-Afrika is die primêre terminologiedokumentering steeds in Engels en by sekondêre termskepping word die Afrikaanse termekwivalente (en definisies) gewoonlik eerste bygevoeg — bloot omdat Afrikaanse termekwivalente in baie gevalle reeds beskikbaar is. Daarna volg die Afrikatale as doeltale. Lesse wat reeds uit verskeie meertalige terminologieprojekte geleer is, is dat Afrikaanse terme en definisies dikwels meer toeganklik is vir die Afrikataalmedewerkers as die Engelse weergawes.

Die ontwikkelaars van Afrikataalterminologie behoort te leer uit die foute wat met die Afrikaanse vaktaalontwikkeling begaan is. Die Afrikaanse vakwoordeboeke en termlyste was merendeels glossariums sonder bykomende, toeligende terminologiese inligting. Die huidige samestellers van vakwoordeboeke, termlyste en termdatabasisse is gelukkig reeds besig om die dokumenteringsproses te verbeter, want verklarings, voorbeeldsinne en ander tersaaklike inligting word tans bygevoeg. Dit is bloot makliker om die inligting verbonde aan 'n konsep behoorlik te ontsluit indien daar 'n behoorlike verklaring van die konsep is en die brontaalterm in konteks verskyn. Rekenaartoepassings vergemaklik ook verskeie (sleur)prosesse wat deel is van die terminologieontwikkelingsproses. Meertalige termlyste oor 'n verskeidenheid vakterreine is deesdae ook aanlyn beskikbaar.

Prolingua se bydrae tot terminologieontwikkeling moet teen hierdie agtergrond beskou word. Maar wie of wat is Prolingua? Prolingua is 'n 'Vereniging

vir Afrikaanse en Engelse Taalpraktisyns'. 'n Mens sou met reg kon vra of daar tans in Suid-Afrika met sy meertalige taalbeleid nog plek vir en 'n behoefte aan 'n vakvereniging sou kon wees wat aan slegs twee van die amptelike tale aandag gee. Die antwoord is onomwonde positief, want daar is steeds taalpraktisyns wat in die huidige meertalige bestel hul daaglikse brood in Afrikaans en Engels verdien.

Die taalpraktisyns wat daaglik met Engels en Afrikaans omgaan, kom met verskeie (gewoonlik Engelse) terme in aanraking wat nie noodwendig reeds in vakwoordeboeke gedokumenteer en gestandaardiseer is nie. Hulle moet dan uiteraard vinnig (gewoonlik Afrikaanse) termekwivalente skep. Die lede van hierdie vereniging bied die nodige ondersteuning aan mekaar en aan ander taalpraktisyns om 'n aanduiding te gee of die termekwivalente reeds in gebruik is, maar nog net nie in vakwoordeboeke opgeteken is nie, of om waar nodig nuwe terme te help skep. Dit blyk dat Afrikaanse terme, wat as deel van meertalige termlyste, vakwoordeboeke of termdatabasisse gedokumenteer word, dikwels reeds in ander vaktaalbronne bestaan en daar is dus nie sprake van nuutskeppings in Afrikaans nie.

2. Prolingua: Agtergrond

Op 8 Mei 1925 word Afrikaans naas Engels en Nederlands as amptelike taal van Suid-Afrika erken. Dit was 'n tyd waarin daar skielik vele uitdagings aan staatsamptenare en munisipale werkers gestel is wat skielik ook in Afrikaans moes werk en skryf. Verskeie vertaal- en terminologiekantore is by staatsdepartemente en munisipale kantore gestig (Alberts 2015a), maar die taalpraktisyns was nie opgeleide vertalers of terminoloë nie — hulle het grade in letterkunde en linguïstiek gehad, maar geen formele vertaal- of terminologieopleiding nie. Niemand het in daardie stadium kennis gedra van vertaalteorieë of termskeppingspraktyke nie — vertaalwerk en termskepping het gevolglik gewoon intuïtief plaasgevind. Daar is van die taalpraktisyns verwag om hierdie praktykgerigte opleiding in die loop van hul normale werkslading te bekom. Daar was egter geen gebrek aan entoesiasme en ywer nie en dikwels was die motivering bloot om te verseker dat Afrikaans sy man teen 'n wêreldtaal soos Engels kon staan (Prolingua 2015: 1).

Een van die grootse probleme was die gebrek aan vakwoordeboeke. Die taalpraktisyns verbonde aan staatsdepartemente en munisipaliteite het gevolglik begin om hul eie termlyste vir interne gebruik saam te stel. Daar was by slegs 'n paar van die taalkantore ook terminologieseksies wat tweetalige vakwoordeboeke saamgestel en gepubliseer het. Die probleem met die interne termlyste was dat daar geen koördinering van werksaamhede of terme was nie wat aanleiding gegee het tot 'n proliferasie van terme wat nie gestandaardiseer was nie.

Op Saterdag 28 Oktober 1950, skaars 25 jaar na die amptelike erkenning van Afrikaans as amptelike taal van Suid-Afrika, is 'n vereniging, die *Trans-*

vaalse Vereniging van Munisipale Vertalers, gestig om hoofsaaklik vertalers verbonde aan verskeie munisipaliteite met vertaalprobleme van Engels na Afrikaans te help.

Die *Vereniging van Munisipale Vertalers* het met ander taalkantore en staatsdepartemente soos die Terminologieseksie van die Taaldiensburo, die Buro van Standaarde, die Taaldiens van die Weermag, die Spoorwegtaalburo, die Poskantoorstaaldiens en provinsiale administrasies saamgewerk. Taalpraktisyns verbonde aan hierdie en ander instellings het ook lede van die Vereniging geword. Die lede het gewoonlik op Saterdag byeengekom om terme te vertaal of te skep. Hulle het op eie koste die vergaderings bygewoon. Aangesien baie werknemers verbonde aan die munisipaliteite destyds lede was van die Vereniging is terminologievergaderings om die beurt by die Stadsraad van Pretoria en by die Stadsraad van Johannesburg gehou.

Munisipale dienste het verskeie departemente en terreine ondersteun waarvoor Afrikaanse terme voorsien moes word, byvoorbeeld *kragopwekking en -verspreiding, riool, afvalwater en watersuiwering, gasfabrieke, brandweer, bus- en tremdienste, tegniese werkwinkels, markte, slagpale, afvalverwydering, vullishope, krematoriums, begraafplase, parke, padbou en -instandhouding, verkeersbeheer, opmeting, waarderings, stedelike beplanning, argitekte, regsadviseurs, rekenmeesters, begrotings, administrasie, personeelaangeleenthede, museums, kunsmuseums en biblioteke* (Prolingua 2015: 2; Alberts 2015a).

Die vakvereniging se hoofdoel was om terme te voorsien, te standaardiseer en vertaalprobleme op te los. Die Vereniging het gou deur die land bekend geword en as gevolg van die belangstelling deur nasionale instansies en munisipaliteite in ander provinsies as Transvaal, is die naam verander na die *Vereniging van Munisipale Vertalers* en die lede is toegelaat om vergaderings in kantoor tyd by te woon. 'n Verteenwoordiger van die Vereniging het ook op die Koördinerende Vaktaakraad (KOVAK) gedien (Alberts 2015a; Alberts 2017: 165-167).

Later is meer klem op terminologiewerk gelê en die Vereniging se naam is verander na die *Terminologievereniging van Suid-Afrika* (Termisa). Die naam is in die negentigerjare verander na *Prolingua* (*pro*: 'n voorvoegsel uit Latyn wat 'voor', 'ten gunste van', 'ondersteun' beteken; *lingua*: uit Latyn wat 'tong' of 'taal' beteken) (Prolingua 2015: 2; Alberts 2015a).

Prolingua is 'n vakvereniging vir taalpraktisyns waar werksvergaderings maandeliks gehou word om terminologieprobleme te bespreek en nuwe terme te skep. Dit is hierdie taalpraktisyns se passie om nuwe terme te skep in 'n veranderende en opwindende globale wêreld en om meer van hul profesie te wete te kom. Hoewel Prolingua tweetalig is, is daar steeds 'n groter behoefte aan die skep van Afrikaanse terme (Prolingua 2015: 2).

Daar word jaarliks tien werksvergaderings gehou wat om die beurt by die kantore van die Tshwane Munisipaliteit in Centurion en by Unisa se Florida-kampus, Roodepoort gehou word. Tydens die werksvergaderings word terminologieprobleme bespreek en die Prolingua-termlys word hersien en bygewerk. Prolingua gee gereeld erkenning aan lede (en nielede) wat uitmuntende

diens op taalgebied lewer en toekennings word dan gegee of erelidmaatskap word aan dié lede toegeken.

Hoewel daar tans akademiese kursusse is vir taalpraktisyns om hulle as vertalers, tolke, redigeerders en terminoloë te bekwaam, werk taalpraktisyns dikwels alleen en het 'n behoefte aan kontak met ander taalwerkers. Buiten die gereelde werksvergaderings is daar ook 'n aktiewe groep wat termprobleme per e-pos bespreek. Prolingua het ook 'n Yahoo-gespreksgroep (Prolingua99@yahoo.com) waar lede mekaar met die vertaling van dringende terme help.

Gassprekers word na die maandelikse werksvergaderings genooi om aspekte van terminologie te behandel (Alberts 2012; Bosman 2015; Jonker 2013; Burger 2016; Van Nierop 2016; Russouw 2017; Van Huyssteen 2017) of verskeie ander taalverwante sake te bespreek, soos tersaaklike wetgewing (Beukes 2014; Malan 2014; Le Roux 2016), taalaspekte (Posthumus 2016; Van Rooy 2017); vertaal- en redigeerkwessies, eenvoudige en korrekte taalgebruik en kompakte skryfwyse (Cornelius 2012; Geldenhuys, Dykman en Viljoen-Smook 2012; Dykman 2015; Viljoen-Smook, Geldenhuys en Coetzee 2017), taalverwante programmatuur (Fourie 2014; Van Huyssteen 2014), forensiese linguistiek (Kotzé 2015), teksredigering (Carstens 2011, 2014; Linnegar 2013, 2014), tekstegnologie (Prinsloo 2011; Roux 2016), die kuberruim en sosiale media (Maree 2014; Von Solms 2013); klassieke vertaling (Vos 2015; Human 2016); kreatiewe skryfkuns (Du Plessis 2016); Afrikaanse spel- en skryfreëls (McLachlan 2011, 2012; Van Huyssteen en Kapp 2017), ens. Soms word werkwinkels of kort kursusse aangebied (Combrink 2015; Dykman 2012, 2013, 2015).

Hoewel Prolingua hoofsaaklik 'n vakvereniging vir Engelse en Afrikaanse taalpraktisyns is, woon taalpraktisyns wat in die ander amptelike tale werk ook die werksvergaderings by. Hulle word dan ook gereeld op die hoogte gehou van publikasies wat vir die Afrikatale van waarde kan wees, soos die onlangse publikasie oor teksredigering in Sesotho (Malete 2016) wat op *Teksredaksie* (Carstens en Van de Poel 2010) en *Text Editing: A Handbook for Students and Practitioners* (Van de Poel, Carstens en Linnegar 2012) gebaseer is. Hierdie Sesotho-weergawe is 'n Suid-Afrikaanse publikasie wat as pionier beskou kan word op die gebied van die standaardisering van 'n inheemse taal.

Prolingua (en sy voorgangers) het met die verloop van etlike jare verskeie termlyste en vakwoordeboeke in 'n Prolingua-termlys versamel. Nuwe terme uit verskillende vakgebiede word steeds in hierdie lys opgeneem. Die lys word gereeld bygewerk en hersien en is 'n waardevolle hulpmiddel vir taalpraktisyns landwyd. Daar word soms ook termlyste saamgestel wat gratis op die webwerf (<http://www.prolingua.org.za>) beskikbaar gestel word (Prolingua 2015: 1).

Verskeie Prolingua-lede is ook medeouteurs van vakwoordeboeke: *Business Dictionary/Sakewoordeboek* (Geldenhuys en Viljoen-Smook 2009), *Modern Political Dictionary/Nuwerwetse Politieke Woordeboek* (Botha, Le Clus en Venter 2011) en *Legal Terminology: Criminal Law, Procedure and Evidence/Regsterminologie: Straf-, Strafproses- en Bewysreg* (CLTAL 2015).

Verskeie lede en/of gassprekers het ook met verloop van jare aan die lede inligting verskaf oor wetgewing en wetsontwerpe wat op die taalberoep van toepassing is, soos die *Nasionale Taalbeleidsraamwerk* (Finale Konsep 2002); die *Nasionale Taalbeleidsraamwerk en Implemeteringsplan* (DKK 2003); die *Taalbeleidsraamwerk van die Gautengse Provinsiale Regering* (2005); die *Kopieregwet* (1978); die *Kopieregwysigingswet* (2002); die *Wet op die Gebruik van Amptelike Tale* (2012); die *Suid-Afrikaanse Taalpraktisynsraad* (2014); en die wysiging van hofbenamings (Le Roux 2016).

Tersaaklike inligting word ook aan lede verskaf deur 'n rubriek *Taalwys* wat per e-pos versprei en in die notules van die werksvergaderings opgeneem word. Lede word ook gereeld ingelig oor publikasies soos die *Landdroshowe se boek met reëls in Afrikaans* (VRA 2017) wat hul werk kan vergemaklik. Sulke publikasies is belangrik vir die regs- en taalprofessies omdat dit die gebruik van Afrikaans in die land se howe bevorder en belangrike regsterme regstegnies binne konteks (regsprosesse, -prosedures en wetlike betekenis) beskikbaar stel.

Verskeie gassprekers het tydens werksvergaderings inligting verskaf oor programmatuur wat gebruik word om terminologie te bestuur (Fourie 2014, Van Huyssteen 2014). Werkwinkels is ook deur lede aangebied oor programmatuur wat taalpraktisyns kan help om hul eie termlyste saam te stel en by te werk (Liebenberg, Delpont en Greenfield 2016).

3. Die Prolingua-termlyste en ander produkte

'n Ondersoek wat in 2015 in opdrag (SAAWK 2015a) van die Dagbreek Trust en die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) afgehandel is oor die stand en status van Suid-Afrikaanse vaktaal (Alberts 2015a) wys daarop dat Afrikaanse vaktaal gedurende die vorige bedeling (en veral vanaf 1950 tot 1994) 'n bloeitydperk beleef het (Alberts 2000a en b; Alberts 2015a en b). Die huidige Suid-Afrikaanse Grondwet (1996) vereis by implikasie dat aandag gegee word aan die ontwikkeling van meertalige vaktaal. Die vaktale van die amptelike tale moet in die toekoms saam ontwikkel word — nie net vir eie behoud en ontwikkeling nie, maar veral weens vakgerigte kommunikasie-behoefte.

Tans is die meeste tweetalige vakwoordeboeke wat in die verlede deur (vak)taalkantore en individue saamgestel is, uit druk. Daar is geen sin daarin om hierdie ou tweetalige vakwoordeboeke in hardebandformaat te herproduseer nie aangesien die klem deesdae op die samestelling van meertalige termlyste en aanlyn terminologiese produkte val. Vaktaalbronne wat wel beskikbaar is, soos termlyste, vakwoordeboeke en termdatabasisse, is by verskillende staats- en semistaatsdepartemente, instansies en in biblioteke of êrens by private persone. Hierdie fragmentering maak dit uiters moeilik om beskikbare vaktaalbronne op te spoor en potensiële vaktalgebruikers is dikwels onbewus dat bronne wat hulle nodig het, weliswaar bestaan (Luther 2005).

As gevolg van die beperkinge wat die huidige talesituasie op veral Afrikaanse vaktaalontwikkeling plaas, is dit ook heel onwaarskynlik dat hierdie vakwoordeboeke ooit herdruk sal word. Aangesien baie belastingbetalergeld, vak- en taalkundige insigte en werk in die samestelling van hierdie vakwoordeboeke gegaan het, sal dit 'n onreg wees as hierdie vaktaalinligting verlore sou gaan. Die vaktaalinligting is steeds geldig. Die terme is grootliks reeds gestandaardiseer en baie min daarvan kan as verouderd beskou word. Die terminologieskat is steeds bruikbaar vir vakspesialiste, joernaliste, leerders, onderwysers en studente en taalpraktisyne soos vertalers, tolke, terminoloë, leksikograwe, taaldosente, onderwysers, ens. Taalpraktisyne wat hoofsaaklik in die Afrikatale werk, maak ook van Engels-Afrikaanse vakwoordeboeke gebruik in termskepping. Hulle kan konseptuele inligting soms makliker uit die Afrikaanse definisies ontsluit en skep terme na aanleiding van die Afrikaanse termekwivalente eerder as die Engelse terme (Alberts 2015a). Hoewel die Prolingua-termlys dus slegs tweetalig is, is dit ook van waarde vir die gebruikers van die ander amptelike tale.

Die gedigitiseerde Prolingua-termlys dateer uit 1984. Van die terme wat vóór 1984 deur Prolingua se voorgangers gedokumenteer is, is ook in die Prolingua-termlys opgeneem. Sommige terme wat in hierdie termlys opgeneem is, is nêrens in vakwoordeboeke opgeteken nie en die lys is dus 'n waardevolle terminologiese bron (Prolingua 2015: 3).

Mnr. Pieter Taljaard, voorheen verbonde aan die Poskantoortaal diens, het na sy aftrede in sy hoedanigheid as voorsitter van Prolingua verskillende instansies besoek, o.a. die Nasionale Terminologiese Diens, Departement van Kuns, Kultuur, Wetenskap en Tegnologie (tans die Terminologiese Koordineringsafdeling, Nasionale Taal diens, Departement van Kuns en Kultuur) en die Suid-Afrikaanse Akademie vir Wetenskap en Kuns om te verneem of Prolingua die data van hul gepubliseerde woordeboeke kon benut deur die data te skandeer en dan in elektroniese formaat aan vertalers en vaktaalgebruikers beskikbaar te stel. Prolingua het die nodige toestemming om die data te benut van die meeste uitgewers, vakwoordeboeksamstellers en vaktaalkantore bekom en die outeursreg is opgehef. Waar moontlik is die inligting elektronies aan Prolingua verskaf (Alberts 2015a).

Die samestelling van die Prolingua-termlys is 'n projek sonder winsbejag en alle wins uit die aanvanklike verkope van die termlys is vir die verbetering van die produk aangewend. Die Prolingua-termlys was aanvanklik slegs vir lede toeganklik en hersiene, bygewerkte termlyste is gereeld elektronies aan lede gestuur en drukstukke kon ook op aanvraag gemaak en verkoop word.

Dit blyk duidelik uit voorgaande dat Suid-Afrikaanse terminologie grotendeels met behulp van belastingbetalergeld ontwikkel, gedokumenteer en gedissemineer is (Alberts 2015a). Die hoofbron vir die Prolingua-termlys was ou vakwoordeboeke en termlyste wat deur staatsinstansies saamgestel is en goedgunstig aan Prolingua beskikbaar gestel is. Daar was dus eintlik geen rede waarom die terme nie gratis aan alle (vak)taalgebruikers beskikbaar gestel kon

word nie. Die belastingbetalers, dus (potensiële) vaktaalgebruikers, het immers reeds daarvoor betaal. Die terminologie in die Prolingua-termlys kan vir ander taalpraktisyns soos vertalers, redigeerders, terminoloë en tolke en vakspesialiste van nut wees en is tans gratis op Prolingua se webwerf <http://www.prolingua.org.za> beskikbaar. Tans word nuwe terme uit verskillende vakgebiede in die Prolingua-termlys opgeneem en gereeld deur die lede hersien of bygewerk.

Die volgende doelwitte van Prolingua rig sy aktiwiteite rondom terminologieontwikkeling:

- om terme sover moontlik te standaardiseer;
- om die gehalte van vertaalwerk te verbeter; en
- om lede die geleentheid te gee om hul vertaal- en terminologieprobleme te bespreek, met die Prolingua-termlys as die uiteindelige uitkoms van hierdie besprekings (Prolingua 2015: 3).

In die afgelope 67 jaar het Prolingua buiten die Prolingua-termlys ook vele ander termlyste saamgestel. Termlyste is saamgestel vir *elektrisiteit; stadsbeplanning* asook *pad en vervoer; name van organisasies; name van kwalifikasies; name van wette en ander terme*. Prolingua-lede het ook aan die saamstel, hersiening en bywerking van etlike vakwoordeboeke meegewerk. Hierdie vakwoordeboeke sluit in: *Brandweerwoordeboek; Elektrotegniese en Elektroniese woordeboek; Kernbedryfwoordeboek; Lewensversekeringswoordeboek; Opmeetwoordeboek; Plastiekwoordeboek; Poskantoorwoordeboek; Stadsbeplanningswoordeboek; Terminologie van vergaderingsprosedure en redevoering; Tweetalige Voorsetselwoordeboek; Water- en rioolwerketerme; Insekname- en Skoenlapperlys*. Heelwat van dié vakwoordeboeke is uit druk en daar is toe begin om hulle elektronies in te lees en gevolglik is heelwat van hierdie vakwoordeboeke steeds by Prolingua in elektroniese formaat beskikbaar (Prolingua 2015: 3).

Prolingua tree ook proaktief op en 'n lys van ongeveer 300 *Gautreinterme* is op die Prolingua-webwerf beskikbaar. Een van die Prolingua-lede, Johan Dorfling, het 'n unieke *Brugwoordelys* saamgestel waarvoor hy self Afrikaanse brugterme geskep het. Hy het 'n behoefte aan Afrikaanse terme vir die brugspel geïdentifiseer en die termlys gevolglik geïnisieer, saamgestel en bygewerk (Prolingua 2015: 3; Dorfling 2002–2011).

Al hierdie vakgebiede waarby Prolingua betrokke is (hetsy deur die bewaring of skep van terme), toon dat elke vakgebied, hoe gespesialiseerd of eenvoudig ook al, oor 'n unieke stel terme beskik wat bepaalde begrippe uniek benoem en belangrik is vir vakkommunikasie. Terminologieontwikkeling is belangrik vir elke vakgebied en staan voorop in die ontwikkeling van enige taal tot funksionele gebruiks- en wetenskapstaal (Alberts 2003: 128; Alberts 2015a).

4. Terminologieontwikkeling deur Prolingua

Terminologie word nie net deur vakkundiges, vertalers, tolke, joernaliste, teg-

niese skrywers en woordeboekmakers benodig nie, maar ook deur onderwysers, leerders, dosente, studente, staatsamptenare, sakelui, medici, regslui, skoenmakers, kuipers en alle ander beroepslui. Vaktaalgebruikers het nie net terme, termekwivalente en definisies of verklarings nodig nie, maar ook gebruiksvoorbeelde wat die gebruik van 'n term in verskillende registers aantoon. Hoewel terminoloë poog om konsepte eenduidig met slegs een term/termekwivalent te benoem (die beginsel van *een konsep een term*), is dit tog soms nodig om sinonieme te verskaf (byvoorbeeld by Afrikaanse terme die Germaanse vorm naas 'n getranslitereerde vorm). Taalpraktisyns benodig soms sinonieme vir 'n bepaalde konteks asook 'n aanduiding van die registers waarin 'n bepaalde term of sy sinonieme gebruik sou kon word. Hulle benodig uiteraard ook die heel jongste terminologie.

Wanneer doeltreffende kommunikasie op en oor 'n spesifieke spesialisveld onmoontlik raak, omdat die taalgebruikers nie oor die nodige terme in die moedertaal beskik nie of nie weet waar om die regte terme te vind nie, is daar geen keuse nie as om na Engels oor te slaan (Luther 2005). Almal kla oor die hegemonie van Engels, maar min pogings word werklik gekoördineerd aangewend om die vaktale van die verskillende vakterreine in die amptelike tale te ontwikkel. Die lede van Prolingua poog om ten minste aandag te gee aan die ontwikkeling van Afrikaanse terme en hierdie terme is beskikbaar en bruikbaar en toeganklik vir gebruik deur enige ander taalgroep wat termekwivalente na aanleiding van die Afrikaanse terme wil skeep.

4.1 Prolingua-lys: hersiening en bywerking

Die Prolingua-termlys is 'n tweetalige alfabetiese lys met Engels as brontaal en Afrikaans as doeltaal. Geen verklarings van terme word voorsien nie, maar verbandswoorde gee wel waar nodig 'n aanduiding van vakgebied of kontekstuele gebruik.

Die Prolingua-lede hersien die Prolingua-lys gereeld tydens werksvergaderings. Die lysie terme wat tydens 'n bepaalde werksvergadering vir hersiening aan bod kom, word vooraf per e-pos aan lede gestuur sodat lede wat nie die vergadering kan bywoon nie ook die geleentheid te baat kan neem om insette te lewer. Na afloop van die werksvergadering word die databasis deur Fritz Wolff, 'n voormalige voorsitter van Prolingua, bygewerk.

Enkele voorbeelde van terme wat aan bod gestel word vir hersiening tydens 'n werksvergadering (Prolingua (IvB) 2017b):

bricks and mortar industry {physical industry}: stene-en-daghabedryf
brickyard (brick works, brickfield): steenmakery
bright v. {vehicle headlights}: helder
brilliant cut diamond: briljant
broadbanding {staff management}: bandverbreding
broad-based {economics and social work}: breëbasis-
broad-based black economic empowerment (BBBEE): breëbasis swart
ekonomiese bemagtiging, BBSEB

broadside {advertising}: voubiljet
brokerage (broking) {eg insurance}: makelaarsbedryf
brokerage {firm}: makelary, makelaarsmaatskappy
brokerage: kommissie, makelaarsloon
broking (brokerage) {eg insurance}: makelary

Enkele voorbeelde van probleemterme wat deur lede voorgelê word vir bespreking tydens 'n werksvergadering (Prolingua (IvB) 2017a):

picketing (groep mense wat staak deur byeen te kom en doodstil met plakkate staan en protesteer): betooglinievorming (Wet op Arbeidsverhoudinge, 66 van 1995/Labour Relations Act 66 of 1995): Artikel 69 5(a) and (6); Vloeidiagram/Section 69 5(a) and 6.

picketing: betooglinievorming

picket *n.* : betooglinie

picket *v.* : om 'n betooglinie te vorm

Voorstelle vir Afrikaanse termekwivalente: **picketing**: staakwag, brandwaglinie (*Pharos*); protesstanery

penalising (Example sentence: In such cases, a court can grant a more penalising division or maintenance order than it would normally do if it were of the view that one spouse is in control of significantly more trust assets.)

Voorstelle vir Afrikaanse termekwivalente: bestraffende, strafbare, penalisierende, straffende (punitive = bestraffende (CLTAL 2015))

lockout/tagout {beroepsveiligheid en -gesondheid} 'Lockout' is wanneer werkers fisies 'n kragkakelaar sluit sodat dit nie aangeskakel kan word wanneer iemand daaraan werk of aan iets wat daaraan gekoppel is nie en 'tagout' is wanneer daar 'n etiket aan die geslote ding geheg word.

lock-out switch: afsluitskakelaar (*Pharos*)

Voorstelle vir Afrikaanse termekwivalente: afsluitingsetiket, afsluitingskaartjie, afsluitingsmerk

master account billing

Voorstelle vir Afrikaanse termekwivalente: meesterrekeningfakturering, meesterrekeningdebitering

Die volgende terme is tydens werksvergaderings bygevoeg, gewysig of geskrap (Prolingua 2016; 2017a):

board *v.*: medies ongeskik verklaar aftree, board [~~someone is boarded/ iemand medies laat aftree~~]

bodies of knowledge: kenniskorpuseorpora, kennisinhoude [Onderwys-terme][Mrt12]

body {eg oil}: lywigheid, viskositeit
body {general}: romp
body [no body/te dun]:—
body language: liggaamstaal, lyftaal
body of knowledge: kenniscorpus, kennisinhoud
[Onderwysterme][Mrt12]
body surfing: branderry
body surfing: branderry, lyfbranderry
boggle [it boggles the mind, the mind boggles at/dit slaan my dronk]:—
bogie {railway truck}: bogie, draaistel
boil off (degum): afkook
bone mass density scan: beenmassadigtheidsondersoek,
beenmassadigtheidskandering
boob tube: buisbostuk, buistoppie, spantoppie
booby prize: fopprys, spotprys
boogie board: lyfplank
boomtown: snel groeiende stad, snel groeiende dorp [Feb2006]
bottom line: ~~ook~~ netto wins
brain drain: kundigheidsverlies, breindrein
brain writing: skryfskrum breinskrif
brainstorm *v.*: dinkskrum [colloquially — harsingsgalop, kopgalop]
brainstorming session: dinkskrum, ideestorm, ideeberaad
brand manager: produkmerkbestuurder [bankwese]
branded bottled water: gebottelde handelsmerkwat
branding: naamreklame, beeldbemarking, handelsmerkvestiging [Mei09],
handelsmerkgewing
breeding camp {for game}: teelkamp [~~hupskamp~~ slegs in nietegniese,
laeregistergebruik]
breëkoparend: martial eagle [Feb2006]

4.2 Termskepping

Prolingua se gebruik om gassprekers by geleentheid na werksvergaderings (of soms na die algemene jaarvergadering) uit te nooi om Prolingua-lede oor verskeie terminologiekwessies toe te spreek, dra daartoe by dat die lede insae kry in verskillende terminologieverwante sake soos termskepping, standaardisering en harmonisering.

Van die aspekte wat in die onlangse verlede veral onder die loep gekom het, is terme van klassieke herkoms, neologismes en die waarde van die standaardisering en harmonisering van terminologie.

4.2.1 Terme van klassieke herkoms

Die invloed van die klassieke tale op terminologieontwikkeling is deur verskeie gassprekers beklemtoon.

Die vertaling van die Bybelse frase '**klinkende metaal**' (1 Kor 13:1) is 'n voorbeeld waar die Griekse en Latynse herkoms van 'n terminologiese frase nagespoor word en veroorsaak dat die terminologiese frase dan later tydens die hersiening van tekste (byvoorbeeld Bybeltekste) gewysig word na byvoorbeeld '**galmende ghong**' (Proefvertaling van 2005) en later (in beplande 2018-vertaling) as '**weergalmende ghong**' (Human 2016). Die terminologiese frase verwys moontlik na akoestiese vase wat deur die Romeine in teaters gebruik is om as klankversterkers vir akteurs en sangers te dien — hierdie klankversterkers is selfs gebruik om donderweer te verklank en sou dus 'weergalmers' genoem kon word. Die Griekse woorde *chalkos echōn* beteken letterlik koper/geelkoper/brons- + *galmend/weergalmend/wat eggo*. Die Afrikaanse term '**ghong**' is veral bekend as die musiekinstrument waarmee etenstye in treine aangekondig word.

Dikwels is woorde of terme alreeds sodanig in 'n taal of tale ingebed dat dit soms moeilik is om te bepaal of hulle uit Grieks of Latyn afgelei is (Bosman 2015), byvoorbeeld:

Eng **catastrophe** / Afr **katastrofe** > **kata** (onder) + **strophe** (draai) = draai daar onder

Eng **atmosphere** / Afr **atmosfeer** > **atmos** (damp) + **sphaino** (bal)

Eng **manipulate** / Afr **manipuleer** > **manus** (hand) = speel (iemand) na my hand

Eng **politics** / Afr **politiek** > **polis** = gemeenskap/dorp/stad

Voorbeelde van terme wat klaarblyklik uit Grieks ontleen is (Bosman 2015), is die volgende:

Eng **metaphore** / Afr **metafoor** > Grieks: **meta** (prefiks: langs of onder) + **pherein/phore** (dra) = dra langs/in die plek van iets anders

Eng **synonym** / Afr **sinoniem** > Grieks: **syn** (prefiks: saam) + **onoma** (naam) = "saam naam", dieselfde naam, die een is soos die ander

Eng **syntax** / Afr **sintaksis** > Grieks **syn** (prefiks: saam) + **tasso** (organiseer) = (woorde) saam rangskik (in 'n sin)

Voorbeelde van terme wat klaarblyklik uit Latyn ontleen is (Bosman 2015), is die volgende:

Eng **suffix** / Afr **suffiks** > Latyn: **sub** (prefiks: onder, end) + **figo** (bind) = "bind by einde"

Eng **suspicion** / Afr **suspisie** > Latyn **sub** (prefiks: onder, end) + **spectare** (kyk) = kyk daar onder (waar ek iets verwag)

Dit is egter veral in die mediese en regsberoep waar terme wat uit die klassieke tale ontleen is, voorkom. Volgens Bosman (2015) is die mediese terme van die 21ste eeu gevestig in liggaamsbelewinge en waarnemings van mense wat 2000 tot 3000 jaar gelede gelewe het. So ervaar mense vandag soortgelyke liggaamlike gebreke en gesondheidstoestande as 2000 plus jaar gelede. Die mense

van die toekoms gaan steeds afhanklik wees van dieselfde geestelike gawe van kommunikasie (taal) as vandag en 2000 jaar gelede. Die moderne mediese terme huisves dus 'n verskeidenheid aspekte tiperend van hoe die mens 2000 jaar gelede die lewe, gesondheid en siektes ervaar het, hoe hulle gedink het, hulle kultuur en tradisies beleef het, hul logiese en filosofiese denke en hulle mitiese of godsdienstige ingesteldheid tot die lewe. Sonder inagneming van bogenoemde aspekte is dit nie altyd moontlik om moderne mediese terme se oorsprong en betekenis te verstaan nie. Voorbeelde van terminologie in die mediese beroep wat duidelik bogenoemde redenasie verteenwoordig is:

melancholia (Afr: melankolie) (letterlik: 'toestand van swart gal'; depressie, swartgalligheid). Konteks: swart = boos; gal hou verband met die tradisionele siening dat die gees/verstand in die maag gesetel is en word dus met die maag se kernfunksie (gal) verbind.

maag word ook met **diafragma (phren)** (spier wat dwars sluit) verbind.

phren: 1 diafragma 2. gees, verstand

hypochondria (letterlik: onder kraakbeen (bors); maag; diafragma — waar die gees/verstand is): siektetoestand van die gees/kop/verbeelding/verstand. Afr: hipokondria; volksetimologie: iepekonders

phrenoplegia (Afr frenoplegie): 1. Verlamming van die diafragma, diafragmaparalise 2. Aanval van geestesiekte

schizophrenia (Afr: skisofrenie): **shiz** = oopkloof/gesplete + **phren** = gees, verstand (letterlik: gesplete gees/verstand)

Mediese terme word gewoonlik saamgestel met behulp van die klassieke stamvorm plus prefikse en suffikse. Voorbeelde hiervan is die volgende (vgl. Bosman 2015):

facioplasty (**facio**: gesig + **plast**: herstruktureer, verbetering): plastiese chirurgie van die gesig; soortgelyk: stomatoplasty (mond-); rhinoplasty (neus-); cheiroplasty (hand-)

sigmoid colon (**sigma** = Griekse letter S, dus sigm(a) (S) + **oid** = soos): laaste deel van die kolon voor die rektum wat s-vormig is.

thyroid gland (Grieks: **thyre(os)** = skild + **iod** = soos): skildklier

Dit is egter nie net op die mediese terrein waar terme van klassieke herkoms benut word nie. Die regsprofessie wemel van Latynse terme en frases. Hierdie terme en frases is dikwels ontoeganklik vir gebruikers. Latynse terme wat onveranderd in Engels en Afrikaanse regstaal ingebed is, word kortliks omskryf ter wille van gebruikers wat nie Latyn magtig is of oor die nodige regs kennis beskik nie (Alberts 2012; CLTAL 2015):

Eng: **prima facie** (on the first impression) Afr: **prima facie** (op die oog af)
Eng: **pro forma** (as a matter of form) Afr: **pro forma** (ter wille van die vorm)

Sommige Latynse regsterme word onveranderd in Engels en Afrikaans gebruik, maar van sinonieme voorsien (Alberts 2012), byvoorbeeld:

Eng: **dolus (intention)** Afr **dolus (opset)** Noord-Sotho **dolus (maikemisetso)**

Voorbeelde van Latynse frases wat in die regte gebruik word (Alberts 2012):

sine die [Latin, without a day being fixed] : **sine die** [Latyn, sonder dat 'n dag vasgestel is]

situs n location or position of something for legal purposes, e.g. the place where a crime or accident occurred [Latin, location] : **situs n** ligging of positionering van iets vir regsdoeleindes, bv. die plek waar 'n misdaad of ongeluk gebeur het [Latyn, ligging]

4.2.2 Hersiening of wysiging van terme

Regstaal is daarvoor bekend dat dit 'n magtige instrument is wat aangewend kan word om in te sluit of uit te sluit. Regsjargon is eiesoortig en veroorsaak groot begripsprobleme vir verskillende doelgebruikers — nie net vir regsgeleerdes nie, maar ook vir leke. Volgens huidige wetgewing het almal die reg op inligting in gewone en verstaanbare taal en dit vereis verder dat inligting aan gebruikers in gewone taal beskikbaar gestel moet word (Cornelius 2012). Voorbeelde van terme waar inligtingsdigtheid verstaan- en leesbaarheid beïnvloed, is

korttermynversekeringspremies	eerder: premies vir korttermynversekering
aankoopkaartbeskermingsfonds	eerder: kaartbeskermingsfonds
kredietopnemerlidmaatskap	eerder: lidmaatskap
herskikkingsooreenkoms	eerder: ooreenkoms om skuld af te betaal

Baie Prolingua-lede werk in staats- en semistaatsdienskantore en moet gevolglik vertrouwd wees met die jongste wetgewing. Verskeie benamings van die land se howe is gewysig en die jongste benamings is vervat in die Wet op Hoër Howe, 2013, Wet 10 van 2013. Voorbeelde van die belangrike howe se benamings is (Le Roux 2016):

Constitutional Court / Konstitusionele Hof
Supreme Court of Appeal / Hoogste Hof van Appèl
High Court of South Africa / Hooggeregshof van Suid-Afrika
Superior Court / Hoër Hof
Circuit Courts / Rondgaande howe
Magistrates' Court / Landdroshof

Die benaming van die land se howe lewer dikwels vir gebruikers probleme op

omdat dié terme dikwels gewysig word sonder dat die meeste mense van die wysigings bewus is en soms selfs foutief in vakwoordeboeke opgeneem word (CLTAL 2015), vgl. in hierdie verband **hooggeregshof** (van 1909) wat gewysig is na **hoë hof** (1997) en sedert 2013 weer bekend is as **hooggeregshof** (SRTAT-databasis 2017).

4.2.3 Neologismes

Prolingua se doelstellings is om terme te voorsien, te standaardiseer en vertaalprobleme op te los. Voorbeelde van Afrikaanse neologismes in die Prolingua-termlys (Prolingua 2015) is:

abdominal crunch {working out}: **buikknak**
backseat driver: **bekdrywer**
booby prize: **fopprys**
coaster {for glass}: **drupmatjie**
comedy of errors: **flaterklug**
cutting-edge technology: **voorpuntegnologie**
cyber harassment : **kuberteistering**
guesstimate: **raairaming**
mouse pad {computer}: **muismat**
pop-up restaurant: **opskietrestaurant**
trending {social media}: **twietgons**
wrap: **koskarossie**
upsized v. {on menu}: **rek 'n porsie** [om porsie van dieselfde item te vergroot]; **byvoeg** [nog iets by die gereg byvoeg, byvoorbeeld uieringe]

Prolingua se gassprekers het egter ook grootliks daartoe bygedra om neologismes onder die aandag van lede te bring. Hierdie nuutskeppings brei nie net die Afrikaanse terminologieskat uit nie, maar hulle bevorder die Afrikaanse taal op 'n kleurrike wyse. Enkele voorbeelde van nuutskeppings in verskillende vakgebiede wat deur gassprekers en medewerkers voorsien is, is die volgende:

Brugterme: Daar is 'n toenemende belangstelling onder Afrikaanssprekendes om brug te speel. Veral tuisspelers en etenstydspelers het die afgelope aantal jare talryker geword en aldus het die behoefte aan 'n saamgestelde termlys al hoe sterker geword. Daar bestaan wel goeie Afrikaanse terminologie, maar betreklik min brugspelers is op die hoogte daarvan omdat dit nêrens versamel is of maklik nageslaan kan word nie. Een van die Prolingua-lede, Johan Dorfling, het 'n *Brugwoordelys* saamgestel waarvoor hy self Afrikaanse termekwivalente geskep het. Hy het die proses om die brugtermlys saam te stel self geïnisieer en die lys verskeie kere hersien en bygewerk. Die oogmerk van die samesteller van die brugtermlys is om die groot leemte aan 'n eenvormige en volledige Afrikaanse woordeskat vir brugspelers aan te vul (Dorfling 2002; 2012).

In die soeke na meer volledige en bruikbare termekwivalente vir Afrikaans is daar onder meer by Vlaams, Nederlands en Duits kers opgesteek om Afrikaans sodoende te verryk en die Germaanse inslag van Afrikaans te behou. Versigtigheid is egter aan die dag gelê om nie terme op te neem wat met die taaleie bots nie. Sommige brugterme wat in Nederlands gebruik word, het nie inslag in Afrikaans gevind nie. Dié terme, wat weliswaar in ander Afrikaanse woordeboeke verskyn, is aanvanklik in die brugtermlys opgeneem, maar in latere weergawes weggelaat omdat dit enersyds nie deur Afrikaanssprekendes gebruik word nie en andersyds nie in die Afrikaanse idioom pas nie. Een voorbeeld is die benaming **blindeman** wat teenoor die **troefmaker** sit. Hierdie speler is in geen opsig "blind" nie, nie letterlik of figuurlik nie. Die speler beweeg in die skadu van die troefmaker en die benaming **skaduspeler** is meer gepas en word in die algemeen so gebruik.

Regstreekse vertaling vanuit Engels is dikwels nie aanvaarbaar of duidelik nie. So beteken "cancelled call" byvoorbeeld volgens die Engelse spelreëls van brug dat 'n **roep ongeldig** is en nie dat dit 'ingetrek' is nie. Die benaming "international match point" is ook verwarrend omdat dit nie 'n werklike beskrywing van die telwyse is nie en 'n direkte vertaling daarvan te lank is om inslag te vind. Die term "international match point" word meermale ook op plaaslike vlak toegepas en is in werklikheid 'n **glytelling**. **Glytelling** is dan ook die term wat by Afrikaanssprekende brugspelers byval gevind het.

Johan Dorfling is in 2014 deur Prolingua by die algemene jaarvergadering vir hierdie waardevolle bydrae tot die Afrikaanse terminologieskat met 'n spesiale toekenning vereer. Die *Brugwoordelys* is op die Prolingua-webwerf (www.prolingua.org.za) beskikbaar.

Regsterme: Die werksaamhede van die Sentrum vir Regsterminologie in die Afrikatale (SRTAT) is by 'n Prolingua-werksvergadering bespreek (Alberts 2012). Die projek het ontstaan toe vertalers en tolke van die destydse selfregerende state in Suid-Afrika by die algemene jaarvergadering van die Suid-Afrikaanse Vertalersinstituut (SAVI) in 1985 aangedui het dat hulle terminologieprobleme ervaar wanneer regsdokumente vertaal word, wetgewing vir die verskillende state opgestel word en daar in die hof getolk moet word aangesien regsterminologie in die Afrikatale óf nie bestaan nie óf ontoereikend is. Daar is eerste met Noord-Sotho begin omdat taalpraktisyns wat in dié taal werksaam is die probleem die eerste aan bod gestel het. Die volgende Noord-Sotho of Afrikaanse nuutskeppings is in die SRTAT-databasis opgeneem:

Engels	Noord-Sotho	Afrikaans
baby snatcher	seutswalesea	babadief
bush lawyer	moitiraramoloa	bosprokureur
parental kidnapping	tšhabišokamotswadi	ouerontvoering
premeditated crime	bosenyipeakanywa	voorbegadte misdaad
primary evidence	bohlatsehlatse	beste getuienis
polygraph	seutoliamaaka	poligraaf

Die Sentrum se regswoordeboek *Legal Terminology: Criminal Law, Procedure and Evidence/Regsterminologie: Straf-, Strafproses- en Bewysreg* is in 2015 deur Juta gepubliseer en tans word die finale afronding aan die Noord-Sothoweergawe van die woordeboek gedoen. Die Sentrum het ook reeds begin met die byvoeging van terme en definisies in die ander amptelike Afrikatale.

Rolprentterme: 'n Gasspreker, Leon van Nierop, het die Prolingua-lede tydens 'n werksvergadering op 'n reis deur die rolprentwêreld geneem deur 'n lys Engels-Afrikaanse terme wat op rolprentstel gebruik word vir bespreking voor te lê. Die lede kon daardeur uitvind wat elke persoon se taak gedurende die vervaardiging van 'n film is en wat die verskillende rolprentprosesse behels. Die aanbieding het ook aangedui wat die titellys, wat aan die einde van 'n rolprent rol, beteken. Dit was 'n interaktiewe sessie waar Prolingua-lede insette kon lewer oor die korrekte spelling van woorde (veral los- en vasskryf) en voorstelle vir vertalings kon maak. Enkele voorbeelde van die rolprentterme wat bespreek is, is die volgende (Van Nierop 2016):

Titels	Titles
datahanteerder	data handler, data coordinator
digitale fasiliteerder	digital facilitator
kinematograaf	cinematographer
klankhengelaar	boom swinger
lokprentredigeerder	trailer editor
lynvervaardiger	line producer
rolverdelingleser	cast reader
steadicam-operateur	steadicam operator
takelaar	grip
uitvoerende vervaardiger	executive director

Ander interessante rolprentterme wat bespreek is, is die volgende:

byklanke	foley
dolly (spoortrollie)	dolly
doof in en doof uit	fade in and fade out
filmmateriaal	footage
fokustrekker	focus puller
glyskoot	tracking shot
jibarm	jib
klankkring	looping of sound
naproduksieklankfasiliteit	post production sound facility
vriesraam	freeze frame

Radiotegnologie en radioterme: Dit is nie net taalpraktisyns wat 'n belangrike bydrae tot Afrikaanse terminologieontwikkeling lewer nie. Dit is duidelik bewys deur die aanbieding oor radiotegnologie en radioterme tydens 'n Pro-

lingua-werksvergadering (Burger 2016). Burger het die SA Radioliga se Engelse studiegids geredigeer, herrangskik en bygewerk. Die studiegids is daarna op die liga se webwerf geplaas en in gebruik geneem. Nadat die SA Radioliga besluit het dat hulle eksamens voortaan slegs in Engels afgelê gaan word, aangesien die handleiding in Engels is, het Burger besluit om die gids in Afrikaans te vertaal. Hy het die hoofstukke wat die basiese grondslag dek, vertaal, en vir die meer tegniese hoofstukke slegs termlyste saamgestel. Op grond van hierdie werk van hom, is daar besluit dat die eksamens steeds in Afrikaans afgelê kan word. Van die 25 hoofstukke in die studiegids is 15 op suiwer elektronikateorie toegespits en dus op verskeie vakgebiede van toepassing. Om hierdie rede gebruik skole in Mpumalanga die studiegids as deel van die graad 12-Elektro-nikaleerplan.

Wat die vertaling bemoelik het, is dat die radio- en elektronikawoordeboeke in Afrikaans reeds verouderd is. Dit is 'n veld waarin terme vinnig ontwikkel en verander. Burger het sover as moontlik probeer om Germaanse terme te skep en anglismes te vermy. Hy het byvoorbeeld verkies om die volgende Afrikaanse terme op te neem:

antenna: **lugdraad** (al is dit 'n skottel op 'n gebou se dak)
circuit: **kring**
compression: **saampersing**
processing: **verwerking**
digital: **syfer-**

Die termlys word gereeld bygewerk en dit is gratis beskikbaar by: <http://zs6ez.org.za/download/RadioWoordelys.pdf>.

Sportterme: Die bekende sportverslaggewer, Johann Russouw, het tydens 'n Prolingua-werksvergadering oor Afrikaanse sportterme gesels wat oor die radio gebruik word. Hy het veral op voetbal (sokker), gholf, krieket, rugby, swem en atletiek gekonsentreer wat vol eiesoortige Afrikaanse nuutskeppings is. Interessante sportterme wat tydens die aanbieding aan bod gekom het, is die volgende (Russouw 2017):

Atletiek: **gewigstoot** (shot put)

Gholf:	afslaan (tee shot)	skoonveld (fairway)	sukkelveld (rough)
	sandkuil (bunker)	setperk (putting green)	baansyfer (par)
Tennis:	snyhou (slice)	valhou (drop shot)	mokerhou (smash)
	gelykop (deuce)	vlughou (volley)	lughou (lob)
	trutol (back spin)	stroop (love)	
	skoon stel(le) (straight set(s))		
Krieket:	gang (gully)	dekpunt (cover)	regby (square leg)
	skerpbly (short leg)	vlakby (silly mid-on)	vlak (silly mid-off)

Rugby:	doelpale (upright) plettervat (crash tackle) rolskop (grubber kick)	laagvat (tackle) skrumskakel (scrumhalf) pynpolisie (first-aid staff)	skepskop (drop kick)
Voetbal/:	sweefbal (floater)	doelskut (striker)	neerbring (tackle)
Sokker	soolknoppie (stud) krulskop (curler kick/banana kick)	wawielskop (bicycle kick)	

Lugvaartterme: Die volgende **lugvaartterme** (waarvan sommige in konteks gebruik is), is deur 'n medewerker aan Prolingua beskikbaar gestel (Van Huysteen 2017):

- gasarm** (*throttle*)
- kekkeldeos** (*transponder*).
- lusvlug** (*loop*)
- presisienaderingsbaanaanwyser** (*precision approach path indicator*)
- sigvlugreëls** (*visual flight rules*)
- spruitstukaftapkraan** (*manifold drain cock*)
- tolvlug** (*spin*)
- vaan** (of ook die **steek**) **van die lugskroef** (*pitch of the propeller*)
- vliegveldvluginligtingsdiens** (*aerodrome flight information service*)
- rol, gier en hei:** Vliegtuie beweeg vry op drie asse. Wanneer die vliegtuig op die denkbeeldige as beweeg wat van voor tot agter in die vliegtuig strek, **rol** (*roll*) die vlieënier met behulp van die **rolroer** (*aileron*); op die horisontale vlerk-tot-vlerk-as **hei** (*pitch*) hy met die **hoogteroer** (*elevator*); en op die vertikale as **gier** (*yaw*) hy met die **rigtingsroer** (*rudder*).
- wannadering, wentel en waas:** In 'n kwaai **dwarswind** (*crosswind*) wat eers **krimp** (*back*) en dan **ruim** (*veer*) moet daar 'n paar keer om die lughawe gewentel (*orbited*) word. Ná enkele wannaderings (*missed approaches*) en 'n **deurlooptanding** (*touch-and-go landing*) laat die vlieënier die **eindnaderingsgebooië** (*landing checks*) loop om die vliegtuig uiteindelik met 'n suksesvolle **volle landing** (*full-stop landing*) terug op moeder aarde op die **laaiblad** (*apron*) te kry.

Die terme en nuutskeppings wat tydens aanbiedings bespreek is, word in die Prolingua-termlys opgeneem.

4.2.4 Standaardisering en harmonisering

Die waarde van die terminologiese beginsels 'standaardisering' en 'harmonisering' is tydens 'n Prolingua-aanbieding bespreek (Alberts 2012). Vakspecialiste en terminoloë poog om ondubbelsinnige, bron- en doeltaalterme vir goed gedefinieerde begrippe te verskaf. Om dié doel te bereik, is dit noodsaaklik om die presiese betekenis van terme te bepaal ten einde gebruikers in staat te stel om dit te verstaan en op 'n universeel aanvaarbare wyse te gebruik (Alberts en

Mollema 2013). Begrippe wat behoorlik nagevors, verklaar, met brontaalterme benoem en van vertaalekwivalente in doeltale voorsien is, kan deur gereelde gebruik in die betrokke vakgebied en tersaaklike tale penetreer. Sulke terme word dan deur herhaalde gebruik gestandaardiseer. Standaardisasie is 'n natuurlike proses en nie iets wat op 'n vak- en taalgemeenskap afgedwing kan word nie. Taalliggame kan byvoorbeeld nie hul gesag gebruik om terme te standaardiseer nie, hulle kan wel die terme beoordeel in terme van die woordvormings-, spel- en skryfbeginsels wat eie aan 'n betrokke taal is. As terme nie gedokumenteer en gebruik word nie, kan hulle nie in die vakgebied en taal penetreer en gestandaardiseer word nie. Terminologie wat nie gestandaardiseer is nie veroorsaak kommunikasieprobleme.

Deur die korrekte aanwending van gestandaardiseerde terminologie kan die wetenskaplike, ekonomiese en tegnologiese kommunikasievermoëns van 'n land se inwoners ontwikkel word. Hoewel terminologiese en terminografiese aktiwiteite nie altyd koste-effektief is nie, is dit steeds van historiese, kulturele, sosiale, funksionele, akademiese en wetenskaplike belang om soveel moontlik terminologieë te sistematiseer, te dokumenteer, te standaardiseer en te dissemineer (Alberts 2003: 128).

Die ander belangrike terminologiese beginsel wat ook beprek is, is 'harmonisering' (Alberts 2012). In Suid-Afrika veroorsaak taalkontak dat die taalgebruikers van verskillende tale onderling by mekaar kan leer en leen en selfs, wanneer dit by termskepping kom, soortgelyke termekwivalente vir dieselfde konsep en brontaalterm (kan) skep. Hierdie termekwivalente se stam kom gewoonlik ooreen, maar die spel- en skryfwyse is aangepas by die betrokke taal se stelsel. Geharmoniseerde terme het 'n beter kans om in verwante tale te penetreer en verminder die proliferasie van termekwivalente in verwante tale. Die fokus is hier op die versameling van brontaalterme en doeltaaltermekwivalente en die standaardisering daarvan deur middel van behoorlike definisies eerder as op die skep van nuwe terme in verwante tale. Dit is dus van belang om by die benoeming van nuwe begrippe (veral in die geval van nuutskeppings) die ander taalgemeenskappe te nader om te verneem of die begrip reeds benoem is. Die termekwivalent behoort dan volgens die woordvormings-, spel- en skryfbeginsels van die betrokke taal oorgeneem te word. Voorbeelde van harmonisering in die SRTAT-databasis is:

- bribe** *v* (en): omkoop (af); -reka (ns); -reka (tn)
- trade** *n* (en): handel (af); kgwebo (ns, ss, tn); ukuhweba (zu)
- law** *n* (en): wet (af); molao (ns, ss, tn); umthetho (xh, zu)

5. Samewerking met ander instansies

Prolingua en sy voorgangers het oor die afgelope 67 jaar daarin geslaag om 'n goeie samewerkingsverhouding met verskeie instansies op te bou. Dit is simbiotiese verhoudings met wedersydse voordele vir die vereniging en die betrokke instansies.

5.1 Suid-Afrikaanse Vertalersinstituut

Prolingua het 'n jarelange verbintenis met die Suid-Afrikaanse Vertalersinstituut (SAVI) wat inderwaarheid uit Prolingua se voorgangers voortgespruit het. Verskeie taalpraktisyns is dan ook lede van sowel Prolingua as SAVI en die lede van albei instansies word voortdurend op die hoogte gehou van hul onderskeie bedrywighede (Prolingua 2015: 5). *Murato*, SAVI se vaktydskrif, wat 'n verskeidenheid artikels oor taal en verwante kwessies en inligting oor instuutsake bevat, word ook onder Prolingua-lede versprei.

SAVI is by die Fédération Internationale des Traducteurs (FIT) (Internasionale Federasie van Vertalers) geaffilieer en het in samewerking met FIT dié instelling se veeltalige vertaalterminologielys in Afrikaans vertaal. Die vertaaltermes is in 2010 gepubliseer met die titel *Veeltalige vertaalterminologie/Multilingual Translation Terminology*. Die uitdagings en struikelblokke wat by die vertaling van die FIT-vertaalterminologielys in Afrikaans ondervind is, is tydens 'n Prolingua-werksvergadering met die lede gedeel (Beukes en Pienaar 2011).

5.2 Suid-Afrikaanse Akademie vir Wetenskap en Kuns

Daar is goeie samewerking tussen Prolingua en die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK). Sommige Prolingua-lede is ook Akademielede. In 2010 is Prolingua se volledige boekery aan die SAAWK geskenk. Heelwat Prolingua-lede het ook vakwoordeboeke in hul privaat besit aan die SAAWK geskenk.

Buiten die skenkings wat van Prolingua en sy lede ontvang is, het die woordeboekargief by die SAAWK alreeds talle algemene woordeboeke in verskeie tale beskikbaar, soos Engels, Duits, Russies, Portugees, ens., onder andere in kombinasie met Afrikaans. Die SAAWK het ook 'n aansienlike aantal vakwoordeboeke in sy besit wat oor verskeie jare deur verskillende taalkantore en individue saamgestel is. Die terme in die vakwoordeboeke en termlyste is steeds baie bruikbaar, maar dit is nie altyd moontlik om die vakwoordeboeke of termlyste te bekom nie. Die meeste vakwoordeboeke is reeds uit druk en sommige is nooit hersien nie. Die vakwoordeboeke sluit omtrent alle vakgebiede of navorsingsterreine in (Alberts 2015a). Daar is selfs 'n Deense vakwoordeboek oor elektronika en een oor bouetegnologie — gelukkig in kombinasie met Engels en Duits. Die Deense terme lyk selfs na Afrikaanse skeppings en sou gebruik kon word om, waar nodig, Afrikaanse terme na aanleiding daarvan te skep. Daar is ook terminologielyste in die amptelike Afrikatale, byvoorbeeld die ou terminologie, ortografie- en spelreëls van die Afrikatale, die jongste spel- en skryfreëls wat deur die Nasionale Taalliggame van PanSAT in 2008 saamgestel is en 'n lys met viertalige staatkunde en politieke terme.

Daar is egter geen sin daarin om 'n woordeboekargief te hê wat nie gedeel word nie. Die navorsingsmateriaal, algemene en vakwoordeboeke is gevolglik behoorlik aangewins en 'n vertrek word tans in die nuwe Akademiegebou

ingerig om die bronne toeganklik te maak.

Die *Afrikaanse Woordelys en Spelreëls* (AWS) van die Taalkommissie van die SAAWK word gereeld by Prolingua onder die loep geneem en lede word op die hoogte gehou van nuwe verwickelinge op die terrein van die Afrikaanse spel- en skryfwyse (McLachlan 2011, 2012; Van Huyssteen en Kapp 2017).

5.3 Die Virtuele Instituut vir Afrikaans

Die SAAWK ondersteun die Virtuele Instituut vir Afrikaans (VivA) en gevolglik beskik die SAAWK oor 'n uitstekende infrastruktuur waarmee terminologiewerk gedoen kan word. VivA is 'n gesamentlike inisiatief van die SAAWK, die Dagbreek Trust, die Sentrum vir Tekstegnologie (CTexT[®]) van die Noordwes-Universiteit en die Afrikaanse Taal- en Kultuurvereniging (ATKV). Alle aspekte ten opsigte van die struktuur, grammatika, dokumentering en ontwikkeling van Afrikaans word reeds deur VivA ondervang, vgl. <http://viva-afrikaans.org>. VivA het ook toegang tot die vakwoordeboeke wat deur Prolingua gedigitiseer is.

VivA is op Afrikaans as digitale taal gerig. Die hoofdoel is om toegang tot terme en ander inligting oor en in Afrikaans te skep. Tydens 'n aanbieding by 'n Prolingua-werksvergadering (Van Huyssteen 2014) is voorgestel dat Prolingua by VivA betrokke kon raak deur saam te werk aan die skep van 'n termbank. Die lede het tydens die vergadering besluit om die aanbieding, VivA se sakeplan, asook inligting oor moontlike samewerking aan alle lede te stuur. Daarna is 'n stemming per e-pos gehou sodat alle lede aan die besluitnemingsproses kon deelneem. Die lede het aangedui dat hulle graag by VivA betrokke sou wou wees, dat daar skakels tussen Prolingua en VivA se webwerwe moet wees en dat beide instansies daardeur wedersyds op die hoogte gehou kan word van aktiwiteite en data kan uitruil.

VivA het onlangs 'n Afrikaanse woordeboektoepassing (toep) bekend gestel wat met die ondersteuning van die Atterbury Trust ontwikkel is. Die gratis toep verleen kitstoegang tot sewe Afrikaanse woordeboeke en woorde-lyste (SAAWK 2015b: 7). Dié soort toepassing is beslis ook moontlik vir termlyste en vakwoordeboeke.

5.4 CTexT

Die Sentrum vir Tekstegnologie (CTexT[®]) op die Potchefstroomkampus van die Noordwes-Universiteit het die oopbron *Autshumato Terminology Management System* (TMS) ('n terminologiebestuurstelselprogram) ontwikkel en die VivA-projek gebruik die TMS om terme, gesegdes, vertalings, etimologie en taaladvies vas te lê. Hierdie inligting word onmiddellik beskikbaar gestel deur die webwerf (<http://viva-afrikaans.org>). Die gebruik van VivA se infrastruktuur maak dit moontlik om termbanke aan VivA te lisensieer. Daar word dan onder-

ling bepaal of die terminologiese data gratis of teen betaling aan gebruikers beskikbaar gestel word.

In 2014 is Prolingua-lede aan die hand van 'n aanbieding en demonstrasies ingelig oor die *Autshumato-vertaalprogram* wat ook deur CTeXT ontwikkel is (Fourie 2014). Die Autshumato-vertaalprogram is deur die Departement van Kuns en Kultuur geïnisieer tot voordeel van al die amptelike tale in Suid-Afrika. Die doel daarmee is om vertaalwerk en die standaardisering van terme te vergemaklik.

5.5 Tersiêre instansies

Verskeie lede is verbonde aan die akademiese departemente en taaldienste van tersiêre instansies en hou hulle kollegas op die hoogte van Prolingua-bedrywighede, aanbiedings en publikasies. Kollegas word gevolglik ook na die aktiwiteite wat aangebied word, uitgenooi en selfs ook versoek om praatjies oor taalverwante sake by werksvergaderings, werksessies en die algemene jaarvergaderings aan te bied. Op hul beurt word die lede op die hoogte gehou van aktiwiteite wat by die kampusse van tersiêre instansies aangebied word, soos taaldebate, slypskole, werksessies, kortkursusse en modules oor vertaling, tolking, terminologie, leksikografie en praktiese demonstrasies van rekenaartoepassings en -programmatuur wat die werk van taalpraktisyns kan vergemaklik. Lede word ook dikwels uitgenooi om aan navorsingsprojekte deel te neem, byvoorbeeld aanlyn vraelyste oor een of ander aspek van taal.

Die Departement van Hoër Onderwys het in 2013 fondse bewillig om 'n veeltalige oopbrontermbank vir Hoër Onderwys tot stand te bring. Die oorhoofse doel van hierdie projek is die vestiging van 'n veeltalige termbank met oop toegang binne Hoër Onderwys. Op sy beurt het die termbank ten doel om terminologieontwikkeling tussen tersiêre instansies onderling te koördineer en om die amptelike Afrikatale (Afrikaans ingesluit) te ontwikkel as tale van akademiese diskoers — spesifiek om studente via hulle moedertale toegang te gee tot kernbegrippe in akademiese dissiplines (Alberts 2015a: 57). Die Prolingua-werksaamhede en -termlys sou ook hier tot sinvolle samewerking kan lei.

5.6 Terminologiekoördineringsafdeling

Die nasionale termbank wat deur die nasionale terminologieskantoor, die Terminologiekoördineringsafdeling (Terminology Coordination Section) van die Nasionale Taaldiens, Departement van Kuns en Kultuur bedryf word, stel terminologie aan alle landsburgers deur die Departement se webtuiste beskikbaar. Daar word hoofsaaklik terminologieswerk vir die verskillende staatsdepartemente gedoen en ander vakterreine en domeine kry nie meer soos vantevore aandag nie.

Terminologieontwikkeling dra by tot die ontwikkeling en bevordering

van tale en is dus 'n direkte reaksie op die Grondwet van die Republiek van Suid-Afrika (1996) ten opsigte van die ontwikkeling van die amptelike tale. Meertalige terminologie is noodsaaklik om die tale tot funksionele tale met volle hoëvlakfunksies te laat ontwikkel. Engels word by die nasionale terminologieskantoor as brontaal aangewend en terme word slegs in Engels gedefinieer en van vertaalekwivalente in die ander tien amptelike tale voorsien.

Tans is daar net een terminoloog per taal by die nasionale terminologieskantoor werksaam. Nuutaangestelde terminoloë kry deesdae slegs indiensopleiding ten opsigte van die metodologie en hantering van die Autshumato TMS (terminologiebestuurstelselprogram). Dit blyk dat daar tans geen behoefte is aan opleiding wat verband hou met terminografiese beginsels en praktyk nie. Terminoloë wat egter nie in die teoretiese en praktiese begronding van terminologie as wetenskap en praktyk opgelei is nie, kan nie sinvolle terminologieontwikkeling bewerkstellig nie (Alberts 2015a).

In die verlede, tydens die voormalige tweetalige bestel, is tweetalige terminolyste en vakwoordeboeke deur die voorgangers van die Terminologiekoördineringsafdeling saamgestel. Toe was dit die nasionale terminologieskantoor se hoofdoel om Afrikaanse terminologie te bevorder. Tans word Afrikaans bloot as een van die doeltale beskou en aangesien daar reeds Afrikaanse terme in verskeie vakgebiede gedokumenteer is, word geen daadwerklike poging aangewend om nuwe terme in dié taal te ontwikkel nie. Die terminoloë by die nasionale terminologieskantoor raadpleeg gevolglik dikwels die Prolingua-termlys wanneer Afrikaanse termekwivalente benodig word.

5.7 Nasionale Leksikografiese Eenhede

Die vereniging werk ook saam met die Nasionale Leksikografiese Eenhede (NLE's), soos die Buro van die Woordeboek van die Afrikaanse Taal (WAT) en die Dictionary of South African English (DSAE). Die twee NLE's hou Prolingua-lede op die hoogte van hul werksaamhede en die lede maak op hul beurt gebruik van die leksikografiese produkte van die NLE's.

Lede het in 2016 deelgeneem aan die Dictionary of South African English se aanlyn ondersoek om die aanlyn weergawe van die woordeboek te verbeter. Die sate wat sedert die 17de eeu in die historiese Engelse woordeboek opgeneem is, bevat taal- en linguistiese data wat waardevol is vir taalpraktisyns. Die doel van die aanlyn vraelys was om hierdie Suid-Afrikaanse kulturele leksikografiese bron deur middel van die elektroniese medium meer dinamies, buigbaar en toeganklik te maak (DSAE 2016).

5.8 Afrilex

Prolingua het 'n jarelange verbintenis met die African Association for Lexicography (Afrilex), 'n vakvereniging wat op leksikografie konsentreer en wat in

1994 gestig is. Verskeie taalpraktisyne is dan ook lede van sowel Prolingua as Afrilex en die lede van albei instansies word voortdurend op die hoogte gehou van hul onderskeie bedrywighede. Prolingua-lede publiseer gereeld in *Lexikos*, Afrilex se vaktydskrif en woon gereeld Afrilex-konferensies en -seminare by en Afrilex-lede bied praatjies aan by Prolingua.

5.9 Uitgewers

Anders as in die geval van algemene woordeboeke, is 'n vakwoordeboek gewoonlik slegs bestem vir 'n klein spesialisgebruikersgroep en om vakwoordeboeke te publiseer is dus nie so 'n betalende bedryf vir uitgewers nie (Hankom 2012).

Om vakwoordeboeke en terminologielyste aanlyn beskikbaar te maak, sal steeds 'n groot koste-implikasie hê. Die grootste uitgawe sal die omskakeling van papier- na elektroniese formaat wees van woordeboeke wat reeds uit druk is. Daar is uiteraard ook koste verbonde aan die bou van 'n webwerf en die instandhouding daarvan. Daar sal ook mannekrag- en koste-implikasies wees om woordeboeke en termlyste wat reeds elektronies beskikbaar is, in die geskikte elektroniese formaat te kry en gereeld met nuwer weergawes te vervang (Luther 2005; Van Huyssteen 2014).

Dit is sekerlik haalbaar, maar die vraag is egter of ander mediahuise en uitgewers by so 'n nasionale plan vir uitgewerye betrokke sal wil raak — dit gaan immers by elke mediahuis en uitgewer oor finansiële oorlewing in 'n uiters mededingende bedryf. Dit sal egter in belang van vaktaalontwikkeling wees indien indringende onderhandelinge met mediahuise en uitgewers gevoer sou kon word om hulle te oortuig dat dit vir hulle sowel as vaktaalgebruikers van groot waarde kan wees as vaktaalprodukte (as drukkopies en aanlyn) steeds gepubliseer kan word. Die Prolingua-termlys verskaf dus in 'n groot mate terminologiese data wat andersins verlore sou wees.

6. Die pad vorentoe

'n Daadwerklike gesamentlike poging behoort deur Prolingua en al bogenoemde instansies aangewend te word om meertalige terminologie in 'n verskeidenheid van dissiplines te ontwikkel, te dokumenteer en nasionaal beskikbaar te stel. Sonder vaktaal kan daar immers nie sprake wees van behoorlike vakkommunikasie nie. Die internasionale terminologienetwerk *TermNet* se leuse is: "There is no knowledge without terminology".

Verskeie (vak)taalkantore (by nasionale, provinsiale, munisipale en (semi)-staatsinstellings, banke, ens.) versamel steeds terminologie vir interne gebruik. Die nasionale terminologiekantoor stel meertalige terminologielyste saam en hou 'n nasionale register by om terminologieprojekte te registreer om samewerking te bewerkstellig en duplisering te voorkom. Privaat instansies, verenigings soos

Prolingua, sentra soos die Sentrum vir Politieke en Verwante Terminologie in Suider-Afrika (CEPTSA) en die Sentrum vir Regsterminologie in Afrikatale (SRTAT), VivA, tersiêre instansies, uitgewers en individue stel steeds vakwoordeboeke of termlyste saam (Alberts 2015a).

Bestaande infrastruktuur, soos taalkantore, termbanke en digitale platforms, en finansiering moet benut word eerder as om nuut ontwikkel of verkry te word. Alle terminologie wat beskikbaar is (selfs terminologieprodukte wat reeds uit druk is), behoort gedigitaliseer te word en gratis aanlyn (byvoorbeeld by die Terminologiekoördineringsafdeling/VivA/tersiêre instansies/Prolingua) beskikbaar te wees of daar behoort koppelvlakke en skakels te wees na alle vaktaalaktiwiteite (Alberts 2015b).

Data behoort gedigitaliseer en oop te wees vir vrye toegang en algemene gebruik, want data wat nie oop gelisensieer is nie, raak onbruikbaar. Die terme wat aanlyn in die oop domein beskikbaar gestel word, moet van 'n hoë gehalte wees en die doeltreffendheid daarvan moet gereeld gemeet word. Daar behoort ook stelsels te wees vir kommentaar op vaktaaldata en wat die hersiening en bywerking daarvan sal vergemaklik.

Dit sou geen doel dien om terminologie by Prolingua in isolasie te ontwikkel en vir eie gebruik te benut nie. Prolingua sou egter sy bestaande infrastruktuur, hulpbronne en bates kon aanwend om 'n enorme bydrae te lewer tot vaktaalontwikkeling, met spesiale klem op Afrikaanse vaktaal. Dit gaan myns insiens hier oor die volgende hulpbronne en bates:

- Vak- en taalkennis: Prolingua beskik oor lede wat 'n magdom vak- en taalkennis in pag het — baie Prolingua-lede is reeds afgetree, maar ander is nog aktief in bepaalde vakdissiplines en tale werkzaam — en hierdie kundigheid (ook van die afgetredenes) behoort benut te word.
- Vakwoordeboeke: Prolingua-lede beskik oor 'n groot versameling vakwoordeboeke en termlyste vir eie gebruik. Die Prolingua-vakwoordeboekversameling is aan die SAAWK se boekery beskikbaar gestel vir navorsingsdoeleindes. Dié vakwoordeboeke is egter tot 'n baie groot mate ontoeganklik, aangesien die boekery fisies in Pretoria geleë is. Boonop is die meeste van dié vakwoordeboeke nie digitaal beskikbaar nie, wat ook 'n risiko met betrekking tot veiligheid en volhoubaarheid inhou. Daarenteen is die Prolingua-termlys aanlyn beskikbaar.

Vervolgens word konkrete aanbevelings gemaak hoe elk van hierdie bates ten beste benut kan word.

Vak- en taalkennis

- Prolingua-lede se kundigheid behoort nasionaal benut te word.
- Die Prolingua-lys is aanlyn beskikbaar op die Prolingua-webwerf en met behulp van skakels/koppelvlakke aan ander stelsels verbind. Die spesifieke doel daarmee is om kommentaar van ander vaktaalgebruikers op terminologiese data te ontvang en self kommentaar te lewer.

- Deel kundigheid deur die vaktaaldata ook aan die nasionale termbank by die Terminologiekoördineringsafdeling, Nasionale Taaldiens, Departement van Kuns en Kultuur beskikbaar te stel sodat Afrikataalekwivalente bygevoeg kan word.

Vakwoordeboeke

- Gaan voort om vakwoordeboeke en termlyste te digitiseer wat nog nie in die Prolingua-lys opgeneem is nie. Die digitisering moet sover moontlik deur middel van skandering en optiese karakterherkenning geskied.
- Noue skakeling met veral VivA is noodsaaklik om duplisering te verhoed.
- VivA kan gedigitiseerde data digitaliseer.
- Die gedigitaliseerde Prolingua-termlyste kan op sowel Prolingua as VivA se webwerwe beskikbaar gestel word, met 'n aanduiding daarby dat Prolingua oor die outeursreg beskik.

Voordele verbonde aan die benutting van VivA se infrastruktuur is die volgende:

- Eenstopdiens: (Vak)taalgebruikers sal 'n eenstopdiens by VivA kan hê. Skakels kan (vak)taalgebruikers met Prolingua en ander bronne verbind.
- Navraagdiens: Terminologiese navrae kan deur VivA se taaladviesdiens gekanaliseer word. Die taaladviseur kan met behulp van bestaande terminologiebanke by VivA termnavrae beantwoord. Prolingua sou ook 'n lys van kundiges, wat bereid sou wees om die taaladviseur by te staan, aan VivA beskikbaar kon stel. Daar sou selfs 'n skakel kon wees na die Prolingua-geselsgroep (Prolingua99@yahoo.com).
- Tegnologie en infrastruktuur: Die nodige tegnologie en infrastruktuur bestaan reeds by VivA. Prolingua het dus nie nodig om enige platforms te onderhou of nuwe tegnologie te ontwikkel nie — dit word reeds deur VivA gedoen.
- Termbank(e): VivA se programmatuur en platforms maak daarvoor voorsiening dat data uit verskillende bronne benut kan word. Die data word met behulp van bestaande koppelvlakke met VivA se stelsel versoen en indien nodig kan 'n doelgemaakte koppelvlak ontwikkel word om nuwe projekte se terminologiese data met die stelsel te versoen. Nuwe projekte kan bloot by die bestaande stelsel gevoeg word sodra die data toeganklik gemaak is en aanlyn beskikbaar gestel word. VivA-terminologie kan ook aan die Afrikatale beskikbaar gestel word vir die byvoeging van die ander amptelike Suid-Afrikaanse tale
- Besikbaarstelling (gratis): Die terminologiese data kan gratis op VivA se webwerf en mobiele toepassing ("toep") beskikbaar gestel word. Dit pas in by VivA se sakemodel om soveel as moontlik inligting gratis beskikbaar te stel. (Dit voorkom ook dat nog geld op tegnologie spandeer word, terwyl dit eintlik aan inhoud spandeer kan word.)
- Besikbaarstelling (betaald): VivA se webwerf beskik ook oor die moontlikheid om sommige woordeboekinhoud teen betaling beskikbaar te stel.

Dit gaan egter spesifiek oor kommersiële woordeboeke (byvoorbeeld algemene woordeboeke soos die WAT en HAT en gepubliseerde vakwoordeboeke, byvoorbeeld deur uitgewers soos Juta en Pharos). In sulke gevalle kan gebruikers deur die webwinkel toegang tot die data bekom; tantieme (70%) word aan woordeboekverskaffers (byvoorbeeld Prolingua) betaal, want die lisensiehouer kry telkens betaling wanneer sy data deur middel van VivA benut word. Die aanbeveling hier is egter dat Prolingua soveel as moontlik van sy vakwoordeboekdata gratis deur middel van VivA aan gebruikers beskikbaar moet stel.

- Kundigheid: VivA het reeds 'n model daargestel om vak-, taal- en tegnologiese kundigheid te verenig. CTeX van die Noordwes-Universiteit verskaf die tegnologiese kundigheid wat ook in die artikel bespreek word en Prolingua sou dus kon bydra deur taal- en vakkundigheid hiertoe toe te voeg.
- Opleiding: Teoretiese vakleksikografiese opleiding sou aanlyn by VivA se webseminare (webinare) aangebied kon word.

7. Samevatting

Taalpraktisyns het daagliks terminologie nodig vir die werksaamhede verbonde aan hul beroepe. Afrikaanse terme is klaarblyklik steeds nodig en waar dit nie aktief deur ander instansies ontwikkel word nie, het Prolingua en sy voorgangers getoon dat die instansie doelbewus poog om 'n nasionale terminologiesiens te lewer en daardeur sorg te dra dat die ontwikkeling van veral Afrikaanse vaktaal nie agterweë bly nie. Prolingua kan deur sy eie inisiatiewe Afrikaanse terme volhoubaar ontwikkel, bewaar en dissemineer. Prolingua kan ook deur samewerking met ander instansies aktief betrokke wees by die voortdurende en volhoubare ontwikkeling van die amptelike tale as gebruiks- en wetenskapstale wat vir alle doelgebruikers van waarde kan wees.

Dit sou interessant wees om te sien in welke mate geïsoleerde terminologieontwikkeling, soortgelyk aan dié vir Afrikaans deur Prolingua, in die ander amptelike tale deur enkele instansies onderneem word.

8. Bibliografie

- Alberts, M. 1998. *Vaktaalontwikkeling in Suid-Afrika*. Referaat gelewer by LVSA-kongres, 1 Julie 1998, Potchefstroom.
- Alberts, M. 2000a. *Audit on South African Terminology Projects*. Julie 2000. Ongepubliseerde dokument. Pretoria: Nasionale Terminologiesiens, Departement van Kuns, Kultuur, Wetenskap en Tegnologie.
- Alberts, M. 2000b. Terminology Management at the National Language Service. *Lexikos* 10: 234-251.
- Alberts, M. 2003. Vaktaalontwikkeling en -opleiding in Suid-Afrika. Botha, Willem (Red.). 2003. *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 127-161. Stellenbosch: Buro van die WAT.

- Alberts, M.** 2012. *Sentrum vir Regsterminologie in Afrikatale (SRTAT) — 'n Oorsig van die projek oor straf-, strafproses- en bewysreg*. Referaat gelewer by die Prolingua-werksvergadering, 10 Julie 2012, Unisa, Florida-kampus, Roodepoort.
- Alberts, M.** 2015a. *'n Oorsig van vaktaalontwikkeling en -opleiding in Suid-Afrika met spesiale verwysing na Afrikaans*. Ongepubliseerde verslag opgestel in opdrag van Die Dagbreek Trust en die Suid-Afrikaanse Akademie vir Wetenskap en Kuns, November 2015.
- Alberts, M.** 2015b. *Die Akademie se Vaktaalbuuro — verlede, hede en toekoms*. Ongepubliseerde verslag opgestel in opdrag van Die Dagbreek Trust en die Suid-Afrikaanse Akademie vir Wetenskap en Kuns, Desember 2015.
- Alberts, M.** 2017. *Terminology and Terminography Principles and Practice. A South African Perspective*. Milnerton: McGillivray Linnegar (MLA Publications).
- Alberts, M. en N. Mollema.** 2013. Developing Legal Terminology in African Languages as an Aid to the Court Interpreter: A South African Perspective. *Lexikos* 23: 29-58.
- Beukes, S.M.** 2014. *SA Language Practitioners' Council Bill*. Referaat gelewer by die Prolingua-werksvergadering, 13 Mei 2014, Unisa, Florida-kampus, Roodepoort.
- Beukes, S.M. en M. Pienaar.** 2011. *Uitdagings en struikelblokke by die vertaling van die FIT-vertaalterminologielys in Afrikaans*. Referaat gelewer by die Prolingua-werksvergadering, 8 Maart 2011, Unisa, Florida-kampus, Roodepoort.
- Bosman, J.P.** 2015. *Ons Griekse-Latynse erfenis/Our Latin Greek Heritage. Ons dink, praat en skryf daaglik in Grieks en Latyn en ons besef dit nie: 'n Vertrekpunt vanuit die mediese taal*. Referaat gelewer by die Prolingua-werksvergadering, 10 Junie 2015, Tshwane Munisipaliteit, Centurion.
- Botha, S., P. le Clus en A.J. Venter (Reds.).** 2011. *Modern Political Dictionary/Nuwerwetse Politieke Woordeboek*. CEPTSA. Auckland Park: Universiteit van Johannesburg.
- Burger, C.** 2016. *Radiotegnologie en radioterme*. Referaat gelewer by die Prolingua-werksvergadering, 12 Oktober 2016, Tshwane Munisipaliteit, Centurion.
- Burger, P. en J. de Jong.** 1997. *Handboek Stijl: Adviezen voor aantrekkelijk schrijven*. Groningen: Martinus Nijhoff.
- Buys, F.** 2014. Afrikaanse universiteit bekendgestel. *Pretoria East Rekord*, November 2014: 4.
- Carstens, W.A.M.** 2011. *Teksredaksie (2010)*. Referaat gelewer by die Prolingua-werksvergadering, 10 Mei 2011, Unisa, Florida-kampus, Roodepoort.
- Carstens, W.A.M.** 2014. *Die storie van Afrikaans: Uit Europa en van Afrika. 'n Biografie van Afrikaans*. Referaat gelewer by die Prolingua-werksvergadering, 8 Oktober 2014, Tshwane Munisipaliteit, Centurion.
- Carstens, W.A.M. en K. van de Poel.** 2010. *Teksredaksie*. Stellenbosch: SUN Media.
- CLTAL (Centre for Legal Terminology in African Languages).** 2015. *Legal Terminology: Criminal Law, Procedure and Evidence/Regsterminologie: Straf-, Strafproses- en Bewysreg*. Kaapstad: Juta.
- Combrink, A.** 2015. Moets, moenies en hoekoms: Taalversorging van akademiese geskifte. Werksessie, 10 Maart 2015, Unisa, Florida-kampus, Roodepoort.
- Cornelius, E.** 2012. *'n Verkenning van die leksikogrammatikale kompleksiteit van regstaal en die vereenvoudiging daarvan*. Referaat gelewer by die Prolingua-werksvergadering, 10 April 2012, Tshwane Munisipaliteit, Centurion.
- Dorfling, J.** 2002–2011. *Brugwoordeboek*. www.prolingua.org.za.
- Dorfling, J.** 2012. *Brugterme en die Brugakademie in Pretoria*. Referaat gelewer by die Prolingua-werksvergadering, 12 Junie 2012, Tshwane Munisipaliteit, Centurion.

- DSAE.** 2016. DSAE Reader Survey. (www.dsae.co.za; <http://dsae.limequery.com/index.php/survey/>). April 2016.
- Du Plessis, M.** 2016. *Storietyd! Terme en begrippe in die letterkunde*. Referaat gelewer by die Prolingua-werksvergadering, 10 Februarie 2016, Tshwane Munisipaliteit, Centurion.
- Dykman, D.J.** 2012. *When You Use those Words, Don't Abuse, Confuse or Misuse those Words*. Werkwinkel oor die korrekte gebruik van Engelse woorde aangebied by die Universiteit van Johannesburg, 11 September 2012, Auckland Park.
- Dykman, D.J.** 2013. *When You Use those Words, Don't Abuse, Confuse or Misuse those Words*. Referaat gelewer by die Prolingua-werksvergadering, 9 Oktober 2013, Tshwane Munisipaliteit, Centurion.
- Dykman, D.J.** 2015. *Powerful Presentations: Guidelines on Effective Speaking*. Werksessie oor die gebruik van Engels aangebied by die Prolingua-werksvergadering, 14 Oktober 2015, Tshwane Munisipaliteit, Centurion.
- Dykman, D.J., J.D.U. Geldenhuys en E.E. Viljoen-Smook.** 2014. *The Write Stuff: The Style Guide with a Difference*. Derde uitgawe. Kaapstad: NB-Uitgewers.
- Fourie, W.** 2014. *Autshumato-projek*. Referaat gelewer na afloop van Prolingua se Algemene Jaarvergadering, 10 Augustus 2014, Tshwane Munisipaliteit, Centurion.
- Geldenhuys, J.D.U. en E.E. Viljoen-Smook.** 2009. *Business Dictionary/Sakewoordeboek*. Vyfde hersiene uitgawe. (www.pharosonline.co.za). Kaapstad: Pharos.
- Geldenhuys, J.D.U., D.K. Dykman en B. Viljoen-Smook.** 2012. *Saketaal, die jongste Sakewoordeboek en 'The Write Stuff'*. 'n Vraag-en-antwoordsessie aangebied by die Prolingua-werksvergadering, 8 Mei 2012, Unisa, Florida-kampus, Roodepoort.
- Hanekom, D.** 2012. *Die uitgewersbedryf en die publikasie van boeke*. Referaat gelewer by die Prolingua-werksvergadering, 14 Februarie 2012, Tshwane Munisipaliteit, Centurion.
- Human, D.** 2016. *Uitdagings om die Bybel beter te verstaan*. Referaat gelewer na afloop van Prolingua se Algemene Jaarvergadering, 10 Augustus 2016, Tshwane Munisipaliteit, Centurion.
- Jonker, T.** 2013. *Sweeftuie en sweeftuigterme*. Referaat gelewer by die Prolingua-werksvergadering, D-Ring 353, Universiteit van Johannesburg, Aucklandpark-kampus.
- Kotzé, E.F.** 2015. *Die aard en werkswyse van die Forensiese Linguistiek*. Referaat gelewer na afloop van Prolingua se Algemene Jaarvergadering gehou op 12 Augustus 2015 by die Kunstgalerie, Voortrekkermonument, Pretoria.
- Le Roux, W.** 2016. Benamings van die land se howe. *Taalwys*, 7 Junie 2016.
- Liebenberg, W., M. Delpont en E. Greenfield.** 2016. *WordFast Classic Workshop*. 'n Inleiding tot WordFast Classic aangebied op 23–25 Mei 2016, Rigellaan, Pretoria.
- Linnegar, J.D.** 2013. *CCC: An Inclusive New Model for Ensuring that Text Editors Intervene in Texts Systematically*. Referaat gelewer by die Prolingua-werksvergadering, 12 Junie 2013: Tshwane Munisipaliteit, Centurion.
- Linnegar, J.D.** 2014. *CCC: An Inclusive New Model for Ensuring that Text Editors Intervene in Texts Systematically*. Miniwerkwinkel aangebied by die Prolingua-werksvergadering, 9 September 2014, Unisa, Florida-kampus, Roodepoort.
- Luther, Jana.** 2005. *Vaktaalontwikkeling*. Ongepubliseerde artikel.
- Malan, K.** 2014. *Amptelike tale as tale vir die notulering van hofverrigtinge*. Referaat gelewer by die Prolingua-werksvergadering, 11 Junie 2014, Tshwane Munisipaliteit, Centurion.
- Malete, N.E.** 2016. *Metheo ya ditokiso tsa sengolwa*. Milnerton: McGillivray Linnegar (MLA Publications).

- Maree, K.** 2014. *Social Media: Our Changed Reality*. Referaat gelewer by die Prolingua-werksvergadering, 12 Februarie 2014, Tshwane Munisipaliteit, Centurion.
- McLachlan, J.D.** 2011. *Taalgebruik*. 'n Taalgebruiksessie aangebied na afloop van Prolingua se Algemene Jaarvergadering, 2 Augustus 2011, Tshwane Munisipaliteit, Centurion.
- McLachlan, J.D.** 2012. *Taalnavrae*. 'n Taalgebruiksessie aangebied na afloop van Prolingua se Algemene Jaarvergadering, 14 Augustus 2012, Tshwane Munisipaliteit, Centurion.
- Mollema, N. en M. Alberts.** 2011. *Terminological Research: A Practical Guide for Court Interpreters*. Referaat gelewer by die Departement van Justisie se konferensie: The Value and Professionalism of the Court Interpreter, 14–15 Oktober 2011, Universiteit van die Vrystaat, Bloemfontein.
- Mtintsilana, P.N. en R. Morris.** 1988. Terminography in African Languages in South Africa. *South African Journal of African Languages* 8(4): 109-113.
- Nasionale Terminologiesiens.** 1998. *Katalogus van vakwoordeboeke/Catalogue of Technical Dictionaries*. (NTD 1998.) Pretoria: Departement van Kuns, Kultuur, Wetenskap en Tegnologie.
- Posthumus, L.** 2016. *Deiksis en taal: Deiksis en die relevansie daarvan vir kommunikasie en vertaling*. Referaat gelewer by die Prolingua-werksvergadering van 8 Maart 2016, Unisa, Florida-kampus, Roodepoort.
- Prinsloo, D.J.** 2011. *Tekstegnologie — Die versameling, analise, manipulering en markering van teks*. Referaat gelewer by die Prolingua-werksvergadering, 11 Oktober 2011, Tshwane Munisipaliteit, Centurion.
- Prolingua.** 2015. *Prolingua 65! 'n Vereniging vir Afrikaanse en Engelse Taalpraktisyne. An Association for English and Afrikaans Language Practitioners*. 28 Oktober 1950–12 Augustus 2015. Pamflet saamgestel ter viering van Prolingua se 65ste bestaansjaar.
- Prolingua.** 2016. *Notule van 'n werksvergadering van Prolingua* gehou op Dinsdag, 13 September 2016, Florida-kampus van Unisa, Roodepoort.
- Prolingua.** 2017a. *Notule van 'n werksvergadering van Prolingua* gehou op Woensdag 12 April 2017, Tshwane Munisipaliteit, Centurion.
- Prolingua.** 2017b. *Items vir Bespreking (IvB)*. Werksvergadering 9 Mei 2017, Tshwane Munisipaliteit, Centurion.
- Roux, J.C.** 2016. *Spraakbronne met onder meer leksikografie, terminologie en vertaling*. Referaat gelewer by die Prolingua-werksvergadering, 12 Oktober 2016, Centurion. Tshwane Munisipaliteit, Centurion.
- Russouw, J.** 2017. *Sport in Afrikaans oor die radio*. Referaat gelewer by die Prolingua-werksvergadering, 8 Februarie 2017, Tshwane Munisipaliteit, Centurion.
- SAAWK.** 2015a. *Notule van die Verkennende Gesprek oor Terminologie-ontwikkeling*. 28 Augustus 2015. SAOU Raadsaal, Serenestraat 278, Garsfontein.
- SAAWK.** 2015b. *Nuusbrief* 53(2). Augustus 2015. www.akademie.co.za.
- Sentrum vir Regsterminologie in Afrikatale (SRTAT).** 2017. *SRTAT-databasis: Regsterminologie: Straf-, Strafproses- en Bewysreg*. SRTAT: Pretoria.
- Van de Poel, K., W.A.M. Carstens en J.D. Linnegar.** 2012. *Text Editing: A Handbook for Students and Practitioners*. Antwerpen, België: Antwerpen Universiteitsdrukkers.
- Van Huyssteen, G.B.** 2014. *VivA! Die Virtuele Instituut vir Afrikaans*. Referaat gelewer by die Prolingua-werksvergadering, 9 April 2014, Tshwane Munisipaliteit, Centurion.
- Van Huyssteen, G.B.** 2017. *Lugvaartterme*. Beskikbaar gestel uit die rubriek in die *Beeld* oor Lugvaartterme. E-pos gestuur op 2017-03-06.

- Van Huyssteen, G.B. en S. Kapp.** 2017. *Die nuwe Afrikaanse woordelys en spelreëls (2017)*. Referaat gelewer by die Prolingua-werksvergadering, 11 Oktober 2017, Tshwane Munisipaliteit, Centurion.
- Van Nierop, L.** 2016. *Wat is in 'n naam? Die betekenis van slottitels in rolprente*. Referaat gelewer by die Prolingua-werksvergadering van 10 Mei 2016, Unisa, Florida-kampus, Roodepoort.
- Van Rooy, B.** 2017. *The Distinction between Innovation and Integration of New Expressions in a Language. Die onderskeid tussen innovasie en inburgering van nuwe uitdrukkings in 'n taal*. Referaat gelewer by die Prolingua-werksvergadering, 9 Mei 2017. Unisa, Florida-kampus, Roodepoort.
- Vereniging van Regslui vir Afrikaans (VRA).** 2017. *Landdroshowe se boek oor reëls in Afrikaans*. Vertaling deur McLachlan, J.D. Pretoria: Maroela Media.
- Viljoen-Smook, E.E., J.D.U. Geldenhuys en W. Coetzee.** 2017. *Business Writing for South Africans*. Kaapstad: NB Uitgewers.
- Von Solms, B.** 2013. *Die Kuberruimte — Die moderne Wilde Weste?* Referaat gelewer by die Prolingua-werksessie, Dinsdag, 14 Mei 2013, D-Ring 353, Universiteit van Johannesburg, Aucklandpark-kampus.
- Vos, C.** 2015. *Die literêre vertaling van die Ilias (Homerus) van Grieks na Afrikaans*. Referaat gelewer by die Prolingua-werksvergadering, 8 April 2015, Tshwane Munisipaliteit, Centurion.

Wetgewing

SA Grondwet:

Constitution of the Republic of South Africa, 1996. (Act No. 108 of 1996.) (Nou: Constitution, 1996.)/
Grondwet van die Republiek van Suid-Afrika, 1996. (Wet 108 van 1996.) (Nou: Grondwet, 1996.)

Outeursreg/Kopiereg:

Copyright Act 98 of 1978/Kopieregwet 98 van 1978.

Copyright Amendment Act 9 of 2002/Kopiereg wysigingswet 9 van 2002.

Nasionale Taalbeleid en Implementeringsplan:

National Language Policy Framework. Final Draft (NLPF)/Nasionale Taalbeleidsraamwerk. Finale Konsep. 13 November 2002. Pretoria: Departement van Kuns en Kultuur.

National Language Policy Framework (NLPF) and Implementation Plan/Nasionale Taalbeleidsraamwerk en Implementeringsplan. 2003. Pretoria: Departement van Kuns en Kultuur (DKK).

Provinsiale Taalbeleidsraamwerk:

Language Policy Framework of the Gauteng Provincial Government/Taalbeleidsraamwerk van die Gautengse Provinsiale Regering. 2005. Johannesburg: Gautengse Provinsiale Regering, Departement van Sport, Ontspanning, Kuns en Kultuur.

SA Taalpraktisynsraad:

South African Language Practitioners' Council Act 8 of 2014/Suid-Afrikaanse Taalpraktisynsraad Wet 8 van 2014.

Gebruik van Amptelike Tale:

Use of Official Languages Act 12 of 2012/Wet op die Gebruik van die Amptelike Tale, Wet 12 van 2012. Staatskoerant No. 35742, 2 Oktober 2012. Kaapstad: Staatsdrukker.

Use of Official Languages Act — Proposed Regulations/Wet op die Gebruik van die Amptelike Tale — Voorgestelde Regulasies. Staatskoerant No. 36392, Staatskennisgewing 411 van 2013, 26 April 2013: 56-67. Kaapstad: Staatsdrukker.

Guessing Verb–Adverb Collocations: Arab EFL Learners' Use of Electronic Dictionaries*

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

Abstract: Collocational studies have recently attracted a great deal of interest. To date, hardly any study has tackled Arab EFL learners' competency in the use of verb–adverb collocations. This study explores the way advanced Arab EFL learners handle verb–adverb collocations using two learner's dictionaries. The subjects (N = 82) were required to look up 22 verbs, 12 frequent and 10 infrequent, and guess three adverb collocates of each verb. The results showed that even advanced EFL learners had considerable difficulty in providing extra adverb collocates of both frequent and infrequent verbs. Dictionary use was effective; the subjects performed significantly better especially with infrequent verbs. Many reasons were posited for this finding, foremost of which included the subjects' deficiency in collocational skills and the lack of sufficient clues in the definitions to facilitate dictionary users' correct use of collocations. Whereas knowledge of the meaning of the stimulus was found to be a significant contributing factor to the subjects' overall collocational competence, basic prior training in dictionary usage did not show any positive impact on their overall performance.

Keywords: VERB–ADVERB COLLOCATIONS, ARAB EFL LEARNERS, FREQUENT VERBS, INFREQUENT VERBS, TRAINING IN DICTIONARY USE, COLLOCATIONS DICTIONARIES, DICTIONARY USE, LEARNER'S DICTIONARIES

Opsomming: Die raai van werkwoord–bywoord-kollokasies: Arabiese EVT-leerders se gebruik van elektroniese woordeboeke. Kollokasienavorsing het onlangs baie belangstelling ontlok. Tot op hede het byna geen studie Arabiese EVT-leerders se vaardighede in die gebruik van werkwoord–bywoord-kollokasies ondersoek nie. Hierdie studie ondersoek die manier waarop gevorderde Arabiese EVT-leerders met werkwoord–bywoord-kollokasies omgaan terwyl hulle twee aanleerderswoordeboeke gebruik. Die proefpersone (N = 82) is versoek om 22 werkwoorde, waarvan 12 gebruiklik en 10 ongebruiklik was, na te slaan en drie bywoordelike kollokasies vir elke werkwoord te raai. Die resultate het getoon dat selfs gevorderde EVT-leerders aansienlike probleme ondervind het om ekstra bywoordelike kollokasies vir sowel gebruiklike as ongebruiklik werkwoorde te verskaf. Woordeboekgebruik was effektief; die proefpersone het aansienlik beter gevaar by veral gebruiklike werkwoorde. Baie redes is voorgestel vir hierdie bevinding, waarvan die belangrikstes die proefpersone se gebrek aan kollokasionele vaardighede ingesluit het, asook die gebrek aan voldoende leidrade in die definisies om woordeboekgebruikers se

* This is a sequel to Alzi'abi (2017). The two studies made use of the same material and subjects but each had its own aims, procedure and results.

korrekte gebruik van kollokasies te vergemaklik. Terwyl daar bevind is dat kennis van die betekenis van die stimuli 'n betekenisvolle bydraende faktor tot die proefpersone se oorkoepelende kollokasionele bedrewenheid gelewer het, het basiese vorige opleiding in woordeboekgebruik geen positiewe invloed op hulle algehele prestasie gehad nie.

Sleutelwoorde: WERKWOORD–BYWOORD-KOLLOKASIES, ARABIESE EVT-LEERDERS, GEBRUIKLIKE WERKWOORDE, ONGEBRUIKLIKE WERKWOORDE, OPLEIDING IN WOORDEBOEKGEBRUIK, KOLLOKASIEWOORDEBOEKE, WOORDEBOEKGEBRUIK, AANLEERDERS-WOORDEBOEKE

1. Introduction

Collocations are crucially important to language competency and fluency (Lewis 2000 and Wray 2002). Hausmann (cited in Busse 1995: 125) has rightly claimed that "learning a language is learning collocations". This notion has also been echoed by compilers of the *Oxford Collocations Dictionary for Students of English* (OCD) who point out that no piece of natural spoken or written English is totally free of collocations (2009: v). Historically, EFL educators have been known to altogether neglect collocations in their teaching repertoire (Bahns 1993). Presently, there has been a surge of both interest in and availability of research covering all aspects of collocations.

Researchers, to date, have mainly focused on verb–noun collocations using EFL learners of diverse linguistic backgrounds including Arabs (Alzi'abi 2012), Afrikaners (Nizonkiza, Van Dyk and Louw 2013), Chinese and Swedish (Wang and Shaw 2008), Dutch (Peters 2016), Germans (Nesselhauf 2005), Israelis (Laufer and Waldman 2011), Lithuanians (Juknevičienė 2008), Japanese and French (Kurosaki 2012), Malaysians (Ang, Rahim, Tan and Salehuddin 2011), Poles (Szudarski 2012), Russians (Siyanova and Schmitt 2008), Spanish (Zinkgräf 2008), Taiwanese (Kuo 2009), Turkish (Akpýnar and Bardakçý 2015) and Vietnamese (Nguyen and Webb 2016). Although contrasting findings may be found in the aforementioned works, there appears to be a consensus among researchers that the correct acquisition and use of collocations has proved to be highly challenging for most EFL learners regardless of their linguistic background (see Laufer 2011 and Sun and Wang 2003).

To date, very few empirical works have focused on the means by which EFL learners acquire collocations and their effective use of dictionaries to extract collocational information. Philip (2007: 2) has called for greater research efforts to assess "whether collocation errors are indeed mislearned or misremembered collocations, or if they are something else entirely — calqued or invented forms — with the 'arbitrary' collocations being avoided instead". Currently, there is little to no research available that has elucidated the means by which EFL learners, particularly those of Arab background, utilise dictionaries for the production of verb–adverb collocations. The present study aims at addressing this research gap and may provide invaluable insight into how

advanced Arab EFL learners process verb–adverb collocations using two of the available learner's dictionaries. The subjects will be involved in dictionary look-ups for verbs to provide additional adverb collocates relating to those verbs using the electronic version of either the *Longman Dictionary of Contemporary English* 5th edition (LDOCE) or the *Oxford Advanced Learner's Dictionary* 9th edition (OALD). These two dictionaries were selected as they were the most popular references among Arab EFL learners.

It should be added that learner's dictionaries are not typically scripted to solely teach collocations since they do not include an exhaustive amount of collocational information. It is, therefore, worth conducting a study to establish whether this information would systematically enable advanced EFL learners to infer which other items, i.e. adverb collocates occur with a certain verb.

In the sections to follow, the concept and meaning of collocation is defined with the assistance of relevant studies to date. An elucidation of the method, procedure, findings and discussion of the present study comes next.

2. What is collocation?

Researchers use a variety of expressions to refer to chunks of two word expressions including the phenomenon of collocation. These include, *inter alia*, *fixed expressions* (Alexander 1984); *formulaic language* or speech (Weinert 1995); *multi-word items* (Moon 1997); *prefabricated chunks* (Williams 1998) and *word combinations* (Howarth 1998). Wray (2002) uses the expression *formulaic sequences* as an umbrella term covering all the aforementioned formulations. Conceivably, the abundance of available terminologies has made it challenging to settle on an exact and satisfactory definition of collocation. Although many terminologies abound, they all encompass the same concept albeit approaching it from a different perspective. However, none to date are entirely foolproof.

One particular definition which may encompass the notably wide elusive concept of collocation and will better fit the goals of the present study has been given by Gries (2008: 3). He states that collocation is "the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance". Although this definition is exhaustive, it must nonetheless be noted that the suggestion of co-occurrence of the items imposes undue constraints on the way in which combination relation works. This is technically referred to as the 'restricted commutability' of the components of the chunk. For instance, any given verb will have a set of adverbs that modify it in order for a specific meaning to be communicated. These restrictions in the substitutability of the collocates are, indeed, the root of the problem in grasping the concept of collocations.

3. Literature review

Researchers in the field have maintained a sustained interest in multiword units; among which collocations have assumed a central position (see Nesselhauf 2003 and Nizonkiza, Van Dyk and Louw 2013). Studies completed to date may broadly fall into four categories: 1) Studies that assess EFL learners' competence in forming collocations correctly (e.g. Farghal and Obeidat 1995; Hussein 1990; Zughoul and Abdul-Fattah 2003). Included in this category are those that focus on the learners' miscollocations and ill-formed productions (e.g. Farghal and Al-Hamly 2007; Granger 1998; Kuo 2009; Kurosaki 2012). 2) Studies exploring the causal relationship between mastery of collocations and the subjects' level of proficiency or linguistic skills (e.g. Akpýnar and Bardakçý 2015 and Nizonkiza 2011). 3) Studies investigating the impact of implicit or explicit instruction, noticing, providing visual enhancement as well as raising collocational awareness on the learners' successful use of collocations (Alfahadi, Zohairy, Momani and Wahby 2014; Farrokh 2012; Szudarski 2012). 4) Studies describing syntactically and semantically parsed corpora in attempts to extract collocations either for pedagogical or lexicographic purposes (e.g. Jaén 2007; Kennedy 2003; Krenn and Evert 2001; Seretan 2011; Smadja 1993). It should be noted that these four categories are not a binary choice since some studies may be found to encompass more than one theme and will not strictly fall into one classification.

3.1 General collocational studies

Hussein (1990) investigated the ability of 200 third- and fourth-year English majors of Arab background to collocate items correctly. The subjects generally performed poorly on a 40-item multiple choice collocation test and failed to collocate more than half of the time. Farghal and Obeidat (1995) involved 34 Arab learners of English in a gap-filling task (22 common collocations) and Arabic–English translation task (23 expressions). Results showed overall poor performance across the board; however, the subjects did well when convergence existed between L1 and L2. Granger (1998) and Lorenz (1999) also reported similar findings with French and German EFL learners, respectively. Comparing the learners' production of the adverbs ending in *-ly* (Granger 1998) and adverbs not ending in *-ly* (Lorenz 1999) with that of native speakers, the researchers found that EFL learners underused adverb collocates. The adverbs used were limited to those congruent with adverb–adjective combinations in their mother tongues, particularly those with a more 'restricted adverb' (see Peters 2016).

Two further studies reported Arab EFL learners' overall ignorance of collocations: Al-Amro (2006) and Alsakran (2011). The former study assessed the collocational knowledge of 51 Saudi advanced English learners through cloze, multiple choice and essay-writing tests. It found a close relation between the subjects' comprehension and production of collocations but no relation was

found between knowledge of collocations and the subjects' language proficiency. Alsakran (2011) involved 38 advanced Arab learners in an appropriateness judgment test and reported subject' incompetence in using collocations in all tests. They, however, performed better in the comprehension tasks with all types of collocations under scrutiny, verb–noun collocations being foremost.

Satriawan (2009) explored the acquisition of adverb collocates and compared the types and tokens of the degree, aspect and manner of adverbs in both natives and non-natives' adverbial collocations. The author examined the latter's adverbial collocations native likeness using three corpora which contained the adverb production of Indonesian TEFL undergraduates and native English speakers. Non-natives were found to use far less amplifiers and down-toners to modify adjectives and lexical verbs versus the native subjects. Moreover, the non-natives were found to use twice as many manner adverbs as natives. Interestingly, 75% of non-natives' adverb collocations (228) of adjectives, adverbs and verbs were deemed acceptable.

The above studies have highlighted learners' problems with the acquisition of collocations and some attributed subjects' lack of collocational competency to the negligence of the lexical approach in EFL teaching programs (Al-Amro 2006). In addition, negative transfer, overgeneralization, unfamiliarity with the structure of collocations (Hussein 1990) and the use of lexical simplification (Farghal and Obeidat 1995) have also proven to be significant contributing factors.

3.2 Studies involving the use of electronic dictionaries for collocations

Dziemianko (2010) investigated the influence of using paper versus electronic versions of *Collins COBUILD Advanced Learner's Dictionary* on the retention of meaning and collocations. Sixty-four upper-intermediate Polish learners of English performed receptive and productive tasks. The electronic dictionary was found to be significantly more effective than the printed version in both tasks. However, this finding is anomalous when compared with those from earlier works where no difference existed between electronic and paper dictionaries. A similar study to Dziemianko (2010) was Laufer's (2011) which explored the effect of dictionary usage on high school learners' production and retention of verb–noun collocations. The subjects were encountering difficulties in locating the right verbs to collocate with some nouns. Very low scores were obtained with regard to the retention of the looked-up collocations. Dictionaries significantly, though moderately, increased the number of correct collocations in the fill-in task.

Alzi'abi (2012) investigated the comparative effectiveness of using two different types of dictionaries to extract collocational information. Twenty verb–noun expressions, ten correct combinations and ten made-up ones were used in two tests in conjunction with the *Cambridge Advanced Learner's Dictionary 3rd edition*¹ (CALD) and OALD. A pretest required 130 second-year Syrian English

majors to judge the appropriateness of the items dictionary-free and a main test which required them to judge the acceptability of the same items but dictionary-aided. The subjects performed poorly on both tests. Interestingly, dictionary use significantly improved subjects' performance but the difference between the two dictionaries was not significant. The subjects were incapable of taking full advantage of the collocational information in dictionaries owing to hasty exploration strategies or lack of dictionary using skills.

Hamad and Laohawiriyanon (2013) investigated Thai learners' knowledge of English collocations following dictionary consultation. A seven-week course exposed 47 first-year university students to twelve high-frequency verb–noun collocations. They used the *Macmillan English Dictionary online* to decipher the meanings of the collocations. The subjects gained significant collocation skills and this was ascribed to the teaching method in the course, which involved intensive involvement tasks and regular practice. Although the findings were found to be consistent with the results from earlier works (Alsakran 2011; Dziemianko 2010; Laufer 2011; Murnani and Salehi 2016), it nonetheless would remain doubtful whether they could be generalisable since only a small number of subjects were included with a limited set of stimuli.

Alzi'abi (2016a) examined the efficacy of OALD in electronic form in relation to the Arab learners' production of verb–noun collocations. Twenty-two low-frequency verbs were used in two tests. The first aimed to ascertain whether the subjects (54 MA students majoring in English) could replace the "etc." in the dictionary definitions of the target verbs with three noun collocates functioning as 'objects' of the verbs under investigation. The second required the subjects to judge, from the definitions, the appropriateness of a set of four noun collocates used with each target verb. The participants did not perform well in Test1 and provided only 40% correct answers. However, they achieved much better scores in the judgment test. The subjects were better judges of noun collocates of the verbs than providers. The dictionary was found to offer little to no help and lacked sufficient clues as to what may correctly substitute for *etc.* The blame was placed on the format of some of the definitions, which were error-conducive.

Most of the aforementioned studies focused on verb–noun and adjective–noun collocations but rarely concerned the other categories. The studies also shared a common major limitation. The researchers generally requested the subjects to provide single collocates of the stimuli that reflected only part of their L2 collocational performance (Fan 2009). Moreover, they were wholly devoted to exploring the use of a particular type of collocations, i.e. verb–noun and/or adjective–noun, and consequently had restricted a balanced assessment of the learners' collocational competence. Furthermore, the frequency of materials used was mostly uncontrolled. It is therefore unclear whether the findings from these studies findings are entirely reliable.

Despite the abundant literature on collocations, there still remains a paucity of research conducted on the way in which learners would acquire and use

verb–adverb collocations. There is also little to no research into the role of electronic dictionaries in helping users utilise verb–adverb collocations. No single study to date, to the best of the author's knowledge, has been solely designed to tackle EFL learners' use of verb–adverb collocations. There is still a pressing need for conducting research into the specific means employed by non-natives to acquire and use verb–adverb collocations and thus elucidate the source of the common difficulties encountered. The findings of such investigations may help devise more effective and targeted teaching methods to be utilised by EFL educators to promote the acquisition of this slightly less researched type of collocation.

4. The study

The primary aim of the present study is to assess the difficulties encountered by Arab EFL learners' in using electronic dictionaries for extracting and using verb–adverb collocational information correctly. In order to realise this aim, three study objectives and related hypotheses are stated as follows:

4.1 Aims

1. To determine, whether or not, Arab dictionary users provide extra adverb collocates of verbs based on dictionary look-ups of these verbs. In addition, it will seek to establish whether a difference in participants' performance level exists when using frequent vs. infrequent verbs. In following the precedent set out by prior works on collocations (e.g. Alzi'abi 2012), Arab dictionary users in this study will not be expected to habitually provide additional collocates for low-frequency verbs. It is hypothesised that subjects will likely respond appropriately half of the time and will perform better with the frequent items that have previously been learnt. Thus, in such cases it is also expected that subjects will encounter less difficulty in extrapolating these collocates correctly from the provided electronic dictionary.
2. To determine whether dictionaries vary in the amount of help they offer users with regard to the correct production of verb–adverb collocations. To date, there is a lack of consensus on this issue with some researchers indicating significant differences between the dictionaries employed on vocabulary acquisition (Dziemianko 2010), whilst others have revealed no significant differences at all (Lew and Radłowska 2010). It is hypothesised that LDOCE and OALD, included in the present study, will show variance in the amount of assistance offered to their respective users. It is further assumed that LDOCE may offer comparatively more help to users since there are greater numbers of examples cited, and more importantly, it contains greater amounts of collocational information than OALD.

3. To establish the effects of having prior basic training in dictionary use and pre-knowledge of the meaning of the stimuli on the subjects' performances. It is hypothesised that these two individual factors, each with its own merit, will contribute positively to the subjects' success in producing correct adverb collocates (see Murnani and Salehi 2016 and Peters 2016).

There is a noteworthy question to be posed at this juncture. Some may argue that dictionaries have little to no influence in encouraging the extrapolation of other adverbial collocates so why should a dictionary be expected to help users guess such collocates that are omitted? Dictionary compilers invariably include a limited list of collocates in the definitions and there is an expectation that users will correctly deduce the omitted possibilities which 'etc.' represents in the definition (see Alzi'abi 2016a and 2016b). It is therefore justifiable to expect that users will be able to correctly guess *some* of the adverbial collocates that are not listed in the dictionary and are deemed acceptable in the English language (see Xu 2010).

To achieve the above stated objectives, a set of randomly selected Arab dictionary users will be tasked to look up verbs in the provided dictionaries and thereafter extrapolate verb–adverb collocations. The findings from the study will be stated in Section 4.5 reinforced by some verbal report data of four subjects, other than those in the population sample. The latter data was gathered in order to examine the way in which Arab EFL learners handle collocational information.

4.2 Subjects

The sample for this study included 88 MA students majoring in English at five Jordanian universities consisting of 45 males and 41 females. The age range of the subjects was between 23 and 34. They had been exposed to English on average for more than 16 years prior to their enrolment to the trial, which took place between April and June 2015. A functional prerequisite to enroll on the MA course was to pass a National Proficiency Examination, equivalent to a TOEFL iBT test. Therefore all included subjects' proficiency level may be deemed as advanced.

The subjects were selected on the grounds of their availability and were randomly divided into two groups; one assigned to work with LDOCE (henceforth LDOgr) and the other with OALD (henceforth OALgr). Although the subjects were selected from a homogenous group sharing a similar academic background, nonetheless, further steps were taken to ensure a robust homogeneity between group members. Each subject from the two groups completed Meara's (1992) vocabulary size test (411), i.e. test No. 11 at level four. The range of proficiency level for the two groups was found to be almost identical: LDOgr (73.29, Sd 3.10) compared to (72.97, Sd 3.18) for OALgr. A t-test showed no significant difference between the two groups ($t=.456$, $p<.560$ with 80 df).

The intention was to exclude any extreme scores, i.e. very high or very low. Thus, the data collected from six subjects (all males), was discarded on the basis of their scores on the initial vocabulary test being below 70. Subsequently, the total number of included subjects in the study was 82, i.e. 41 subjects per group.

4.3 Materials

Twenty-two lexical verbs with a wide range of collocability, 12 high-frequency items and 10 low-frequency items, were selected as the vehicle of this research. The first category, the frequent verbs, was first selected in accordance with CALDⁱⁱ classification of frequent headwords which comprised three groups — four verbs in each group. These included 'Essential': basic common words considered important to learners (improve, increase, mention and watch); 'Improver': words to help improve beyond basic English (declare, gain, organise and slow) and 'Advanced': words to enable learners to articulate English at a more sophisticated level (defeat, deserve, oppose and pause). This classification approximately coincides with that of MEDAL, which may further be divided into three categories: 'very high frequency', 'high frequency' and 'quite high frequency'. It is necessary to add that LDOCE classification was deliberately not adopted in order to avoid any potential confusion. A verb such as *improve* was labeled *S2 W1*, i.e. belonging to the two thousand most frequent words in spoken English and the first thousand words most frequently found in written English.

A distinctive characteristic of the stimuli in the aforementioned category was the richness of potential adverb collocates for each verb — a minimum of ten adverb collocates each in OCD. This was found to be the average number for a good set of frequent verbs and was determined following a thorough search. A random set of twelve frequent verbs, with at least ten adverb collocates each, was selected out of the frequent verbs previously chosen. It should be noted that many frequent verbs had no adverb collocates, at least not in OCD, e.g. 'seem', 'remain', etc. Some were assigned only one adverb, e.g. 'sell'; two adverbs, e.g. 'tend'; three adverbs, e.g. 'ban'; four adverbs, e.g. 'suppose' or five adverbs, e.g. 'spend'. However, some frequent verbs were not at all included in OCD, e.g. 'steal'.

The second category included ten verbs. All were infrequent items in accordance with the categorisation of learner's dictionaries earlier outlined. None were found on the *Academic Word List* of 5000 words. The verbs were *flinch*, *halve*, *impair*, *knot*, *pedal*, *relish*, *resent*, *retort*, *snort* and *worsen*. They had been randomly selected on the grounds that they were included in OCD and occurred with at least seven adverbs each. This number was found to be a common feature among many comparable verbs. It was again selected following a close examination of the adverb collocates of all infrequent verbs in OCD.

It must be remarked that although many other infrequent verbs exist in

English, nonetheless, such verbs have not been included in OCD, e.g. 'abstain', 'surmise', etc. The tendency was to select verbs that had the same number of adverb collocates as above, i.e. a minimum of ten adverbs; however, the least frequent verbs were mostly found to have fewer adverb collocates. It is also noteworthy that infrequent verbs with ten or more adverb collocates were seldom found. Evidently that low-frequency verbs tend to have fewer adverb collocates than the high frequency items.

Some stimuli were polysemous (e.g. *declare*, *halve*, *organise*, etc.) or homonymous (e.g. *defeat*, *gain*, *snort*, etc.). Whenever this was found to be the case, the procedure was to ascertain which adverb collocates in OCD associated with the sense of the stimulus under consideration. For example, in the case of *pause*, the sense selected was 'stop speaking', the one with which all the adverbs listed in OCD could be used most often. However, none of the stimuli was assigned two senses in OCD; only the verbs *snort*, *pedal* and *halve* were assigned two or more senses each in LDOCE and OALD and the verb *knot* was assigned three senses in both. Remarkably, the two dictionaries had followed approximately the same policy in entering and ordering the senses of multiple meaning words — at least for the first, and possibly, the most common sense. In totality, they were found to coincide in all cases, save for *deserve*, *flinch* and *mention*, where LDOCE included multiple senses and OALD contained only a sole sense for each.

It has to be mentioned that OCD usually lists sets of adverb collocates which occur with all senses of the verb in question. When the various senses of a certain verb collocate with particular adverbs, separate sets are provided, each occurring with one of the relevant distinct senses. In fact many verbs in OCD, e.g. 'abandon', 'absorb', 'appreciate', 'beat', etc. had more than one distinct sense where an entirely different set of adverbs was supplied for each sense. Being a specialised dictionary, OCD usually lists a larger number of collocates than learner's dictionaries. Sometimes the verbs which had a good number of adverb collocates in OCD were allocated only one or two adverb collocates in the learner's dictionaries, e.g. the verb 'store'. A randomly selected item, 'state', had some 35 adverb collocates in OCD compared to five in LDOCE, and seven in OALD. In very few cases, these adverbs were approximately the same — seven out of ten adverb collocates of 'regret' (v) in OCD also 'exited' in LDOCE.

One more factor that played a role in the selection of the stimuli was the existence of a minimum of three more adverb collocates for each verb, which were not included in any of the dictionaries used. Lastly, for all verbs, adverbial particle collocates such as 'back', 'in', 'off', 'on', 'onwards', etc. were excluded as these might confuse the subjects, with the exception of 'away', being used with *pedal*.

The stimuli were used in two tests: a pre-test (Test1) and a main test (Test2). Test1 presented the stimuli in a random order and the subjects were requested to provide three adverb collocates for each verb in the box provided. This helped us assess their collocational knowledge in advance of dictionary

use. Test2 required them to perform the same task as above but this time dictionary aid was allowed. They were required to look up verbs and read avidly all the information offered to them. They were then asked to write three adverb collocates of the target verb, i.e. three adverbs other than those in the dictionary. The aim of this was to ascertain that subjects were able to provide extra adverb collocates of the verb based on consultation of dictionary information.

Two forms were designed for Test2 — one for LDOgr, viz. subjects working with LDOCE and another for OALgr, subjects working with OALD. Each form contained a list of the stimuli with the assigned senses, in case these were multimeaning verbs, e.g. *pedal1*: the first sense of *pedal*, *snort1*: the first sense of *snort*, etc. The subjects were given a desktop or a laptop with either OALD or LDOCE ready to use. It was ascertained that all subjects could use computers.

4.4 Procedure

There was an orientation session to acquaint the participants with the dictionaries. The researcher explained their distinctive features, innovative search facilities and the way information was presented, particularly collocational information. The subjects' attention was specifically directed to the additional examples and the special sections for collocations.

Prior to taking the tests, the subjects were presented with a plain list of the stimuli, frequent and infrequent in random fashion. The subjects were also required to give the meanings of the verbs, either in L1 or L2, in case they were familiar with them. This question was meant to assess their pre-knowledge of the target verbs and pinpoint the relationship between knowledge of the stimuli and adverb collocates provision. They were expected to be acquainted with the meaning of the high-frequency verbs but not the low-frequency ones. In addition, they were asked to indicate by means of [Yes] or [NO] whether they had received any prior training in using dictionaries. The latter was meant to determine whether instruction in dictionary usage could be a contributing factor to their performance.

Test1 was the initial test followed by Test2. In Test1, the subjects were asked to write down three adverb collocates of the target verbs relying on their own knowledge of the items. Additional instructions and examples were provided in writing. The two forms of Test1 were randomly distributed to the subjects who were granted access to the dictionary corresponding to the form they had received. In Test2, the subjects were asked to look up the verbs, examine the definitions and examples with the intention of providing three extra adverb collocates of the target verbs. The tests were administered during class hours. The researcher remained ready to help and deal with any queries raised during the test. They were tested in small groups due to limited availability of computers. For practical reasons, the time required to perform the task was not recorded but in most cases it was less than one hour for both tests. The subjects were informed that the tests were intended for research purposes

and they were briefed on the explicit aim of the study at a later date.

With regard to data analysis, a comprehensive collocations list (henceforth CCL) was compiled and comprised all the adverbs frequently collocating with the stimuli. Over 420 adverbs were included and served as a reference for checking the responses. The collocates of each stimulus varied in number, ranging from seven items for infrequent verbs such as *pedal*, *relish* and *resent* to thirty for frequent verbs such as *oppose* and *organise*. CCL was drawn up from several collocations dictionaries, both printed and electronic versions. These included OCD, *The BBI Dictionary of English Word Combinations*, the *Free Online Collocations Dictionary*, the *Macmillan Collocations Dictionary for Learners of English* and the *LTP Dictionary of Selected Collocations*. Even though *The BBI Dictionary* did not contain many adverb collocates of the stimuli, it comprised a handful of collocates relating to *increase*, *improve*, *oppose* and *retort*, which were missing from OCD. There was no need to elicit any other collocates from native speakers to avoid any controversial items.

The first objective of this study was to find out whether the subjects could provide additional appropriate adverb collocates of verbs following dictionary consultation. To begin analysing the data, the mean scores for correct answers in Test1 and Test2 had to be computed. The responses, about 5000 for each test, were all checked against CCL and marked. It was necessary to opt for a unified maximum score for all tests including knowledge of meaning. This made it easier to correlate scores and provided results, which could be easily interpreted. Number 10 was deemed appropriate for this very purpose. Consequently, each correct response was awarded $1 \times 10 / 66$, where 10 represented the maximum score possible and 66 stood for the total number of collocates of the stimuli, i.e. three for each verb. For example, if a subject got 30 correct responses, the mark was $30 \times 10 / 66 = 4.54$ points out of ten. Spelling and grammatical mistakes such as *'fastly' instead of 'fast' or *'steadly' instead of 'steadily', were ignoredⁱⁱⁱ.

It goes without saying that some subjects left blank spaces, most likely because adding new collocates was an extremely challenging task and these blanks were given nil. An answer was considered correct if it was on CCL, provided it did not exist in the particular dictionary used by the responder in the case of Test2. There was no need to forward the responses to native English judges to determine whether the responses were appropriate since CCL comprised collocates provided by renowned lexicographers who had exerted commendable efforts to compile the material used for drawing up CCL. Nonetheless, in very few cases where the responses were not found on CCL, they were checked against the *British National Corpus*. However, none appeared to be a well formed adverb collocates of the stimuli; that is, they did not occur in five different texts in the corpus (see Wang and Shaw 2008).

Regarding the pre-test, the objective was to assess the subjects' ability to provide appropriate collocates for the stimuli prior to dictionary use. Comparing the mean scores for the dictionary-based assessment (Test2) and the dic-

tionary-free assessment (Test1) would establish whether the dictionary use had any effect on their performance. Therefore, it was essential that the marking process exclude any identical answers on both tests. A close comparison of the individual responses to both tests showed no duplicate collocates. This suggests the subjects were answering by extrapolating, though incorrectly in many cases, from dictionary information or leaving it blank.

The second issue concerned the subjects' performance with regard to individual dictionaries. The primary aim was to determine which dictionary aided the subjects to perform better. The amount of help dictionary users received from each dictionary was assessable by comparing the mean scores for the two groups on Test2. Submitting these scores to statistical tests revealed if there were any significant difference between the groups and whether a particular group excelled at any of the two types of item. The analysis of the responses with regard to individual entries along with the verbal report data (see Section 4.5) could help display part of the subjects' behaviour and unveil their look-up strategies as to handling collocational information. Likewise, it enabled us to explore the way dictionaries aided or possibly inhibited the subjects' performance in producing collocates. This exercise revealed some underlying causes for subjects' mal-production of collocations.

The third issue was to find out whether knowledge of meaning of the stimuli and dictionary training on the part of the subjects had any proportionate impact on their performance. A total of 1804 responses were marked, 902 for each group to evaluate the subjects' pre-existing knowledge of the meaning of the stimuli. As indicated above, it was found useful that a unified average score would be used for all tests. The total score for the knowledge of meaning test was ten. Each time the meaning of the target verb was correctly expressed, a point of $1 \times 10 / 22$ is awarded, where 22 represented the number of the stimuli. For example, if the subject correctly provided the meanings of 13 verbs, the mark would be $13 \times 10 / 22 = 5.90$. Regarding dictionary training, the data showed that only about 55% (45 subjects) had received some kind of dictionary training. Analysing the scores for knowledge of meaning and dictionary training along with Test2 scores with the appropriate statistical tests revealed the potential effect for these variables on the subjects' performance.

As indicated above, four subjects, other than those mentioned in Section 4.2, accepted to tape-record their thoughts and their decision-making processes during the actual completion of the task. A short practice session with verbs other than the target ones was conducted to familiarise them with the task. The participants were requested to verbally report their cognitive processes while performing the task; this research method should be specified as 'think-aloud protocol'. The verbalisation of their thought process could shed light on the cognitive processes involved in their look-up operations and unveil the strategies employed when utilising the dictionaries.

Before stating the results, it must be underlined that the tests employed in this study can never be wholly accurate. Some external and unaccounted for

factors may have affected the reliability of the results including the subjects' misinterpretation of the meaning of the verbs and poor concentration. The results below must therefore be interpreted with these limitations in mind.

4.5 Results

The first objective of this study related to the subjects' ability to produce adverb collocates of verbs upon dictionary look-ups of the target verbs. To confirm this issue it was necessary first to obtain the scores for responses in Test1, i.e. prior to dictionary consultation as displayed in Table 1.

Table 1: Mean production of collocates per item type in Test1 (max. 10)

	frequent items	infrequent items	total score
Mean	1.00	.639	.626
Sd	.765	.490	.500

Based on the evidence presented, the subjects overall performance from both groups and with both types of item was poor. Their respective scores were well below the chance performance and only less than 10% of the responses were correct. This confirms our earlier hypothesis; particularly in the case of low-frequency items, bearing in mind it was a dictionary-free task. Subjects' poor performances could be attributable to the difficulty of the required task. Notwithstanding the very low scores, as shown above, the subjects performed better, however, not to the expected standard, with the highly frequent items. A t-test showed a highly significant difference between the scores for the frequent and infrequent items ($t= 11.86$, $p= 0.000$, with 80 df). As hypothesised earlier, this outcome is neither illogical nor unexpected, since most subjects indicated they had not learned the meanings of the least frequent items and therefore their responses were mostly presumptive.

To accomplish the principal aim of this study, the scores for responses to Test2 were obtained as shown in table 2 below.

Table 2: Mean production of collocates per item type in Test2 (max. 10)

	frequent items	infrequent items	total score
Mean	1.00	.946	.908
Sd	.664	.542	.506

These figures clearly demonstrate the subjects' poor performance despite using the dictionary. Only about 10% of the responses were appropriate. The scores

were all well below an average level and this was evident in both types of item despite the use of the dictionary. The subjects encountered great difficulty with both types of item. Consequently, they performed unsatisfactorily in most cases. The above figures indicate that the participants faced equal difficulty with both types of item. Minimal improvement was achieved on the overall score in the dictionary-based task. However, this does not contrast with the Test1 scores. A t-test showed a significant difference between the scores for the low-frequency and high-frequency verbs in Test2 ($t= 2.73, p= 0.008$, with 80 df). This suggests that the subjects were performing better with the frequent items and possibly experiencing significantly more difficulty with the infrequent verbs. This confirms the earlier result where subjects performed better with the frequent items in Test1.

A cursory look at the figures in Table 1 and Table 2 shows only little difference between the overall performances in both tests. However, a t-test revealed a significant difference between them ($t= 3.777, p= 0.000$, with 80 df). This was evident in the slightly higher score for Test2, i.e. improvement by 0.282.

Some difference was found between the scores for the low-frequency items in Test1 and Test2, i.e. 0.325, which implies that the subjects who utilised the dictionaries performed slightly better. Two more t-tests were conducted to identify any statistically significant effect for dictionary use on the subjects' performance per item type in both tests. No significant difference was spotted between their scores for the high-frequency items in Test1 and Test2 ($t= 0.03, p= 0.97$) but the difference was highly significant in the case of the low-frequency items ($t= 8.88, p= 0.00$). Their score for the least frequent items as a result of dictionary use constituted some improvement, though surprisingly moderate. Apparently, the subjects found more help in the dictionary with low-frequency items. Notwithstanding, this additional help did not produce a major difference in terms of the overall scores — the amount of gain brought about being comparatively small and modest.

The second objective was to check whether LDOCE and OALD varied in the amount of help they had offered Arab EFL learners to produce verb–adverb collocations. To come up with some evidence in support of this issue, the data in Table 2 was broken down to provide a much more comprehensive picture of the subjects' performance as shown in Table 3.

Table 3: Means collocate production per item type and group (max. 10)

	frequent items		infrequent items		total score	
	Mean	Sd	Mean	Sd	mean	Sd
LDOgr	0.89	0.62	0.88	0.49	0.83	0.46
OALgr	1.12	0.69	1.00	0.58	0.97	0.54

Consistent with the above results, the scores of the two groups are still well below average level and therefore may be regarded as unsatisfactory. OALgr seemed to fare better than LDOgr with both types of item. A t-test was carried out in order to determine if any significant difference exists between the overall performances of the two groups and it was found to be not significant ($t= 1.23$, $p= 0.220$, with 80 df). Two more t-tests showed no significant difference between the two groups for either the high-frequency items ($t= 1.59$, $p= 0.114$) or the low-frequency items ($t= 0.19$, $p= 0.84$). This undoubtedly would rule out any group difference for both types of stimuli, frequent and infrequent verbs. Clearly, no difference existed between LDOCE and OALD relating to extrapolating verb-adverb collocational information.

The third objective of this study was to establish whether the subjects' performance in Test1 showed a relationship between knowledge of the meaning of the stimuli and their dictionary training. Table 4 below presents the subjects' scores for familiarity with meaning.

Table 4: Mean familiarity with the stimuli per item type (max. 10)

	high frequency items	low frequency items	total score
Mean	9.32	2.67	6.30
Sd	1.04	1.88	1.05

As anticipated, the subjects knew the meanings of almost all frequent items but this was not so for the least frequent items. A Pearson correlation test revealed a positive correlation between knowing the meanings of the stimuli and adverb production task on Test1, i.e. no dictionary use ($r= .239$, $p= .031$, with 80df). Knowing the meaning of the target verbs proved to be effective prior to dictionary use and possibly had helped the subjects to perform slightly better with the frequent verbs. This corroborates the hypothesis made earlier that knowing the meaning of items might positively affect the subjects' production of collocations.

As to the influence of dictionary training on the production of adverb collocates, Table 5 demonstrates the two groups' results per dictionary training and item type.

Table 5: Means relating to dictionary training per group and item type

	Training (n=45)				No training (n=37)			
	high frequency items		low frequency items		high frequency items		low frequency items	
	Mean	Sd	mean	Sd	mean	Sd	mean	Sd
OALgr	1.23	.718	1.06	.545	.994	.656	.939	.643
LDOgr	.924	.658	.953	.498	.824	.582	.800	.439

The figures indicate a very slight difference between the scores regarding training in dictionary use and item type. OALgr subjects with dictionary training seemed to have fared better, though modestly, than their counterparts relating to both types of items. A t-test showed no significant difference in both groups between the overall performance of those who had received dictionary training (1.01, Sd .507) and those who had not (.915, Sd .480). Dictionary training was not a positive influencing factor with the subjects' scores. A further t-test showed no significant effect for dictionary training on either group ($t= 0.68$, $p= 0.49$). Neither LDOgr nor OALgr who claimed to have received dictionary training produced more appropriate adverb collocates. It has to be stressed that a Pearson chi-square test showed no significant statistical difference between the number of the subjects with dictionary training and those without in both groups (analysis returned a value of 0.49 and the associated P-value was 0.82). In short, this data refutes the hypothesis that those subjects who received some basic training in dictionary use would perform better. However, this cannot be entirely rejected since many previous studies have provided evidence to the contrary. Dictionary training proved to be effective regarding vocabulary acquisition and collocation competence.

4.6 Discussion

This study was set out to achieve three objectives. The primary of which was to assess Arab dictionary users' uptake of collocational knowledge after dictionary consultation, and whether such uptake varied as to the frequency of the base of the collocation, i.e. the verb. Unlike previous works (e.g. Farghal and Obeidat 1995) where learners were not fully aware of collocations, the Arab EFL subjects in the present study had a good level of awareness of collocations as they avoided the inclusion of the adverb collocates existing in the entries in their answer sheets. Contrastingly, however, they failed to provide more collocates of the base. The data in Table 2 clearly shows that the subjects performed very poorly with regard to both types of item. Only approximately 10% of responses were found to be appropriate, which would be an unimpressive yet not unexpected output. This corroborates with the hypothesis earlier made that the subjects would not likely perform well in this given task. However despite this assumption proving correct, their scores were still much lower than expected. This finding strongly underlines that Arab EFL learners experience great difficulty in producing accurate verb-adverb collocations, and to some extent, the findings are consistent with those of earlier works (e.g. Laufer 2011 and Sun and Wang 2003), which explored EFL learners' difficulties with the use of the various types of collocations. It remains therefore plausible that similar findings could be recorded had the same stimuli and dictionaries been used with subjects of different linguistic backgrounds.

Difficulty in using collocations is not Arab-learner specific. Studies which concerned the use of adverb collocates i.e. 'amplifiers' (e.g. Granger 1998 and

Lorenz 1999), pointed towards French and German EFL learners' underuse of restricted adverb collocates. In a study by Satriawan (2009) a slightly more positive picture was indicated with approximately 75% of the adverb collocates of adjectives, adverbs and verbs of Indonesian subjects were considered to be acceptable. However, Satriawan's focus was on the use of all adverbial collocations rather than verb–adverb expressions specifically. The study also concerned degree adverbs, which would mostly be of open or free collocability.

Interestingly, dictionary use appeared to be statistically significant despite the extremely low scores (cf. Alzi'abi 2012; Dziemianko 2010; Laufer 2011). Dictionary assistance was particularly effective in the case of the infrequent items. Although the underlying reason for this finding remains ambiguous, it may be that the subjects might have benefited from the adverb collocates of equivalent verbs in L1. Nonetheless, no responses indicated cases of literal translation from L1. Arguably, the subjects may have performed better had the items been controlled for congruency between the adverb collocate of the target verbs in English and Arabic. Farghal and Obeidat (1995) found that Arab learners were collocating more correctly when the adverbial collocations in both English and Arabic overlapped and conversely ill-formed structures were produced when the collocations diverged (see also Satriawan 2009 and Yamashita and Jiang 2010).

Several reasons are postulated for the subjects' significant overall low performance. Firstly, the subjects' low exposure to collocational information and training despite their enrolment on advanced level English language courses (cf. Lew and Radłowska 2010). That is to say, they had not developed the basic linguistic competence and requisite skills for extrapolating extra verb–adverb collocations from the dictionary information. At face value, this basic explanation is true. However, the situation is rather more complex when considering that the subjects had succeeded in identifying the adverb collocates of the target verbs (see Alzi'abi 2017) and had avoided the inclusion of these items in their responses. This suggests that their comprehension and production of collocational competencies varied markedly, with much lower scores for production. This in turn reinforces the author's widely held belief and many researchers' empirical findings that EFL learners will significantly perform better at comprehension than production tasks (Alsakran 2011; Hamad and Laohawiriyanon 2013; Hill and Laufer 2003; Jaén, 2007; etc.).

In fact, the subjects who provided the verbal report data were at times astounded by their inability to add any new item. Lamenting their ineffective strategy in learning words, one particular subject remarked in Arabic "I now realise how wrong I have been in learning words out of their appropriate context". The subject also added "It is always important to consider the items that precede and follow any word I learn". This deficiency may also be attributable to the lack of exposure to this particular type of collocation as well as ineffective learning strategies.

Another significant reason that might account for the extremely low scores

obtained was the lack of sufficient collocational clues in the dictionaries provided to enable the subjects to infer correctly what adverbs could co-occur with the target verbs. This assumption is supported by Rundell's (1999: 50) argument that some dictionaries lack the subtlety of information required to enhance users' proficiency. Though it is true that learner's dictionaries have not traditionally been scripted to provide an exhaustive amount of collocational information, the latest editions do incorporate a richer source of collocations in their various forms. Nonetheless, for the EFL learner, there is still a pressing need for the inclusion of more hints as to what other items could occur with the "base" items. Previously, Alzi'abi (2012 and 2016a) found that a paucity of collocational clues existed in traditional dictionaries, namely noun collocates of verbs. Dictionary compilers habitually surrender to the oft-repeated use of "etc." to indicate the remaining list of collocates, thus leaving it to EFL learners to complete the often insurmountable task of completing the list through guesswork. Conversely, this contrasts the finding from Lew and Radłowska (2010), whose subjects found most of the collocational information provided in LDOCE. This particular finding could be attributable to their inclusion of much fewer stimuli, i.e. 13 items representing nine types of collocations. It is also likely that these items were each assigned greater amounts of collocational information.

A third potential reason identified was the difficulty of the test employed in this study. Evidently, requesting the subjects to provide three adverb-collocates proved tasking. In many cases, the subjects left blank spaces due to their inability to provide any adverb collocates of the given stimuli. Their limited vocabulary likely precluded the production of correct collocates, bearing in mind that some of the possible adverb collocates of the stimuli were infrequent items, e.g. 'in passing', 'stoutly', 'solemnly', 'adamantly', 'strenuously', 'vehemently' and the like. This problem was also encountered by Lithuanian EFL learners (Juknevičienė 2008) where lack of complete mastery of academic vocabulary had led to deviant noun collocates of frequent verbs. Not surprisingly, the analysis of the erroneous responses indicated that some only loosely-associated with the target verbs and mostly included highly frequent adverbs, e.g. 'nicely', 'quickly' etc. Such adverbs should not be considered entirely inappropriate. They may well be labeled as all-verb adverbs, but not the type that native speakers would habitually collocate with the verbs under consideration, i.e. the type Sinclair (1987) called random 'co-occurrences' (see also Siyanova and Schmitt 2008). Heidler (2011) found that international undergraduates' use of adverbs was correct from a grammatical standpoint but it was not in conformity with what native speakers would otherwise use. The think-aloud subjects were frequently found to stumble over the target verbs and provided nothing that could be deemed acceptable.

Closely associated with the factor of test difficulty was the unpredictability of collocations. The way some adverbs collocated with verbs might have bewildered the subjects. For example, why could something *improve* 'slowly',

'rapidly' rather than 'quickly'; *worsen* 'rapidly' but not 'quickly'; *knot* 'securely' not 'safely'? Likewise, why 'fully' and 'thoroughly' rather than 'completely' could modify and collocate with '*deserve*' and 'specifically' rather than 'particularly' with *mention*? Such unacceptable collocates along with many others were cited by the subjects.

Analysis of the findings from the think-aloud protocols showed the subjects were keen to extract information anywhere in the explanation to enable them to answer correctly. They frequently read the explanation more than once in the hope of inferring the adverb collocates of the target verb. It is likely that all other subjects did the same. However, it is possible that some subjects were unable to handle the information or decipher the meaning of some infrequent verbs and, as a result, had failed to provide any correct responses. Rundell (1999) points out that dictionary look-ups are not always straightforward operations. These require taught skills as well as due diligence on the part of the user. If not performed efficiently, instead of facilitating the user, dictionary look-ups may lead to erroneous inferences. Some subjects might have understood *impair* to possibly have a positive connotation and provided inaccurate collocates such as 'proudly' and 'successfully', thus giving a score of 0.60. This may clarify why the scores for some low-frequency verbs were extremely low, e.g. *flinch* (0.16), *relish* (0.28), *resent* (0.24), etc. When the subjects were able to grasp the meaning of the low-frequency verb, the results were relatively better, e.g. *pedal* (1.9) and *snort* (1.7). It is worth noting here that the best score correlated to *pedal* whereas the lowest score was for *flinch* and *halve*, each with a score of 0.0. Incongruously, the score for *improve*, proved to be a unique case where there was a decrease in the score following dictionary consultation. The score in Test1 was 1.74 and following dictionary consultation during Test2 the score decreased to 1.54. The subjects who responded to the verb appropriately in Test1 could not predict other adverb collocates with dictionary assistance. This suggests that dictionary using instruction should be incorporated in any curriculum to adequately prepare the users for successful look-up operations.

A close examination of the responses including the verbal report data along with the relevant dictionary definitions unveiled some common strategies adopted by the subjects. First, some collocates were selected on the grounds that they were synonyms or possibly near-synonyms of collocates included in the explanation. Examples of this rather unsuccessful strategy included 'tidily' for 'neatly' at the entry for *organise*; 'badly' for 'severely' at *impair*; 'particularly' for 'specifically' and 'repeatedly' for 'frequently' at *mention*; 'calmly' for 'coolly', 'humorously' and 'ironically' for 'sarcastically' and 'furiously' for 'angrily' at *retort*; 'surely' and 'lastly' for 'certainly' and 'finally', respectively, at *gain*; 'not especially' for 'not particularly' at *relish*; 'calmly' for 'silently' at *watch*; etc. Such occurrences were far more common than could be covered here. More evidence for this strategy emerged from further comments made by the subjects in the concurrent verbal report data. One subject commented "Oh, yes! The best thing to do, I think, and to be on the safe side, is to

look for synonyms of the collocates [the ones in the entry]". Responding to *deserve* in LDOCE, another subject stated: "I am sure 'totally', being a synonym of 'fully', is 100% appropriate".

Another strategy, related to the above, was the use of adverbs, which associated with the verb synonyms of the target verbs. The subjects mistakenly generalised many adverb collocates of verbs synonymous with the target verbs. For example, some subjects cited the adverbs 'clearly' and 'briefly' as responses to *declare*; the two adverbs would be strong collocates of the verb 'state' which was used in the explanation of the verb *declare*. Similarly, the adverbs 'fast', 'massively' and 'suddenly', were given by some as collocates of *increase*. These would collocate with verbs synonymous with *increase*, such as 'expand' and/or 'grow'. Interestingly, this was not limited to Arab learners; Argentinean advanced EFL learners also employed a similar approach. Zinkgräf (2008: 91) ascribed errors in the production of adverb collocates by the subjects to the "semantic overlap between appropriate forms and possible synonyms of either the base or the collocate".

In many a case, the subjects provided an antonym of the adverb collocates found in the dictionary explanation of the relevant verb. For example, responding to *improve* some subjects gave 'indirectly' because the entry contained 'directly'. Sometimes they were coining their own versions of antonymous words ending up with odd items. For example, when presented with *deserve*, they cited 'unrichly'. It is possible that they coined this antonym and retained the original adverb 'richly' to ensure that it would be a correct answer since 'poorly' would be a different word. Further examples included 'unreally' at *deserve* and 'undramatically' at *oppose*, as the entries for these verbs included 'really' and 'dramatically', respectively.

In the same vein, the data also included some peculiar and amusing responses for which no clear underlying reasons could be given. For instance, one subject collocated '*unemploymentally' with *halve*, should such a response be taken seriously? Other responses included 'gastro-intestinally', 'emotionally', 'spiritually' and 'psychologically' to collocate with *impair*; 'presidentially' with *worsen*; 'gazely' with *watch* and so on. It could somehow be understood that 'psychologically' was probably provided as a near-synonym of, or possibly, co-hyponym of 'mentally' and 'visually', respectively. It is difficult to propose an explanation for the rest. One is really astounded by such answers and can offer nothing in explicating them. Other examples included the use of the noun 'access' by two OALgr members as a collocate of *gain*, since this appeared in the explanation. It is again quite difficult to expound on this point. Was it because the subjects were simply confused or because of an underlying failure to come up with more adverb collocates?

One more strategy was the utilisation of the adjectives in the definitions of certain target verbs. The subjects derived adverbs from these items as collocates of the relevant verbs. For instance, the adjective 'illegal' at the entry for *declare* in LDOCE (declare sth illegal...*The war was declared illegal...*) spurred some to

form the adverb 'illegally' and collocate with *declare*. This is a substantial piece of evidence in support of the subjects' scrutiny and exploitation of the information in the definitions. These inferences coupled with the inaccurate perception that the adjectives would be appropriate collocates when changed into adverbs led the subjects to erroneous responses in many cases. In any case, these findings reflected a deficiency in the subjects' vocabulary repertoire. Had they acquired larger vocabulary items, the task might have been easier for them.

In conformity with that which has already been discussed, at times the subjects were found to spot certain collocates of a particular verb in the explanation and generalised them to others. However, it could not be ascertained whether they did so since they assumed these would also collocate with the latter verbs or whether they were just filling the blank with any adverb they had previously come across as a desperate attempt to find collocates to be used with the stimuli. This again indicates the subjects' poor collocational competence. According to earlier studies (e.g. Zughoul and Abdul-Fattah 2003), this behaviour might also be a direct result of ineffective learning and teaching strategies of collocations and vocabulary items in general.

The second issue concerned whether a particular dictionary would prove more helpful in the production of adverb collocates of verbs. Very small differences existed between the overall scores of LDOgr and OALgr, which also held true for both types of item. Although dictionary use is generally considered effective, neither dictionary in this study has proven to be comparatively more useful than the other for providing additional adverb collocates of the target verbs. The scores of OALgr for both types of item were slightly higher than those of LDOgr; however, the differences were not overall significant, which would suggest the two dictionaries did not vary in the amount of help they offered users for either type of stimulus. This contradicts the prediction that LDOCE would be more helpful than OALD owing to the extra examples, which might supply more clues. This might be quite frustrating for the compilers. Alzi'abi (2017) has found that the subjects meticulously researched the examples in their endeavours to spot the adverb collocates of the target verbs.

For the third and final objective of the study, familiarity with the meaning of the verb significantly facilitated the provision of some of its adverb collocates. Evidence to confirm this finding emerged from two sources: Firstly, a remarkable correlation between familiarity with the meaning of the verbs and the production of extra adverb collocates with the dictionary-free task, i.e. Test1; secondly, the statistically significant difference between the scores for the frequent and infrequent items in Test1. The subjects performed better, albeit marginally, with the verbs with which they were previously familiar. Therefore, one is inclined to accept the assumption that pre-knowledge of meaning could improve the subjects' collocational competence. Nonetheless, the low scores certainly do not significantly diminish the importance of this relationship.

Dictionary training on the other hand did not significantly influence the

subjects' performance. OALgr subjects with dictionary training slightly outperformed their LDOgr counterparts but the difference, however, was not significant. This suggests that the hypothesis about dictionary training must be rejected based on the findings from the present study. It is likely that the subjects did not receive proper instruction into the best way to deploy dictionaries of any type. They might have been taught how to utilise bilingual dictionaries or possibly paper dictionaries rather than electronic versions. Further concrete evidence pointing to the latter case came from comments made by some subjects during the dictionary orientation session prior to the tests. It might be true that the training in dictionary usage was ineffective. But what could the well-trained users do with the dictionaries that lacked clear collocational clues to help them achieve better scores? The likely answer might be "very little". This claim had been partially substantiated by the subjects' full search for information and their utilisation of the adverbs and adjectives that existed, though wrongly in some cases, to present adverb collocates of the target verbs.

Generally, this finding about dictionary training concurs with others from previous research where no accurate picture has been established of the relationship between training in dictionary usage and receptive and productive competences. Since some earlier studies have provided contrary evidence, one cannot simply draw a hasty conclusion regarding the usefulness of dictionary training skills for effective dictionary use. Further specified research is required to elaborate upon this matter.

Before concluding, it is important to set forth some of the limitations of the present study. Firstly, the sample size was not large enough to be representative of all advanced Arab EFL learners. Perhaps the inclusion of learners from other Arab countries would have made the findings more generalisable. Secondly, the test was unpredictably difficult. The decision to use a mixture of frequent and more sophisticated verbs in one test was perhaps unwise. Merging these into one test might have had a negative effect and severely hindered the subjects' effective performance. One further important limitation was the absence of any objective criterion for evaluating subjects' pre-knowledge of the stimuli, which had made it challenging to generalise the respective findings.

5. Conclusion and future research

In this research three objectives have been accomplished. Firstly, to determine Arab learners' capacity for the provision of adverb collocates of verbs. The results showed their severe difficulty in providing additional collocates of both frequent and infrequent verbs. Despite this apparent failure, they performed significantly better with dictionary assistance, and particularly so with the infrequent verbs. Several reasons might lie behind their extremely modest achievement including deficiency in the collocational competence, difficulty of the given task and lack of sufficient collocational clues in the dictionary. Their deficient vocabulary stock might also have inhibited their effective perform-

ance. The second objective was to establish which dictionary, if any, gave more help to users. Although dictionary use appeared to be a significant influence on the subjects' performance, the two dictionaries were almost identical in regard to the amount of collocational help they offered the subjects. Notwithstanding the huge number of collocations (147,000) introduced to LDOCE, this information is still far from being sufficient to meet EFL learners' needs. The data did not support the assumption that LDOCE users might excel and outperform those of OALD. Thirdly, to determine whether familiarity with the verbs and dictionary training could positively influence the subjects' collocational performance. There was some clear evidence to substantiate the claim that prior-knowledge of the meaning of verbs might facilitate the provision of extra collocates of target words. This finding confirms the notion that knowledge of a word will likely enhance the acquisition of its collocates. Teachers and material writers may capitalise on this highly important outcome. However, the evidence gathered regarding dictionary training did not confirm the claim that training in dictionary use alone would positively impact the users' performance.

Overall, the present findings are in line with those from earlier studies (e.g. Farghal and Obeidat 1995; Laufer and Waldman 2011; Wang and Shaw 2008) that advanced EFL learners have encountered serious but rather soluble problems in using almost all collocational patterns, verb-adverb collocations being no exception (see Kuo 2009). On this basis, researchers, dictionary compilers and syllabus designers are all invited to work towards a practical long-term solution to dictionary users' serious deficiency in collocation competence.

These findings point to important pedagogical implications. They highlight the importance of explicit teaching of collocations because these unpredictable chunks could hardly be heuristically acquired. The inadequacy of vocabulary teaching methods might lie, to a certain extent, behind the subjects' inefficiency in collocating items correctly. Siyanova and Schmitt (2008: 454) rightly suggest that EFL learners' collocational intuitions can be developed by "instituting a fundamental change in our teaching pedagogies". Particular emphasis should be placed on formulaic expressions, and more importantly, on learners' unawareness of collocational restrictions in the early stages of language learning. Likewise, learners have to have adequate practice in using synonymous items to remain alert to the subtle nuances in such types of words. More importantly, dictionary definitions need to be reformulated in order to facilitate and boost correct inferences of the remaining collocates of the defined item.

This study helps us identify some of the challenges faced by learners' with collocations. Although the present findings may not be Arab dictionary user-specific, there remains a need for further research to be conducted to include other nonnative subjects to ascertain whether identical findings could be recorded using a similar study design. The difficulty of the given task might have aggravated subjects' difficulties with verb-adverb collocations. It would

be interesting to administer the same test to native speakers to compare their production with that of the subjects included in the present study in order to ascertain whether the test was unduly challenging. The finding that the frequent items were not easier to handle than the infrequent ones is surprising and also merits further follow-up research. Lastly, as indicated above, the data contained no signs of subjects producing false collocations because they were attempting to find a one-to-one correspondence between the Arabic collocates and their equivalent translations in English. It is therefore worth conducting a large-scale study to investigate the approaches adopted by Arab EFL learners to handle congruent verb–adverb collocations.

References

A. Dictionaries (All learners' dictionaries are on CD-ROMs)

- Benson, M., E. Benson and R. Ilson (Eds.).** 1997. *The BBI Dictionary of English Word Combinations*. Second revised edition. Amsterdam/Philadelphia: John Benjamins.
- 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*. Ninth Edition. Oxford: OUP. (OALD9)
- Hill, J. and M. Lewis (Ed.).** 1997. *LTP Dictionary of Selected Collocations*. London: LTP.
- Mayor, M. (Ed.).** 2009. *Longman Dictionary of Contemporary English*. Fifth Edition. Harlow: Pearson Education. (LDOCE5)
- McIntosh, C. et al. (Eds.).** 2009. *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press. (OCD)
- Rundell, M. (Ed.).** 2007. *Macmillan English Dictionary for Advanced Learners*. Second Edition. Oxford: Macmillan. (MEDAL2)
- Rundell, M. (Ed.).** 2010. *Macmillan Collocations Dictionary for Learners of English*. Oxford: Oxford University Press.
- Walter, E. (Ed.).** 2008. *Cambridge Advanced Learner's Dictionary*. Third edition. Cambridge: Cambridge University Press. (CALD3)

B. Other literature

- Akpýnar, K. and M. Bardakçý.** 2015. The Effect of Grouping and Presenting Collocations on Retention. *The Electronic Journal for English as a Second Language* 18(4): 1-22. <http://www.tesl-ej.org/wordpress/issues/volume18/ej72/ej72a4/>.
- Al-Amro, M.** 2006. *Saudi Learners' Knowledge and its Relationship to their Vocabulary Size and Writing Quality*. Unpublished Ph.D. Thesis. Colorado: Colorado State University.
- Alfahadi, A., S. Zohairy, M. Momani and M. Wahby.** 2014. Promoting Awareness of Teaching Collocations Techniques to Beginners (Adjective–Noun Collocations). *European Scientific Journal* 10(10): 389-396.
- Alexander, R.J.** 1984. Fixed Expressions in English: Reference Books and the Teacher. *English Language Teaching Journal* 38(2): 127-132.

- Alsakran, R.A.** 2011. *The Productive and Receptive Knowledge of Collocations by Advanced Arabic-speaking ESL/EFL Learners*. Unpublished Ph.D. Thesis. Colorado: Colorado State University.
- Alzi'abi, S.E.** 2012. On Arab EFL Learners' Use of Collocational Information in Electronic Dictionaries. *Damascus University Journal* 28(3+4): 63-99.
- Alzi'abi, S.E.** 2016a. Help or Hindrance: "Etc" in Learner's Dictionaries? A Study of Verb–noun Collocations. *European Journal of Applied Linguistics* 4(2): 301-327.
- Alzi'abi, S.E.** 2016b. Etc. The Long-Lasting Defining Device: Unravelling the Mystery. *Lexikos* 26: 36-59.
- Alzi'abi, S.E.** 2017. On Findability of Verb–Adverb Collocations in Electronic Dictionaries. *Indian Journal of Applied Linguistics* 43(1-2): 88-127.
- Ang, L.H., H.A. Rahim, K.H. Tan and K. Salehuddin.** 2011. Collocations in Malaysian English Learners' Writing: A Corpus-based Error Analysis. *The Southeast Asian Journal of English Language Studies* 17: 31-44.
- Bahns, J.** 1993. Lexical Collocations: A Contrastive View. *ELT Journal* 47(1): 56-63.
- Busse, U.** 1995. *Kollokationen*. Ahrens, R., W-D. Bald and W. Hüllen (Eds.). 1995. *Handbuch Englisch als Fremdsprache*: 123-128. Berlin: Erich Schmidt.
- Dziemianko, A.** 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.
- Fan, M.** 2009. An Exploratory Study of Collocational Use by ESL Students — A Task Based Approach. *System* 37(1): 110-123. <http://dx.doi.org/10.1016/j.system.2008.06.004>.
- Farghal, M. and H. Obeidat.** 1995. Collocations: A Neglected Variable in EFL. *International Review of Applied Linguistics in Language Teaching* 33(4): 315-331.
- Farghal, M. and M. Al-Hamly.** 2007. Lexical Collocations in EFL Writing. *The Journal of Asia TEFL* 4(2): 69-94.
- Farrokh, P.** 2012. Raising Awareness of Collocation in ESL/EFL Classrooms. *Journal of Studies in Education* 2(3): 55-74.
- Granger, S.** 1998. Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. Cowie, A.P. (Ed.). 1998. *Phraseology: Theory, Analysis and Applications*: 145-160. Oxford: Oxford University Press.
- Gries, S.** 2008. Phraseology and Linguistic Theory: A Brief Survey. Granger, S. and F. Meunier (Eds.). 2008. *Phraseology: An Interdisciplinary Perspective*: 3-25. Amsterdam: John Benjamins.
- Hamad, M. and C. Laohawiriyanon.** 2013. Effects of Online Dictionary Use and Intensive Involvement Tasks on Learning Verb–Noun Collocations. *Proceedings of the International Conference on Humanities and Social Sciences, April 27th, 2013, Faculty of Liberal Arts, Prince of Songkla University*: 325-337. <http://tar.thailis.or.th/handle/123456789/785>.
- Hill, M. and B. Laufer.** 2003. Type of Task, Time-on-task and Electronic Dictionaries in Incidental Vocabulary Acquisition. *IRAL* 41(2): 87-106.
- Heidler, L.** 2011. *NNS Use of Adverbs in Academic Writing*. Unpublished M.A. Dissertation. Denton, Texas: University of North Texas.
- Howarth, P.** 1998. Phraseology and Second Language Proficiency. *Applied Linguistics* 19(1): 24-44.
- Hussein, R.** 1990. Collocations: The Missing Link in Vocabulary Acquisition amongst EFL Learners. *Papers and Studies in Contrastive Linguistics* 26: 123-136.
- Jaén, M.M.** 2007. A Corpus-driven Design of a Test for Assessing the ESL Collocational Competence of University Students. *International Journal of English Studies* 7(2): 127-147.

- Juknevičienė, R.** 2008. Collocations with High-frequency Verbs in Learner English: Lithuanian Learners vs. Native Speakers. *Kalbotyra* 59(3): 119-127.
- Kennedy, G.** 2003. Amplifier Collocations in the British National Corpus: Implications for English Language Teaching. *TESOL Quarterly* 37(3): 467-487.
- Krenn, B. and S. Evert.** 2001. Can We Do Better than Frequency? A Case Study on Extracting PP-Verb Collocations. *Proceedings of the ACL Workshop on Collocation: Computational Extraction, Analysis and Exploitation, July 2001, Toulouse, France*: 39-46.
- Kuo, C.-L.** 2009. An Analysis of the Use of Collocation by Intermediate EFL College Students in Taiwan. *ARECLS* 6: 141-155.
- Kurosaki, S.** 2012. *An Analysis of the Knowledge and Use of English Collocations by French and Japanese Learners*. Ph.D. Thesis. London: University of London.
- Laufer, B.** 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, B. and T. Waldman.** 2011. Verb–Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61(2): 647-672.
- Lew, R. and M. Radłowska.** 2010. Navigating Dictionary Space: The Findability of English Collocations in a General Learner's Dictionary (LDOCE4) and Special-purpose Dictionary of Collocations (OCD). Ciuk, A. and K. Molek-Kozakowska (Eds.). 2010. *Exploring Space: Spatial Notions in Cultural, Literary and Language Studies. Volume 2: Space in Language Studies*: 34-47. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lewis, M.** 2000. Language in the Lexical Approach. Lewis, Michael (Ed.). 2000. *Teaching Collocation: Further Developments in the Lexical Approach*: 126-154. Hove, England: Language Teaching Publications.
- Lorenz, G.** 1999. *Adjective Intensification — Learners versus Native Speakers: A Corpus Study of Argumentative Writing*. Amsterdam: Rodopi.
- Meara, P.** 1992. *EFL Vocabulary Tests*. Swansea: Centre for Applied Language Studies.
- Moon, R.** 1997. Vocabulary Connections: Multi-word Items in English. Schmitt, N. and M. McCarthy (Eds.). 1997. *Vocabulary: Description, Acquisition and Pedagogy*: 40-63. Cambridge: Cambridge University Press.
- Murnani, Z.T. and H. Salehi.** 2016. Impact of Using Electronic Dictionary on Collocation Learning and Retention of Iranian EFL Learners. *International Journal of English Language and Literature Studies* 5(1): 55-65.
- Nesselhauf, N.** 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24(2): 223-242.
- Nesselhauf, N.** 2005. *Collocations in a Learner Corpus*. Amsterdam/Philadelphia: Benjamins.
- Nizonkiza, D.** 2011. The Relationship between Lexical Competence, Collocational Competence, and Second Language Proficiency. *English Text Construction* 4(1): 113-146.
- Nizonkiza, D., T. van Dyk and H. Louw.** 2013. First-year University Students' Productive Knowledge of Collocations. *Stellenbosch Papers in Linguistics Plus* 42: 165-181.
- Nguyen, T.M.H. and S. Webb.** 2016. Examining Second Language Receptive Knowledge of Collocation and Factors that Affect Learning. *Language Teaching Research*. doi: 10.1177/1362168816639619.
- Peters, E.** 2016. The Learning Burden of Collocations: The Role of Interlexical and Intralexical Factors. *Language Teaching Research* 20(1): 113-138.

- Philip, G.** 2007. Decomposition and Delexicalisation in Learners' Collocational (Mis)behaviour. Davies, M., P. Rayson, S. Hunston and P. Danielsson (Eds.). 2007. *Online Proceedings of the Corpus Linguistics Conference, July 27th–30th*: 1-11. Birmingham: University of Birmingham. http://ucrel.lancs.ac.uk/publications/CL2007/paper/170_Paper.pdf.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-53.
- Satriawan, A.** 2009. *A Comparison between NNS and NS Adverbs in their Adverbial Collocations*. <http://sastra.um.ac.id/wp-content/uploads/2010/01/089-Archangela-Yenny-FKIP-UAJ-EFL-Learners-.pdf>.
- Seretan, V.** 2011. *Syntax-based Collocation Extraction, Text, Speech and Language Technology*. Springer: Dordrecht.
- Sinclair, J.** 1987. The Nature of the Evidence. Sinclair, John (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 150-159. London: Collins ELT.
- Siyanova, A. and N. Schmitt.** 2008. L2 Learner Production and Processing of Collocation: A Multi-study Perspective. *Canadian Modern Language Review* 64(3): 429-458.
- Smadja, F.** 1993. Retrieving Collocations From Text: Xtract. *Computational Linguistics* 19(1): 143-177.
- Sun, Y.-C. and L.-Y. Wang.** 2003. Concordancers in the EFL Classroom: Cognitive Approaches and Collocation Difficulty. *Computer Assisted Language Learning* 16(1): 83-94.
- Szudarski, P.** 2012. Effects of Meaning- and Form-focused Instruction on the Acquisition of Verb-noun Collocations in L2 English. *Journal of Second Language Teaching and Research* 1(2): 3-37.
- Wang, Y. and P. Shaw.** 2008. Transfer and Universality: Collocation Use in Advanced Chinese and Swedish Learner English. *ICAME Journal* 32: 201-232.
- Weinert, R.** 1995. The Role of Formulaic Language in Second Language Acquisition: A Review. *Applied Linguistics* 16(2): 180-205.
- Williams, G.** 1998. Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles. *International Journal of Corpus Linguistics* 3(1): 151-171.
- Wray, A.** 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Xu, X.** 2010. Study on the Effect of Dictionary Use on Second Language Incidental Vocabulary Acquisition: An Empirical Study of College English Vocabulary Learning Strategy. *Journal of Language Teaching and Research* 1(4): 519-523.
- Yamashita, J. and N. Jiang.** 2010. L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly* 44(4): 647-668.
- Zinkgräf, M.** 2008. V+N Miscollocations in the Written Production of University Level Students. *ELIA: Estudios de lingüística inglesa aplicada* 8: 91-116.
- Zughoul, M. and H. Abdul-Fattah.** 2003. Translational Collocational Strategies of Arab Learners of English: A Study in Lexical Semantics. *Babel* 49(1): 59-81.

A Usability Evaluation of the Prototype *Afrikaanse idiome-woordeboek**

Liezl H. Ball, *Department of Information Science, University of Pretoria,
South Africa (liezl.ball@up.ac.za)*

and

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

Abstract: The *Afrikaanse idiome-woordeboek* is a prototype e-dictionary of Afrikaans fixed expressions developed with the intention to test the functionalities of the e-dictionary. This dictionary is based on the function theory of lexicography. The e-dictionary makes use of various technologies. When digital tools are developed it is important to consider the usability of the tool. Usability evaluation was done on the *Afrikaanse idiome-woordeboek* to determine with what success it can be used. Discount usability methods, viz. heuristic evaluation and usability testing were used. This article reports on the findings from the usability tests which are discussed under the categories of content, information architecture, navigation, access (searching and browsing), help, customisation and the use of innovative technologies to manage data in e-dictionaries for search and display. The usability evaluation showed that the users did not always use the e-dictionary as the designers intended. Various recommendations are made to the designers of the *Afrikaanse idiome-woordeboek*, as well as for the design of e-dictionaries in general. Recommendations appropriate to e-dictionaries in general are made regarding usability evaluation, information architecture, searching in e-dictionaries, the data that can be included in e-dictionaries and training of users of e-dictionaries.

Keywords: E-DICTIONARIES, ONLINE DICTIONARIES, ELECTRONIC DICTIONARIES, USABILITY EVALUATION, USABILITY TESTS, DISCOUNT USABILITY, DICTIONARY LITERACY

Opsomming: 'n Bruikbaarheidsevaluering van die prototipe van die *Afrikaanse idiome-woordeboek*. Die *Afrikaanse idiome-woordeboek* is 'n prototipe e-woordeboek van Afrikaanse vaste uitdrukkings wat ontwikkel is met die doel om die funksionaliteit van die e-woordeboek te toets. Die woordeboek is gebaseer op die funksieteorie van leksikografie. Die e-woordeboek maak gebruik van verskeie tegnologieë. Wanneer digitale werktuie ontwikkel word is dit belangrik om die bruikbaarheid van die tuig te oorweeg. Bruikbaarheidsevaluering is op die *Afrikaanse idiome-woordeboek* gedoen om te bepaal met watter mate van sukses dit gebruik kan

* An shortened version of this article was presented at the Twenty-second Annual International Conference of the African Association for Lexicography (AFRILEX), hosted by the School of Languages and Literatures: African Language Studies Section, Rhodes University, Grahamstown, South Africa, 26–29 June 2017, while the original went through the editorial process for publication in *Lexikos*.

word. Die metodes van afslag-bruikbaarheidsevaluering, naamlik heuristiese evaluering en bruikbaarheidstoetsing is gebruik. Hierdie artikel doen verslag oor die bevindinge van die bruikbaarheidstoetsing en dit word bespreek onder die kategorieë inhoud, inligtingargitektuur, navigasie, toegang (soek en snuffel), hulp, aanpassing en die gebruik van innoverende tegnologieë om data in e-woordeboeke te bestuur vir soek- en vertoondoeleindes. Die bruikbaarheidsevaluering het gewys dat die gebruikers nie altyd die e-woordeboek gebruik het soos die ontwerpers bedoel het nie. Verskeie aanbevelings is gemaak vir die ontwerpers van die *Afrikaanse idioom-woordeboek*, so wel as vir die ontwerp van e-woordeboeke in die algemeen. Aanbevelings toepaslik vir e-woordeboeke in die algemeen is ook gemaak oor bruikbaarheidsevaluering, inligtingargitektuur, soek in e-woordeboeke, die data wat in e-woordeboeke ingesluit kan word en die opleiding van gebruikers van e-woordeboeke.

Slutelwoorde: E-WOORDEBOEKE, AANLYN WOORDEBOEKE, ELEKTRONIESE WOORDEBOEKE, BRUIKBAARHEIDSEVALUERING, BRUIKBAARHEIDSTOETSING, AFSLAG-BRUIKBAARHEIDSEVALUERING, WOORDEBOEKGELETTERHEID

1. Introduction

There are many exciting opportunities that technology brings to the field of lexicography. In the first place, much more data can be included in an e-dictionary. This has many advantages; for instance, e-dictionaries can include or link to more data (De Schryver 2003: 157), multimedia can be incorporated (Lew 2012: 344) or words can be written out and do not need to be abbreviated (De Schryver 2003: 157). Information technology also offers many advantages in terms of access to data; for instance, the speed with which data can be retrieved is a considerable advantage (Verlinde and Peeters 2012: 147) and various search features can be included to enable more efficient search (Lew 2012: 345, 351; Verlinde and Peeters 2012: 147). Bothma (2011) also suggests various modern technologies that could be used to enhance e-dictionaries, such as annotations, linked data, filtering and recommendations.

Technology can also allow e-dictionaries to be customised to the extent that it only gives information that is relevant in a specific situation. For example, a person who is writing a text needs different information than a person who is reading or a person who needs detailed background information about an item. This idea is formalised in the function theory of lexicography (see, for example, Bergenholtz 2011; Bergenholtz and Bergenholtz 2011; Bergenholtz, Bothma and Gouws 2015; Bergenholtz and Gouws 2007; Bergenholtz and Tarp 2003; Nielsen 2011; Tarp 2007, 2008, 2011).

When digital tools are developed, it is vital that these tools can be used effectively and efficiently by users; in other words, the usability of a tool is important. Usability can be seen as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO 9241-11 1998). Usability becomes more important as products become more complex and can be critical to the success of a product (Tullis and Albert 2008: 5-7). Usability evaluation is the process where data about how users will use or do use a product are gathered

and whether it is suitable and acceptable to users (Preece, Rogers and Sharp 2011: 433). Different methods can be used to do a usability evaluation, such as usability testing, inspection methods, inquiry methods, analytical modelling or simulations (Fernandez, Insfran and Abrahão 2011: 796-797; Shneiderman and Plaisant 2010: 134-135, 138-149). The usability evaluation method must be applicable to the product and the environment that the product will be used in; for example, products that will be employed in dangerous environments should ultimately be tested in real-life settings, not only in laboratories (Shneiderman and Plaisant 2010: 133).

Discount usability is an approach where simplified methods are used to evaluate a product (Nielsen 1993: 17; Nielsen 2009). Nielsen (2009) suggests that excellent results can be obtained if only a few participants are used in usability testing. About five participants are enough to point out the main usability problems (Nielsen 2000; 2009; 2012a). Furthermore, in discount usability, experts are used to conduct heuristic evaluations and prototypes, rather than complete systems are evaluated (Nielsen 2009). It has also been argued that sophisticated facilities are not needed (Preece, Rogers and Sharp 2011: 478; Nielsen 2012b).

There are many different aspects that a usability evaluation can assess. Some of the aspects that can be evaluated are navigation (e.g. Calisir et al. 2010), the structure or organisation of information (e.g. Hasan, Morris and Proberts 2012), aesthetics and design problems (e.g. Hasan, Morris and Proberts 2012; Neilson and Wilson 2011); search features (Hasan, Morris and Proberts 2012); task completion (effectiveness) (e.g. Hamel 2012; Molich et al. 2010; Weir, Anderson and Jack 2006); time taken to complete a task (efficiency) (e.g. Lavie, Oron-Gilad and Meyer 2011; Molich et al. 2010) or user satisfaction (e.g. Lavie, Oron-Gilad and Meyer 2011; Roberts et al. 2013).

Though there are guidelines and criteria for evaluating websites, Ball (2016) has developed criteria specifically to evaluate e-dictionaries; see also Ball and Bothma (2018) for a summarised version. This set of criteria takes into consideration standard usability issues, but incorporates aspects from lexicography and developments in information technology. The main criteria are based on issues relating to content, information architecture, navigation, access (searching and browsing), help, customization and the use of innovative technologies to manage data in e-dictionaries for search and display.

2. Description of the *Afrikaanse idiomewoordeboek*

The *Afrikaanse idiomewoordeboek* is a prototype e-dictionary of Afrikaans fixed expressions and contains a subset of fixed expressions of the Afrikaans language. This e-dictionary was developed with the intention to test certain theories and the functionality of the e-dictionary and as such, visual design and aesthetics were purposefully ignored. The design of this dictionary is based on the function theory of lexicography and presents several dictionaries that are created from one large database (Bergenholtz, Bothma and Gouws 2011: 36). The different dictionaries are monofunctional and provide data that are rele-

vant to specific situations. There are 15 fields for each expression in the database and different fields are presented in each dictionary. The five different dictionaries are based on different functions and are listed below.

- Function 1 — *Ek wil basiese inligting oor die idioom hê* — I want to have basic information about an expression
- Function 2 — *Ek wil die betekenis van 'n idioom verstaan* — I want to understand the meaning of an expression
- Function 3 — *Ek wil 'n idioom in 'n teks gebruik* — I want to use an expression in a text
- Function 4 — *Ek wil alles oor 'n idioom weet* — I want to know everything about an expression
- Function 5 — *Ek wil 'n idioom na Engels vertaal* — I want to translate an expression into English

Figure 1 shows the homepage of the *Afrikaanse idiome-woordeboek* and the five functions are given underneath the basic search field.

Figure 2 shows the basic data for the expression *kabaal opskop* (go through the roof) (function 1) which includes the following fields: expression, meaning, style and related expressions in English. Figure 3 shows all the data for this expression (function 4) and includes all the fields in the database that may be of interest to a user (fields for the editor's usage are not displayed).

The e-dictionary also makes use of various technologies such as advanced search and display options, browsing, multimedia in various articles, links to external sources that provide more data and customisation options. Each of these will be discussed below.

The advanced search and display options in the *Afrikaanse idiome-woordeboek* are shown in Figure 4. This allows a person to specify exactly in which fields in the database to search and exactly which fields to display in the article. A person can for example, choose to search for an expression with a certain meaning (e.g. *dronk* – drunk), by selecting to search in the meaning field (*betekenis*) and then to display the expression (*idioom*), its meaning (*betekenis*) as can be seen in Figure 5 as well as any other selected fields.

The ability to browse through the expressions in the *Afrikaanse idiome-woordeboek* is shown in Figure 6.

An example of an image (multimedia) that has been added to an article is shown in Figure 7.

A person can follow a link to an external source to see the context in which an example sentence has been used, see Figure 8.

The *Afrikaanse idiome-woordeboek* allows customisation in the sense that a user can save a selection of search and display fields and so create a custom dictionary specific to that user's requirements. Figure 9 shows where a selection has been saved.

Figure 10 shows how this saved selection can be loaded and a different search can be done with the options that have been saved.

Welkom by die idiome-woordeboek!

Soektaal

A A A

Soek op enige woord(e) of term(e) in die idioom 

Ek wil basiese inligting oor die idioom hê Ek wil die betekenis van 'n idioom verstaan
 Ek wil 'n idioom in 'n teks gebruik
 Ek wil alles oor 'n idioom weet Ek wil 'n idioom na Engels vertaal



Figure 1: Five different dictionaries based on functions

Ek wil basiese inligting oor die idioom hê

Idioom
Idioom: 'n Kabaal opskop
Betekenis: Te kere gaan; Uitvaar teen iets/iemand; 'n Lawaai veroorsaak
Styl: Informeel

Idiome in ander tale
Go through the roof
Raise hell
Raise the roof
Raise Cain

[Ek wil basiese inligting oor die idioom hê](#) | [Ek wil die betekenis van 'n idioom verstaan](#) | [Ek wil 'n idioom in 'n teks gebruik](#) | [Besigtig alle besonderhede](#) |

Figure 2: Basic data about an expression (function 1)

Ek wil alles oor 'n idioom weet

Idioom
Idioom: 'n Kabaal opskop
Terme: kabaal opskop
Betekenis: Te kere gaan; Uitvaar teen iets/iemand; 'n Lawaai veroorsaak
Taal: Af
Styl: Informeel
Verwysings: SW, TFW, HAT

Grammatika
Grammatika: Meerwoordige item

Voorbeeld
Voorbeeld: "Die Durandts van Bordeaux, kollega," sê hy, "is nie mense wat 'n kabaal opskop oor 'n gr
Outeur: Susanna M Lingua
URL: https://books.google.co.za/books?id=JhIbBwAAQBAJ&pg=PT140&lpg=PT140&dq=kabaal+opskopy&sig=65MRFUW_ThcX4n65ieDTMoFdejs&hl=en&sa=X&ved=0ahUKEwidluTJhKvJAhUE0hoKHemIDA

Voorbeeld: "Die kraai storm op hom af en takel hom met pote en vlerke toe terwyl hy 'n oorverdower
Outeur: Chris Barnard
URL: <https://books.google.co.za/books?id=V2RRBAAAQBAJ&pg=PT35&dq=kabaal+opskop&hl=en&sa=X&ved=0ahUKEwidluTJhKvJAhUE0hoKHemIDA>

Agtergrond
Agtergrond: Die woord kabaal kom oorspronklik uit Hebreeus (kabbâlâh). Dit het eers beteken Bybel
oproer".
URL: https://books.google.co.za/books?id=Im5SBAAAQBAJ&pg=PT269&lpg=PT269&dq=kabaal+opskop&source=bl&ots=M_vqWGv5m0&sig

Idiome in ander tale
Go through the roof
Raise hell
Raise the roof
Raise Cain

Figure 3: A partial screen capture of all the data about an expression (function 4)

Gevorderde soek- en vertoonopsies 

Soekterm Styl: ▼

Ek wil in die volgende velde soek: [\[Kies alle\]](#)

Idioom Variasie Betekenis Terme Grammatika Voorbeeld
Outeur Sinoniem Agtergrond

Ek wil die volgende inligting vertoon: [\[Kies alle\]](#)

Idioom Variasie Betekenis Terme Grammatika Voorbeeld
Outeur Sinoniem Agtergrond

Figure 4: Advanced search and display options

Gevorderde soek- en vertoonopsies

Idioom
Idioom: Bokke aanja
Betekenis: Dronk

Idiome in ander tale
[Be drunk](#)

[Ek wil basiese inligting oor die idioom hê](#) | [Ek wil die betekenis van 'n idioom verstaan](#) | [Ek wil 'n idioom in 'n teks gebruik](#) | [Besigtig alle besonderhede](#) |

Figure 5: An expression meaning 'to be drunk'



Figure 6: Browsing through a list of expressions

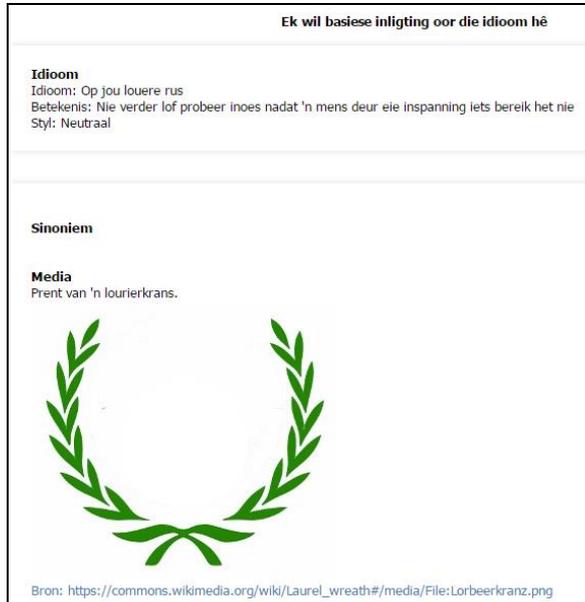


Figure 7: Multimedia used to illustrate a concept

Ek wil alles oor 'n idioom weet

Idioom
Idioom: Swaarde kruis
Terme: swaarde kruis
Betekenis: Teen iemand veg
Taal: Af
Styl: Neutraal
Verwysings: SW, IW

Grammatika
Grammatika: Meerwoordige i

Voorbeeld
Voorbeeld: Louis Trichardt Hoërskool het die afgelope naweek swaarde gekruis met Nylstroom Hoërskool. Nylstroom het hul tuisveld verdedig en Triegies moes op die rugbyfront die knie buig.
URL: <http://www.zoutnet.co.za/articles/sport/29845/2015-03-20/triegies-en-nylstroom-kruis-swaarde>

Triegies en Nylstroom kruis swaarde

Sport - Date: 20 March 2015
Written by: News Correspondent / **Viewed:** 214

Louis Trichardt Hoërskool het die afgelope naweek swaarde gekruis met Nylstroom Hoërskool. Nylstroom het hul tuisveld verdedig en Triegies moes op die rugbyfront die knie buig.

Die O/15-span speel gelykop 5-5, terwyl die eerstespan naelskraap 20-15 verloor. Triegies se hokkiespanne was gelukkig vir Nylstroom in eie munt terugbetaal met hulle skitterende spel. Die eerstespan was die wenner van die wedstryd 3-2, terwyl die O/15's 6-0 wen en die O/14's 1-0.

Die junior senior sokkerspanne van Triegies wen ook hul wedstryde, albei 2-1.

By die sokkiespan het Triegies opgemaak vir die verloor teen Hoërskool Ellisras. Die O/14-span wen 7-3.

Figure 8: Follow a link to see an example sentence in context

Gevorderde soek- en vertoonopsies ?

Soekterm Styl: Alle ▼

Ek wil in die volgende velde soek: [Kies alle]

Idioom Variasie Betekenis Terme Grammatika Voorbeeld Outeur Sinoniem Agtergrond

Ek wil die volgende inligting vertoon: [Kies alle]

Idioom Variasie Betekenis Terme Grammatika Voorbeeld Outeur Sinoniem Agtergrond

 **Stoor hierdie soek- en vertoonparameters**

Verskaf 'n naam vir hierdie soektog

Vul epos adres in

Figure 9: Save search and display options

The screenshot shows a web interface for loading a saved selection. The main heading is "Laai persoonlike soektogte". Below it, there is a form with a label "Vul epos adres in" and a text input field containing "jane@aivuw.co.za". Below this input field is a button labeled "Laai persoonlike soektogte". At the bottom of the interface, there is a search bar with the label "Soekterm" and the text "perde", a dropdown menu labeled "Navorsing", and a "Soek" button.

Figure 10: Load a saved selection

The technologies in the *Afrikaanse idiomewoordeboek* discussed above enables a person to get only information that is relevant to a specific situation. Usability evaluation can show whether the technologies incorporated in this e-dictionary are indeed used successfully.

3. Methodology

The usability of the *Afrikaanse idiomewoordeboek* was evaluated through a heuristic evaluation and usability tests.

Heuristic evaluation is when experts evaluate a system according to a set of criteria (Nielsen 1993: 155; Schneiderman and Plaisant 2010: 134-135). Criteria to evaluate e-dictionaries were developed by Ball (2016) and were used to do an evaluation on the *Afrikaanse idiomewoordeboek*; a summary of the criteria is available in Ball and Bothma (2018). The heuristic evaluation of the *Afrikaanse idiomewoordeboek* involved one expert and was done during the week of 11 to 15 April 2016 and completed on 10 May 2016.

Usability testing is a method where participants are asked to complete several tasks to determine whether a product can be used successfully (Molich and Dumas 2008: 264; Preece, Rogers and Sharp 2011: 438). The participants are observed and can be recorded (Krug 2005: 135; Preece, Rogers and Sharp 2011: 478). One method that is often used in usability testing is the think-aloud method. Participants are encouraged to articulate their thoughts whilst performing the tasks to give the researchers insight into their actions (Schneiderman and Plaisant 2010: 143). At the end of a usability test the participant is often asked to complete a questionnaire to obtain subjective data (Fernandez, Insfran and Abrahão 2011: 802; Preece, Rogers and Sharp 2011: 477; Schneiderman and Plaisant 2010: 139).

Seven participants took part in the usability testing of the *Afrikaanse idiomewoordeboek* that was done from 16 April to 4 May 2016. The participants were selected by purposive sampling, and people with fairly similar demographics, backgrounds and experiences were chosen. Though the discount usability theory discussed earlier indicates that good results can be obtained from using few participants, the researchers recognise that a bigger sample would probably point out more issues. However, it was considered that seven participants would be sufficient for a first evaluation, also considering that a prototype was being evaluated. The evaluation of a final product could consider using more participants. The participants were met at locations that suited the participants. Though the tests were conducted in a laboratory type environment (office, study, etc.), as much as possible was done to reduce the artificial nature of the test. Each of the 16 tasks that the participants had to do contained a scenario that described a real information need. The participant could therefore imagine him-/herself in a certain situation and make an appropriate response. During the tests the participants were observed by a researcher and recorded. The participants themselves were recorded by a standalone camera and the computer screen was recorded using the screen capturing software, BB FlashBack Express. This software allowed for the participant to be recorded with the computer's webcam and the screen to be recorded at the same time. After completing the tasks, each participant was asked to complete a questionnaire with 35 questions, including both open- and close-ended questions. Both the tasks and questionnaire were designed to evaluate the e-dictionary according to the criteria for e-dictionaries.

The 16 tasks are listed in the table below. The tasks listed here are translated from the original that was given to the users in Afrikaans. The purpose of the task is given in an additional column. The tasks are listed here as the discussion that follows refers to specific tasks. The questionnaire that was given to participants is added as an appendix.

	Task	Purpose
1	You are looking for the meaning of the expression <i>uit die lug val</i> (an expression meaning "to appear unexpectedly").	The purpose of the task was to acquaint the participant with the dictionary and set them at ease.
2	You are busy reading a newspaper article and come across the expression <i>te berde bring</i> (an expression meaning "to broach a subject"). You were under the impression it means "to put something away", but from the article it appears that the expression might mean something else. What is the real meaning of the expression <i>te berde bring</i> ?	The purpose of the second task was to see how a user would find the meaning of an expression.

3	<p>You are busy writing a letter for the local newspaper in which you want to use the expression <i>te kort skiet</i> (an expression meaning "not to have enough"), but you are uncertain how to use it in the following sentence: "<i>Soos verlede jaar, skiet hul weer te kort aan oorspronklikheid.</i>" (Just like last year, they lack originality.)</p> <p>You specifically want to know if you can change the word order (e.g. "<i>skiet ... te kort</i>") or must the word order remain "<i>te kort skiet</i>".</p>	<p>The purpose of task 3 was to see how a user would find out how to use an expression in a text, specifically how to check what the word order of an expression must be.</p>
4	<p>You hear the expression <i>dans na sy pype</i> (an expression meaning "to do everything someone else wants") in a conversation and are interested in the expression and want to find out everything about it, also where it comes from.</p>	<p>The purpose of task 4 was to see how a user would find out everything there is in the dictionary about an expression.</p>
5	<p>You are doing translation for a magazine and want to find a good translation for the expression <i>swaarde kruis</i> (an expression meaning "to fight") to use in an English text. Confirm the exact meaning of the English expression.</p>	<p>The purpose of task 5 was to see how a user would use the dictionary to find a translation and confirm the meaning of the expression.</p>
6	<p>You read the following on a billboard next to the road: <i>"Konferensie neem menswaardigheid en mynbou onder die loep"</i> ("Conference investigates human dignity and mining") You are curious as to what the expression <i>onder die loep neem</i> (an expression meaning "to scrutinise") means, what <i>loep</i> ("lens") is and what it looks like.</p>	<p>The purpose of the task was to see if participants would find the inclusion of multimedia in an article useful.</p>
7	<p>You had used the expression <i>voelers uitsteek</i> (an expression meaning "to try and find out what someone thinks") in a text, but wonder if there might be a better expression for the context.</p>	<p>The purpose of the task was to see if a user can browse in a dictionary by following internal links, for example, a synonym.</p>
8	<p>You remember that in one of your previous searches in the dictionary you found an example sentence about a banting diet. What would you do to see this sentence in context?</p>	<p>The purpose of this task was to test if a user could use the advanced search and display options, and if the users could follow an external link.</p>

9	You work for a publisher and want to find out what example sentences from Deon Meyer's books were used in the dictionary.	The purpose of task 9 was to see how the users would use the advanced search features to search for example sentences from an author.
10	You are looking for an expression that means that someone did not keep a secret.	The purpose of task 10 was to see how participants would search for an expression when only the meaning is known.
11	You think that there is an expression that means that someone is drunk that contains the word <i>bokke</i> ("goats"). You are trying to find this expression.	The purpose of task 11 was to find out how users would do a complex search. In this task, it would be searching by meaning and a known term.
12	You are busy with a project and are doing research about the background of expressions. You are interested to display the meaning and the background information. You are not interested to display the rest of the fields. It is sufficient to search in the 'expression', 'variation' and 'term' fields. Set up a search like this and save these search and display options. Test the search by looking for the background information of the expression <i>groot kokkedoor</i> (an expression meaning "an important person").	The purpose of the task was to see if users could save selected search and display options in the advanced search. The users first had to select the appropriate search and display options, save the options, and then test them with a search.
13	Use the search that you saved in 12 to get the background information for the expression <i>kabaal opskop</i> (an expression meaning "to make a noise").	The purpose of the task was to load the saved advanced search options from task 12 and use them in a new search.
14	You are helping a school child with a task and must find an expression that starts with the letter M. Choose one.	Task 14 was designed to test whether a user will find it useful to browse through the dictionary.
15	You remember the expression <i>geld soos bossies</i> (an expression meaning "a lot of money"). Try to find if it is acceptable to use it in a speech.	The purpose of task 15 was to see how users would find specific information after being exposed to the functions as well as the advanced search.
16	You are uncertain what the field 'terms' in the advanced search means. Consult the 'help' to find out how it works.	The purpose of the task was to see if users could find and use the 'help' function of the dictionary.

Table 1: Tasks for the usability testing of the *Afrikaanse idioome-woordeboek*

4. Findings

The findings from the usability testing will be discussed according to the main categories of the evaluation criteria for e-dictionaries (Ball 2016).

4.1 Content

When evaluating the content in an e-dictionary, the following should be considered: whether the e-dictionary can give relevant data, the level of detail and complexity of the data given, the credibility of the e-dictionary, the writing and editorial style used and the usage of multimedia.

To determine whether the *Afrikaanse idiome-woordeboek* gives relevant data to a user, the researchers firstly considered whether the tasks could be completed successfully, and secondly, whether data irrelevant to the task are withheld. (It is important to consider whether irrelevant data are withheld, because if too much data are given there is the possibility that the user is overwhelmed.) In terms of task completion, five of the seven participants could complete all the tasks successfully, but two participants completed all tasks except the fifteenth task correctly. At some points during the tests, the researcher had to provide some guidance if participants did not know how to proceed, but then the tasks could be completed. This was taken into consideration during the analysis of the tests. Due to this success rate of task completion, the researchers found that the e-dictionary is effective as it gives the data that are necessary to complete specific tasks. From the data collected in the questionnaire it was clear that most of the participants were positive that the dictionary provided relevant data for the task at hand and did not give too much data. However, during the tests participants reacted differently to the amount of data given. One participant specifically expressed that the perfect amount of data was given, whereas others expressed disappointment that not more data were given. One suggested that a brief summary of the expression is displayed with the option to see more data. In some cases, participants who had displayed more data found it easier to complete the tasks.

When looking at the level of detail and complexity of data that an e-dictionary provides, the use of external sources to provide more data to a user is evaluated. In the eighth task participants were required to follow an external link to see an example sentence in context (see Figure 8). Only one participant opened the link immediately, all other participants had to be prompted. Even on prompting, one of the participants did not open the link but commented that that is what (s)he would have done to see the sentence in context. Though few made comments about the use of the external sources to provide context for example sentences, most indicated in the questionnaire that the use of external sources is useful. Most participants agreed that the level of complexity and detail in the e-dictionary is appropriate.

Most participants agreed that the e-dictionary is credible, but two were not convinced. One of these participants indicated that the example sentences used in an e-dictionary should be grammatically correct and trustworthy. Though no task was designed specifically to determine the perceived credibility of the e-dictionary, one of the examples sentences that the participants encountered in doing task 3, is written in a mixture of Afrikaans and English (*Waar my opvoeding te kort skiet moet my attitude maar opmaak* — "My attitude should make up for where I lack education") and three participants specifically commented about this during the tests. Another participant had noticed example sentences, taken from newspapers or other such sources, that express political or religious viewpoints and indicated that data in an e-dictionary should be neutral.

The writing and editorial style used in an e-dictionary can help a user to make sense of the data in the e-dictionary, particularly when words are not abbreviated. The *Afrikaanse idiome-woordeboek* included some abbreviations where references to other dictionaries were made. Two participants were particularly confused by the use of these abbreviations. In addition, the participants in general were negative about the labels in the *Afrikaanse idiome-woordeboek*. One commented that some of the labels were written in old-fashioned Afrikaans and another thought that the labels for the different dictionaries (functions) were too long.

The purpose of the sixth task was to determine whether the participants would find the inclusion of multimedia useful (see Figure 7). In this task the participants had to determine what the word *loep* ("lens") means and all the participants referred to the image included in the article to answer the question. One participant even enlarged the image to see it more clearly.

4.2 Information architecture

To make a judgement about the information architecture of an e-dictionary, the way in which the data are organised or structured to give access to it should be evaluated, as well as the way in which data are organised on page or article level.

The data in the *Afrikaanse idiome-woordeboek* are organised according to different functions (see Figure 1). Task 2 to 5 were specifically designed to evaluate the use of these functions in the *Afrikaanse idiome-woordeboek*. Participants did not struggle to find the functions. Initially, however, most participants would type in the search string and press the 'Enter' key without considering or selecting a function. The *Afrikaanse idiome-woordeboek* does not work with the 'Enter' key, which forced the participants to select an appropriate function and then click the search button. The five functions are not only on the homepage, but repeated on the subsequent pages so that a user can change options easily. Most participants changed from one function to another easily, but some participants went back to the homepage to redo the search to try to get to the rele-

vant data if they were not successful at first. Only one participant specifically tried to work out what the difference between the different functions are. At first (s)he assumed they were the same, but later noticed that different fields were displayed in different options.

Tasks 2 to 5 will be discussed in more detail. For these tasks (where the function and the task matched clearly) most participants used the correct function and could complete the task easily. To complete task 2, the most appropriate function to consult was function 2 and four participants used this function. Two participants left the default function selected, but one of these participant changed to function 2 and then to function 4 once the article was displayed. Another participant selected function 4. The most appropriate function to do task 3 was function 3 and four participants did this. Two other participants tried the advanced search first, before also using the basic search and function 3. One participant tried various search strings and opened the articles with either function 3 or 4. To do task 4, a user has to select function 4 and six participants did this. Only one participant used the advanced search options, but did not select any display fields. Upon opening the article, (s)he saw no data and selected function 4 to display data. The purpose of task 5 was to see if the participants would use function 5 to get a translation for a specific expression. All participants expect one used function 5 to start the search. This participant chose function 4. Once the article was open, some participants immediately opened function 4 to see more data.

In other tasks (specifically tasks 7 and 15) that were designed to test other issues, some participants found the functions confusing. This happened specifically if a participant did not select the correct function or expected a function to contain data that it did not. For example, in the *Afrikaanse idioem-woordeboek*, synonyms are displayed under functions 1 and 4, but not 3. This caused some confusion.

In terms of the layout of a page, participants did not have trouble to find the correct data in the article and most indicated in the questionnaire that the data in the article are clearly organised. Most also indicated that they thought it useful to be able to filter data in an article. This could refer to the use of functions that filter pre-selected fields or to the advanced search options where display fields can be selected (see Figure 4). Though the responses from the questionnaire were predominantly positive, it was evident that the advanced search and displayed options were confusing to most participants. Tasks 9 and 12 tested the use of the display options specifically, but their use in other tasks was noted. The display options were sometimes ignored, all fields were selected, inappropriate fields to the task at hand were selected or the same fields as the search fields were selected. In the article of an expression, the *Afrikaanse idioem-woordeboek* does not display fields when there are no data for those fields in the database. For example, if no synonyms have been added in the database, this field is omitted in the article. This caused confusion for some participants; for example, if such a display field had been selected specifically and then did not display.

From these usability tests it was clear that most participants struggled to filter data on a page by using the advanced display options. However, during the tests it was clear that some participants wanted to filter out irrelevant data. For example, one participant wanted to filter out the grammar and reference data and another participant wanted to filter out the translations.

4.3 Navigation

Under the category of navigation, the ease with which a user moves through the e-dictionary is evaluated, as well as whether the user knows where (s)he is at all times. As links are the primary mechanisms through which a user navigates, they are also evaluated.

The ease with which a user can navigate will influence how quickly the user will find the relevant data for a specific task. This allows for a judgement about the efficiency to be made. The time it takes a user to get to data in an e-dictionary is very important. In this study, the users were asked to think-aloud during the tests which gave valuable data, but negatively influenced the time it took to do a task. Consequently, in this study it was not sensible to look at the time it took to complete tasks, but rather to look at the steps that participants took to get to the relevant data. The participants themselves indicated that they could find data easily, but felt that they had to scroll too much or click through too many levels. The scrolling was probably caused by the fact that the results were listed below the advanced search, and as such participants typically had to scroll down after a search to see the results. There were many comments about the fact that the search results were not prominent. Though the hierarchy of the site is fairly shallow (typically two levels), the amount of clicking was influenced by two factors. The search results themselves do not contain any additional descriptive data. A user first has to open the expression to see the article to find the data to answer the question in the task. Secondly, the search field is only on the homepage. The user then always has to navigate back from the article to start another search. Based on this it appears that participants did not find the *Afrikaanse idiome-woordeboek* efficient to use.

Participants indicated that they knew where they were in the e-dictionary, but were not as positive about knowing where to go next in the e-dictionary. This could be as a result of the fact that most participants did not know how to store the advanced search and display options in task 12 or use them in task 13. Very few participants found the help option in task 16. Some participants also felt that the link to the homepage could have been more prominent.

Though participants were negative about the labels as mentioned under the section on content, they indicated that they knew where the links were going to lead them and it did not lead them to unexpected places.

4.4 Access (searching and browsing)

The way in which a system provides access to the data in the e-dictionary

should be evaluated. Access is typically given through searching or browsing options. The basic as well as advanced search options in an e-dictionary should be evaluated, as well as any browsing mechanisms. Filters can help users to narrow the content down to something specific. If a large number of results is retrieved, users should be able to manipulate the results. Lastly, the processing speed can be evaluated, though generally, systems are fast enough that this is becoming negligible.

In the usability tests, tasks 1 to 7 and task 15 could have been done by using the basic search. As has been mentioned before, the basic search field is only on the homepage (see Figure 1). As a result, most participants did not think that it is easy to find or that it is easy to change a search and search for something new. Two participants had made typing errors whilst searching and neither picked it up immediately when no results were retrieved.

Tasks 8–11 were designed to test the use of the advanced search features (see Figure 4). Although the data from the questionnaire made it clear that most participants think advanced search features can help a person to find relevant data, they were not convinced that the advanced search features in the *Afrikaanse idiome-woordeboek* are easy to use or easy to understand. From the usability tests it was also clear that the advanced search and display options were confusing and generally used incorrectly. In addition, some of the comments from the participants were that there are too many options, it is time consuming and the layout should be improved to communicate more clearly how the advanced search and display fields should be used. Most participants used the advanced search incorrectly, by indicating in the search field what data they wanted to display, instead of selecting where (i.e. in what field) they wanted to search. As the display fields are the same as the advanced search fields, these were often ignored or the same fields were selected. If the article then did not display any data or not the relevant data, it caused confusion.

One participant assumed that the basic and advanced search options were linked and used them together; for example, (s)he would select fields under the advanced search, but search with the basic search. Two participants specifically commented that the difference between the sections on the homepage were not clear.

Task 14 tested the use of the browse option, where all the expressions are listed in an alphabetical list (see Figure 6). Most participants did not see the browse option initially and tried to search for an expression that starts with an 'M' by using the basic or advanced search. Once the researcher had pointed out the browse option, they could use it and from the data from the questionnaire it is clear that the participants were positive about the ability to browse through the dictionary. One of the participants had used the browsing option in previous tasks when (s)he had not been successful in using the search options. One of the suggestions for improvement was that it would be helpful to only type in the first couple of letters and then jump to the expression(s) that contains those letters when browsing.

Another point of evaluation in terms of browsing is the ability to browse

between items in the e-dictionary. The purpose of task 7 was to test whether users could browse between expressions, by asking users to find an appropriate synonym for an expression. Some participants did not follow the link for the synonym. Others followed the link successfully to confirm the meaning of the expression. Others followed various internal links (links for translations and synonyms) to confirm that they had selected a valid synonym. In a different task (task 5) one participant also relied heavily on internal links to confirm his/her answer.

Filtering in the *Afrikaanse idiome-woordeboek* is made possible through functions, as well as the advanced search and display options. These have been discussed already under the category information architecture.

In some systems it is possible to search in the results that have been retrieved or to manipulate the results to some extent to refine them further; one participant commented that (s)he would have liked to be able to search in the results. However, most participants felt that the search results were logically arranged. Some participants were frustrated that the search results were not saved. This means that if a user retrieves more than one result, open one and go back to view the other, (s)he has to redo the search.

4.5 Help

The purpose of task 16 was to see whether participants could find the 'help' option easily and find the information in the documentation useful. Although most participants struggled to find the help option (an icon of a question mark), most thought that the information provided sufficient help. There were two incidents in tasks before task 16, where participants did not know how to proceed and consulted the 'help' for guidance. Neither found the information they were looking for in the documentation.

4.6 Customisation

The *Afrikaanse idiome-woordeboek* allows a user to store selections for the advanced search and display options (see Figures 9 and 10). Task 12 tested whether participants could save a selection and task 13 tested whether the participants could retrieve that selection and use it in a search.

Most users struggled to find where to store the selection and had to be shown. Some of the participants were also confused about what was saved and how it was being done. In order to save a search a person has to enter their email address, which is used to uniquely identify the user. This was also confusing, as some participants thought that something was emailed to them. Another participant was sceptical about putting in an email address in a system. Some participants thought that the results were being saved in task 12 and that in task 13 they had to search in the results of task 12. One participant was

confused that upon loading a saved selection, the check boxes for that selection were not automatically checked. Another participant redid the search in task 12 and saved the selection a second time and noticed that the system does not validate the names of the saved selection as (s)he had used the same name in both instances.

4.7 Innovative technologies

No other technologies, apart from those already discussed, were tested in the *Afrikaanse idiome-woordeboek*.

5. Recommendations

5.1 Usability evaluation

The usability tests on the *Afrikaanse idiome-woordeboek* clearly showed that the users did not always use the functionality of the e-dictionary as the designers had intended. This underscores the importance of usability evaluation when an e-dictionary is designed that makes use of innovative technologies. Users do not necessarily have the same point of view as the designers of a system. Usability evaluation will help to show how users really use e-dictionaries and how they understand the options that are included.

This study also supports the idea that usability evaluations do not need to be elaborate or expensive exercises, but that simplified usability evaluation can be very effective. Only seven participants took part, the study was conducted in simple and accessible environments and various important usability issues were identified.

5.2 Information architecture

As was explained earlier, information architecture refers to the way in which the information in a product is organised or structured. In the case of the *Afrikaanse idiome-woordeboek* the information was organised in different functions and presented as different dictionaries to the user. From the usability tests it is clear that the difference between the functions were not always clear and the users did not always know what to expect from the functions. The researchers suggest that more is done in e-dictionaries to communicate the underlying structure or information architecture to the users. One way to communicate more clearly with the users is to write labels, headings or other micro-content that are easy to understand. For example, the participants in these usability tests were negative about the labels used in the *Afrikaanse idiome-woordeboek*.

A suggestion for the labels for the different functions on the homepage is suggested in Figure 11.



Figure 11: Suggested labels for the functions on the homepage

5.3 Searching in e-dictionaries

From the usability tests, it seems that users expect more from the search algorithm and the results themselves before opening a relevant article. For example, one user tried to determine the correct usage by using two different search strings and to make a judgement based on the number of retrieved results, trusting that the most results would be the correct option. The two users who had made typing errors and did not notice it at first should be considered. Implementing a fuzzy matching function to help users who mistype could be a valuable addition to an e-dictionary. Auto-completion of text could also be helpful, and might be something that users expect as it is not uncommon in search systems. There were also various suggestions that the results should give a little more data. This could, for example, help a user know that the correct item is selected to open. It is therefore recommended that more is done to improve the searching algorithms in e-dictionaries, as well as explore different ways to display the search results. In addition, more advanced search features can be added and different ways to present the advanced search options can be tested. For example, though the *Afrikaanse idioom-woordeboek* allows a user to search in different fields, it does not allow a user to define Boolean operators between fields explicitly, nor does it allow truncation. Even different layouts for the advanced search fields can be tested to see if other presentations make it easier to understand.

Apart from search features, it is recommended that some sort of browsing option is included. A user has to have a clearly defined information need and understand the information space for a search to be successful. If a user struggles with the search and a browsing option is available, then there is at least another access point to the content. This idea is supported by the case of the participant who had made a typing error and then tried the browse option to get to the relevant expression.

5.4 Data for e-dictionaries

At various points during the usability evaluation different participants had commented that they had expected or would have liked to see more data. The researchers therefore recommend that lexicographers should not hesitate to include more data in e-dictionaries. However, there were participants who felt the amount of data given was sufficient. As a result it is suggested that more data should always be accompanied with the ability to filter out unwanted data to avoid overload. Most participants also indicated that it is helpful to be able to filter data on a page. It can also be useful to display more data with a button or link that opens more data.

There are many sources available for lexicographers to include more data in e-dictionaries (Bothma 2011: 80), yet care should be taken that users know when something has been approved by lexicographers and when something is from an untested external source (Heid, Prinsloo and Bothma 2012: 285; Tarp 2012: 264). This is confirmed in these usability tests as the participants were sceptical about some of the example sentences that were included.

As some users seem to want more information, it is suggested that different ways to incorporate or give access to external sources are explored.

5.5 Education and training

Even if usability evaluation and proper design can produce e-dictionaries that are easy to use, the value of education and training should not be underestimated. Good user skills are imperative to successful dictionary consultation (Lew 2013: 16). Particularly if an e-dictionary makes use of advanced search features or other innovative technologies, users should be educated to be able to make use of these. In his paper, Lew (2013) discusses skills that are necessary for the successful consultation of e-dictionaries and how it is related to digital literacy. He discusses how some traditional skills necessary for dictionary use change in the digital environment; for example, e-dictionaries can have a wider variety of structures or data in an article can be hidden. Furthermore, he discusses various skills that are purely relevant for a digital environment, such as searching with wildcards or using Boolean operators. In his conclusion Lew (2013: 29) remarks that an "important concern is finding an appropriate context for teaching e-dictionary skills." Ball and Bothma (2017) suggest that information literacy programmes should be considered as a platform where dictionary use can be taught. Information literacy programmes are already used to teach users how to search in online environments such as databases and how to use search techniques such as Boolean operators and wildcards effectively. Dictionary literacy can become a subset of information literacy.

6. Conclusion

The *Afrikaanse idiomewoordeboek* incorporates various technologies to allow a user to get only relevant information for a specific situation. The usability testing of this e-dictionary showed that users do not necessarily use a system as the designers intend and various usability issues could be identified with discount usability methods. Various recommendations regarding the design of the *Afrikaanse idiomewoordeboek*, as well as e-dictionaries in general, have been made. If the *Afrikaanse idiomewoordeboek* is modified according to the suggestions in this article, further usability tests can be done to see if it can be used more efficiently to get only relevant information for a specific situation.

Usability evaluations do not need to be complicated and expensive exercises. This article has shown that using a few users in usability tests as suggested by the discount usability approach is effective in finding usability issues in a design. If lexicographers are serious about designing modern e-dictionaries that incorporate technology that is usable, they should embrace usability evaluation as practice. The discount usability methods of heuristic evaluation and usability testing are both appropriate for lexicographers. Lexicographers can perform heuristic (expert) evaluations by using the criteria explained in Ball and Bothma (2018) and also do their own usability tests by creating tasks and observing users as was explained in this article. The usability evaluations should be used at various stages in the development of e-dictionaries. Iterative evaluations after each design phase should be done. Discount usability methods are purposefully simple and non-expensive so that a product can be evaluated repeatedly throughout the design process. In the initial stages, prototypes can be evaluated and then progressively more complete products. If prototypes are evaluated then changes can be made more easily. For example, in the case of the *Afrikaanse idiomewoordeboek*, a prototype was evaluated. As it is a prototype, it will be easier to implement the necessary changes than if the evaluations were only done after the project has been completed.

The best type of prototype to use for usability evaluation can be a point of future research. The designers of the *Afrikaanse idiomewoordeboek* purposefully did not pay attention to the visual design of the system, as the purpose of the prototype was to be able to test the functionality of the e-dictionary. However, it is unclear if the lack of visual appeal or visual cues had an influence on the users' ability to use the e-dictionary effectively. For example, would improving the layout of the fields in the advanced search without changing the functionality change users' use of the advanced search? Or would better visual cues improve their understanding of the information architecture? It is therefore still an open question as to what extent the functionality of an e-dictionary can be tested without taking visual design into consideration.

More usability evaluations of e-dictionaries and more research of best practices of usability evaluations of e-dictionaries can lead to advanced information tools that address users' needs effectively.

7. References

- Ball, L.H.** 2016. *An Evaluative Study to Determine to What Extent Technology Can be Used in e-Dictionaries to Provide Relevant Information on Demand*. Master's Thesis. Pretoria: University of Pretoria. <https://repository.up.ac.za/handle/2263/61338>.
- Ball, L.H. and T.J.D. Bothma.** 2017. The Importance of Usability Evaluation when Developing Digital Tools for a Library — a Case Study. *Proceedings of Bobcatsss 2017, 25–27 January 2017, Tampere, Finland*: 137-142. <http://tampub.uta.fi/handle/10024/101894>.
- Ball, L.H. and T.J.D. Bothma.** 2018. Establishing Evaluation Criteria for e-Dictionaries. *Library Hi-Tech* 36(1). (In press.)
- Bergenholtz, H.** 2011. Access to and Presentation of Needs-adapted Data in Monofunctional Internet Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 30-53.
- Bergenholtz, H. and I. Bergenholtz.** 2011. A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 187-207.
- Bergenholtz, H., T.J.D. Bothma and R.H. Gouws.** 2011. A Model for Integrated Dictionaries of Fixed Expressions. Kosem, I. and K. Kosem (Eds.) 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users: Proceedings of eLex 2011, Bled, Slovenia, 10–12 November 2011*: 34-42. Ljubljana: Trojina, Institute for Applied Slovene Studies.
- 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 R.H. Gouws.** 2007. Korrek, volledig, relevant. Dít is die vraag aan leksikografiese definisies. *Tydskrif vir Geesteswetenskappe* 47(4): 568-586.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bothma, T.J.D.** 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 71-102.
- Calisir, F., A.E. Bayraktaroğlu, C.A. Gumussoy, Yi. Topcu and T. Mutlu.** 2010. The Relative Importance of Usability and Functionality Factors for Online Auction and Shopping Web Sites. *Online Information Review* 34(3): 420-439.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Fernandez, A., E. Insfran and S. Abrahão.** 2011. Usability Evaluation Methods for the Web: A Systematic Mapping Study. *Information and Software Technology* 53(8): 789-817.
- Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.).** 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.
- Hamel, M.-J.** 2012. Testing Aspects of the Usability of an Online Learner Dictionary Prototype: A Product- and Process-Oriented Study. *Computer Assisted Language Learning* 25(4): 339-365.
- Hasan, L., A. Morris and S. Probeta.** 2012. A Comparison of Usability Evaluation Methods for Evaluating e-Commerce Websites. *Behaviour & Information Technology* 31(7): 707-737.
- Heid, U., D.J. Prinsloo and T.J.D. Bothma.** 2012. Dictionary and Corpus Data in a Common Portal: State of the Art and Requirements for the Future. *Lexicographica* 28(1): 269-291.
- ISO 9241-11.** 1998. *ISO 9241-11:1998 Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) — Part 11: Guidance on Usability*. http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883.

- Krug, S.** 2005. *Don't Make Me Think!: A Common Sense Approach to Web Usability*. Second edition. Berkeley: New Riders.
- Lavie, T., T. Oron-Gilad and J. Meyer.** 2011. Aesthetics and Usability of In-Vehicle Navigation Displays. *International Journal of Human-Computer Studies* 69: 80-99.
- Lew, R.** 2012. How Can We Make Electronic Dictionaries More Effective? Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 343-361. Oxford: Oxford University Press.
- Lew, R.** 2013. Online Dictionary Skills. Kosem, I., J. Kallas, P. Gantar, S. Krek, M. Langemets and M. Tuulik (Eds.). 2013. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17-19 October 2013, Tallinn, Estonia*: 16-31. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.
- Lindgaard, G. and J. Chattratichart.** 2007. Usability Testing: What Have We Overlooked? *CHI 2007: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 28 April-3 May, 2007*: 1415-1424. New York: ACM.
- Molich, R., J. Chattratichart, V. Hinkle, J.J. Jensen, J. Kirakowski, J. Suaro, T. Sharon and B. Traynor.** 2010. Rent a Car in Just 0, 60, 240 or 1,217 Seconds? — Comparative Usability Measurement, CUE-8. *Journal of Usability Studies* 6(1): 8-24.
- Molich, R. and J.S. Dumas.** 2008. Comparative Usability Evaluation (CUE-4). *Behaviour & Information Technology* 27(3): 263-281.
- Neilson, C.J. and V. Wilson.** 2011. "We Want it Now and We Want it Easy": Usability Testing of an Online Health Library for Healthcare Practitioners. *Journal of the Canadian Health Libraries Association* 32(2): 51-59.
- Nielsen, J.** 1993. *Usability Engineering*. Cambridge: Academic Press.
- Nielsen, J.** 2000. *Why You Only Need to Test with 5 Users*. <http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>.
- Nielsen, J.** 2009. *Discount Usability: 20 Years*. <http://www.nngroup.com/articles/discount-usability-20-years/>.
- Nielsen, J.** 2012a. *How Many Test Users in a Usability Study?* <http://www.nngroup.com/articles/how-many-test-users/>.
- Nielsen, J.** 2012b. *Traveling Usability Lab*. <http://www.nngroup.com/articles/traveling-usability-lab/>.
- Nielsen, S.** 2011. Function- and User-related Definitions in Online Dictionaries. Kartashkova, F.I. (Ed.). 2011. *Ivanovskaya lekskograficheskaya shkola: traditsii i innovatsii (Ivanovo School of Lexicography: Traditions and Innovations. A Festschrift in Honour of Professor Olga Karpova)*: 197-219. Ivanovo: Ivanovo State University.
- Preece, J., Y. Rogers and H. Sharp.** 2011. *Interaction Design. Beyond Human-Computer Interaction*. Third edition. Chichester: John Wiley and Sons.
- Roberts, M.J., E.J. Newton, F.D. Lagattolla, S. Hughes and M.C. Hasler.** 2013. Objective versus Subjective Measures of Paris Metro Map Usability: Investigating Traditional Octolinear versus All-Curves Schematics. *International Journal of Human-Computer Studies* 71: 363-386.
- Shneiderman, B. and C. Plaisant.** 2010. *Designing the User Interface. Strategies of Effective Human-Computer Interaction*. Fifth edition. Boston: Pearson Education.
- Tarp, S.** 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Tarp, S.** 2008. The Third Leg of Two-Legged Lexicography. *Hermes — Journal of Language and Communication Studies* 40: 117-132.

- Tarp, S.** 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 54-70.
- Tarp, S.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* (28)1: 253-268.
- Tullis, T. and W. Albert.** 2008. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Burlington: Morgan Kaufmann.
- Verlinde, S. and G. Peeters.** 2012. Data Access Revisited: The Interactive Language Toolbox. Granger, S. and M. Paquot. (Eds.). 2012. *Electronic Lexicography*: 147-162. Oxford: Oxford University Press.
- Weir, C.S., J.N. Anderson and M.A. Jack.** 2006. On the Role of Metaphor and Language in Design of Third Party Payments in eBanking: Usability and Quality. *International Journal of Human-Computer Studies* 64: 770-784.

Appendix: Questionnaire used in the usability evaluation

The questionnaire that was given to the users in the usability evaluation of the *Afrikaanse idiomewoordeboek* was translated from Afrikaans by the researcher and is given here.

		Strongly disagree	Disagree	Neutral	Agree	Strongly agree
	CONTENT					
1	The dictionary has correct and relevant information for the tasks that I had to do.					
2	The amount of information presented in the dictionary is too much and even overwhelming at points.					
3	The level of complexity of the information in the dictionary is appropriate.					
4	The amount of detail given for each item in the dictionary is appropriate.					
5	The links to external sources that give more information about a specific item are useful.					
6	The use of multimedia (e.g. audio files, images or videos) helped me to complete my tasks more effectively.					
7	The dictionary appears trustworthy .					
8	I understand the labels used for buttons, headings, section dividers, etc. in the dictionary.					
9	Please provide any additional comments about the amount and relevance of the information in the dictionary with regard to the tasks that you had to do.					

INFORMATION ARCHITECTURE						
10	The overall organisation of the dictionary is easy to understand . (For example, I knew which sections in the dictionary to choose for different tasks.)					
11	The overall organisation of the dictionary makes it easy to use . (For example, the different functions in the dictionary made it easier to complete my tasks.)					
12	The information for each idiom is clearly organised.					
13	I had to scroll too much to find specific information on a page.					
14	I had to click through too many levels to find the information I was looking for.					
15	It helps to manipulate and filter information on a page to show exactly what I want to know.					
16	Please provide any additional comments about organisation and structure of the dictionary.					
NAVIGATION						
17	I could easily find the information I was looking for.					
18	It was always clear where in the dictionary I was. (I never felt lost.)					
19	It was always clear where I should go next (i.e. navigate) as I looked for information.					
20	The links in the dictionary are labelled in such a way that I understood where the link would lead.					
21	The links in the dictionary did not take me to unexpected places.					
22	Please provide any additional comments about navigating around the dictionary.					

ACCESS (Searching and browsing)						
23	The search field is easy to find.					
24	It is easy to change my current search and search for something new.					
25	The advanced search features are easy to use.					
26	The advanced search features help me to be very precise and find specific information. (For example, I was looking for grammatical information specifically as opposed to just the meaning.)					
27	The search options are overwhelming and difficult to understand.					
28	The search options are too time-consuming.					
29	I found the ability to browse through the items in the dictionary useful.					
30	When there are a lot of results from a search, they are logically arranged (i.e. it makes sense).					
31	I can easily manipulate search results.					
32	Please provide any additional comments about searching in the dictionary.					
HELP						
33	It is easy to find the 'help' section of the dictionary.					
34	The 'help' section in the dictionary provides sufficient and understandable help.					
35	Please provide any additional comments about the 'help' in the dictionary.					

Polyseme Selection, Lemma Selection and Article Selection

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: In linguistics, more specifically in the field of lexical semantics, a lot of attention has been given to polysemy and homonymy. The identification of and distinction between polysemy and homonymy should not be regarded as unproblematic. The lexicographic practice has traditional ways of presenting and treating polysemy and homonymy. This paper focuses on approaches in both linguistics and lexicography to polysemy and homonymy. Examples from the lexicographic practice are given. It is then shown that the traditional lexicographic presentation and treatment of homonymy and polysemy in dictionaries with a text reception function, does not really assist the users adequately in their search to find the appropriate meaning of an unfamiliar linguistic expression. It is shown that different dictionaries often have the same lemma selection but not the same selection of polysemes. It is important that a dictionary should correctly coordinate a meaning and a specific linguistic expression. Consequently, a new approach is suggested for the presentation and treatment of homonymy and polysemy. Negotiating criticism expressed in both linguistics and lexicography, it is proposed that the lexicographic practice, in the case of dictionaries for text reception, should abolish the traditional distinction between homonyms as well as the presentation of the different senses of a polysemous word in a single article. Each meaning, whether the only meaning of a lexical item or one of any number of different senses, should be the only item giving the meaning in an article.

Keywords: ARTICLE SELECTION, DICTIONARY USER, HOMONYMY, LEMMA SELECTION, LEXICOGRAPHY, LINGUISTICS, MEANING, POLYSEME SELECTION, POLYSEMY, TEXT PRODUCTION, TEXT RECEPTION

Opsomming: Poliseemseleksie, lemmaseleksie en artikelseleksie. In die taalkunde en meer spesifiek die leksikale semantiek het polisemie en homonimie reeds heelwat aandag gekry. Die vasstelling van asook onderskeid tussen polisemie en homonimie is nie onproblematis nie. In die leksikografiese praktyk is daar tradisionele maniere waarop polisemie en homonimie aangebied en bewerk word. Hierdie artikel wys op benaderings in sowel die taalkunde as die leksikografie tot homonimie en polisemie. Voorbeelde uit die leksikografiese praktyk word aangebied. Daarna toon hierdie artikel aan dat die tradisionele aanbieding en bewerking van polisemie en homonimie in woordeboeke met teksbegrip as funksie nie werklik die teikengebruiker help in sy/haar

soektog na die verklaring van betekenis van 'n onbekende leksikale item nie. Daar word aangetoon dat verskillende woordeboeke dikwels dieselfde lemmakeuse het maar nie dieselfde poliseemkeuse nie. Dit is belangrik dat 'n woordeboek die korrekte koördinasie van 'n uitdrukking en sy betekenis moet bied. Gevolglik word 'n nuwe benadering in hierdie artikel voorgestel. Na aanleiding van kritiek uit sowel die taalkunde as die leksikografie word voorgestel dat daar in die leksikografiese praktyk, in die geval van woordeboeke vir teksbegrip, afstand gedoen word van die tradisionele onderskeid tussen verskillende homonieme asook van die aanbieding van verskillende polisemiese waardes van 'n woord in een artikel. Elke betekeniswaarde, hetsy die enigste betekenis van 'n uitdrukking hetsy 'n polisemiese waarde, moet die enigste aanduiding van betekenis in 'n artikel wees. Daar word geargumenteer ten gunste van 'n fokus op artikelkeuse met lemmakeuse en poliseemkeuse as onderafdelings.

Sleutelwoorde: ARTIKELKEUSE, BETEKENIS, HOMONIMIE, LEKSIKOGRAFIE, LEMMAKEUSE, POLISEEMKEUSE, POLISEMIE, TAALKUNDE, TEKSBEGRIP, TEKSPRODUKSIE, WOORDEBOEKGEBRUIKER

1. Introduction

In the treatment of lexical meaning in dictionaries lexicographers often adhere to a distinction between homonymy and polysemy, originating from the field of linguistics. This has a direct influence on the ordering and presentation of items giving the meaning of expressions in these dictionaries — irrespective of whether this expression is a single word, a multiword item, e.g. a fixed expression or a loan word combination, or a lexical item smaller than a word, e.g. a stem or an affix. The user is confronted by a lexicographic tradition according to which homonyms, that is lexical items with identical form but unrelated meanings, are entered as separate lemmas. Polysemy, where a single lexical item has different related senses, is treated by including all the senses in a single dictionary article with that lexical item as the lemma.

In a dictionary that has to satisfy a communicative function, the question arises whether adherence to the traditional treatment and presentation of homonymy and polysemy has any significance for the intended target user. This question will be addressed in the present paper. The emphasis will be on dictionaries used for text reception purposes but suggestions made in this paper may also be relevant to the other communicative functions, such as text production and translation.

2. Homonymy and polysemy

In the field of linguistics homonymy and polysemy are regarded as determining two different types of lexical ambiguity, cf. Lyons (1977: 550). In the study of lexical semantics and the presentation and treatment of lexical meaning in dictionaries both linguists and lexicographers have shown significant interest in polysemy and homonymy, cf. in this regard among others Zgusta (1971),

Lyons (1977), Jeffery (1982), Stock (1984), Geeraerts (1986; 2013), Gouws (1983; 1989) and more recently Bergenholtz and Agerbo (2014a, b, c). In linguistics and lexicography homonymy is usually regarded as two or more lexical items with the same form but synchronically unrelated meanings, e.g. *gum* (=secretion of some shrubs and trees) and *gum* (=flesh around roots of the teeth), whereas polysemy is seen as the different related meanings (or senses) of a single lexical item, e.g. *fence* (= a barrier or railing ... enclosing an area of ground // a large upright obstacle in ... showjumping // a guard ... in machinery) (*The Concise Oxford Dictionary*).

In their dictionaries lexicographers frequently adhere to the distinction between homonymy and polysemy. Homonyms are presented as separate lemmas, each with a distinguishing homonym marker (**gum**¹ and **gum**²) whereas the different senses of a polysemous lexical item, also referred to in this paper as different polysemes¹, are presented in a single article:

fence ...1. ...
 2. ...
 3. ...

From a linguistic point of view valid criticism against a too easy classification of words as being homonyms has been expressed by Lyons (1977: 550-569). Lyons indicates typical criteria introduced by linguists to distinguish between polysemy and homonymy. One such a criterion is the lexicographer's knowledge of the historical derivation of words. Homonyms are regarded as having developed from formally distinct lexical items. Lyons shows that the etymological criterion is not always decisive, among others because the historical derivation is not always certain. In addition, Lyons regards information about the origin of words as often irrelevant in the synchronic analysis of language.

A second criterion to which Lyons (1977: 551) refers and that is often used by linguists in distinguishing between polysemy and homonymy is relatedness versus unrelatedness of meaning. Lyons shows that there are problems in using this criterion of relatedness, e.g. the fact that relatedness of meaning can be regarded as a matter of degree. In this regard the intuition of native speakers may not be sufficient for a distinction between polysemy and homonymy.

Formal identity is a further criterion in the definition of homonymy (Lyons 1977: 557). This identity regards the phonological and the orthographic presentation of lexical items. He clearly argues that the condition of formal identity is not strong enough. Turning to lexicography Lyons (1977: 565) indicates that in traditional lexicographic practice a comparatively weak notion of syntactic identity is used when deciding whether two words are homonyms. Lyons (1977: 567) regards polysemy, a product of metaphorical creativity, as essential in the functioning of language whilst homonymy is not essential. Some of the suggestions Lyons makes to circumvent the problems of distinguishing between polysemy and homonymy will be discussed in a following section of this article. Important here is his critical approach to the traditional

distinctions between polysemy and homonymy.

Although seen as two types of ambiguity the traditional distinction between homonymy and polysemy is not always unambiguous — in terms of an Old Danish fixed expression it is an elastic tape measure. Lexicographers do not all agree on distinguishing criteria for polysemy and homonymy. Wahrig (1973) rejected meaning as a distinguishing criterion between polysemy and homonymy in favour of a more unambiguous grammatical criterion. For him homonymy prevails when one and the same orthographic word is used for different lexical words with the same word class, or when within the same word class different inflection paradigms occur. Although such an approach to homonymy might be clearer and can perhaps be employed in a less ambiguous way its lexicographic application can often lead to a collection of different meanings in a single dictionary article of a given word. Bergenholtz and Agerbo (2014c) not only question the criteria to distinguish between polysemy and homonymy but also the actual existence of homonymy and polysemy, albeit that their proposals maintain a traditional lexicographic presentation and treatment of homonymy and polysemy — because such a model "is closer to the solution that dictionary users are familiar with". Whether such an approach is convincing remains to be seen.

In this paper the existence of homonymy and polysemy as concepts in the field of linguistics is acknowledged. The assumption is also that the polysemy of a lexical item prevails when the meaning of that word is assessed in isolation. As soon as a word is introduced to actual language use, the context activates a single sense and the other senses disappear for that context. With the exception of intended or unintended ambiguity a word in any given context has only one meaning.

3. Lexical meaning and contextual meaning

Can a lexical word that is characterised by a specific content connected to a specific expression have more than one meaning? In this paper we will argue against such an approach. Many linguists and lexicographers would have a different opinion. Their question would rather be: How many meanings can a word have? This is for example seen in the title of the contribution of Louw (1995). Louw adheres to a traditional approach that distinguishes between contextual and lexical meaning and he only regards the latter type as relevant to lexicography and should therefore be accounted for in dictionaries. Louw (1995) argues rightly that only the meaning of a word should be presented in dictionaries. The different uses of a given word should not be elevated to and presented as the meaning of the word. Louw argues against an inflation of polysemous senses that do not represent the meaning of the word but merely its use. According to Louw such a distinction between meaning and use, more precisely between lexical and contextual meaning, makes it possible to reduce the number of meanings, i.e. polysemous senses, of a given word. Louw regards

such a reduction as helpful because a long list of polysemous senses can easily lead to dictionaries that lack user-friendliness. Louw argues against dictionaries that present the uses of a word as its different meanings:

They define usage rather than lexical meaning. ... It is of the utmost importance to distinguish lexical meaning from contextual meaning. ... However, if we fail to distinguish lexical meaning from contextual meaning we are bound to accept that whenever a word is used, a meaning comes from that usage. (Louw 1995: 358f)

This paper does not try to contradict Louw's distinction between contextual and lexical meaning or to argue against his approach that the use of a word should not be regarded as its meaning. This paper merely responds to Louw's question, i.e. "How many meanings can a word have?" It will be shown that the number of meanings of a word should not influence the way in which that word and those meanings are presented and treated in a dictionary.

Linguists do not agree on the nature and extent of polysemy. An approach that differs significantly from that of Louw (1995) is found in Panman (1982) who does not acknowledge the occurrence of either homonymy or polysemy in the case of lexical words. According to Panman homonymy and polysemy only prevail between words in context:

The present study argues that it is preferable to regard the phenomena as relations between word-tokens rather than between lexical items. (Panman 1982: 105)

An even more radical approach is found in Hanks (2000) and Hanks and Bradbury (2013). The latter contribution starts with the leading question: "Do words have meaning?" and the answer that is provided is a clear "No". Seemingly in contrast to Louw (1995) and more in agreement with Panman (1982) the basic assumption in Hanks and Bradbury is that a word only has meaning and can only have a meaning in a concrete context. According to Hanks and Bradbury (2013: 28) one cannot give a meaning to a lexical word in an isolated abstract construction.

This article maintains that scholars like Louw, Panman and Hanks and Bradbury primarily argue as linguists and not as lexicographers. A lexicographer is interested in the information needs and reference skills of the potential target user of a given dictionary. Whether a specific item in a dictionary can be regarded according to one or more linguistic theories as real meaning or not is not a question that has relevance for the user of the dictionary. A dictionary user needs help, and a lexicographer wants to provide help. Lexicographers achieve this by making a tool, a dictionary, available to the users. The potential dictionary user may for example not understand a word he or she reads in a text and then consults a dictionary for assistance to obtain an explanation of the meaning of the given word. Such items giving the meaning can be presented in different ways in different dictionaries, also because these dictionaries may differ with regard to their user types and may have fundamentally different usage situations. This contribution will primarily be concerned with the com-

municative situation where a mother-tongue speaker or a foreign language speaker encounters text reception problems.

In this article the authors do not try to show how an ideal item giving the meaning in a cognitive dictionary should look in a case of a dictionary that would like to inform as broadly as possible on the meaning of a word and how this meaning or these meanings can be related historically or systematically to other meanings. This is the topic in papers like that of Halas (2016) that concludes with a proposal for a comprehensive systematic presentation. Halas discusses the English verb *drop* of which the treatment contains six main meanings and twenty four submeanings or submeanings of a submeaning. In comparison to her breakdown of the meaning of this word the divisions within Wittgenstein's "Tractatus" look relatively simple! The practical value of this work is not easily detectable. The approach in this paper is that such minuscule divisions do not have a place in the lexicographic practice. To determine systematic knowledge about the word, this breakdown is so opaque that users will have extreme difficulties to obtain a clear overview of the coherence. The current paper rather looks at the lexicographic practice, especially in dictionaries for text reception, and at alternative ways in which meanings can be linked to expressions.

4. Polysemy in the lexicographic practice

Publications dealing with polysemy often use a single example or very few examples, often with a strong focus on verbs, cf. Louw (1995) and Halas (2016). Verbs frequently have numerous different uses and these uses, as indicated by Louw (1995), are listed in dictionary articles as senses to illustrate an apparent high occurrence of polysemy. The current paper also employs only a few examples but they are taken from different Danish as well as English dictionaries and the chosen lemmas are nouns and not verbs. The expressions represented by these lemmas are to a far lesser degree prone to a situation where the use of a word can be elevated to its meaning.

4.1 Polysemy in three Danish dictionaries

The following section contains a comparison of the selection of polysemes in three recent Danish monolingual dictionaries. Two of these dictionaries are polyfunctional. They strive to supply the user with assistance to solve information problems in different situations, e.g. with regard to text reception, text production and knowledge about words in the Danish language. This paper only focuses on the selection of polysemes from the perspective of text reception problems where only the item giving the meaning is of importance. Text reception is the only function of the third dictionary that has mother-tongue speakers of Danish with a reception problem as target user group. Monofunctional dictionaries like this one are very helpful as online products but also in

the printed format. The first printed Danish dictionary was such a monofunctional dictionary (although there had been an earlier printing of the first volumes of a Danish dictionary that was never completed). This monofunctional dictionary (Leth 1800) was compiled for young people who did not understand all the expressions in the religious texts they read. Yet the majority of Danish dictionaries since then have been polyfunctional, e.g. the *Politikens Nudansk Ordbog* (2005) and *Den Danske Ordbog* (2017), in contrast to the monofunctional *Den Danske Betydningsordbog* (2017). These three dictionaries do not display the same lemma selection and number of lemmas.

- NUD (*Politikens Nudansk Ordbog* 2005) has more or less 47 000 lemmas
- DDO (*Den Danske Ordbog* 2017) has 98 639 lemmas
- DDB (*Den Danske Betydningsordbog* 2017) has 118 212 lemmas

The comparison of the polyseme selection in these three dictionaries is done by means of a limited number of lemmas. These lemmas, as is the case in the discussion of English dictionaries in paragraph 4.2, come from the article stretches of the letters *s* and *t*. Although the problem of text reception is only discussed in more detail with regard to the first example, i.e. the lemma *stjernekygger* (= *stargazer*), a comparable approach was followed for the other examples but due to space restrictions the results have not been presented here. A question to be answered is what the meaning of *stjernekygger* is in the following sentences obtained from a Google search on 6 February 2017:

- (1) Hun ser sød ud, men hvis denne lille pige var en **stjernekygger** under fødslen, så har hun uden tvivl været noget af en udfordring for moderen. (= She looks sweet but if this little girl had really been a stargazer during her birth, she would undoubtedly have been a challenge to her mother.)
- (2) Stjernekyggere og romantikere har gode chancer for at se stjernesky, når stjerneskyssværmen Geminiderne er aktiv i perioden 4.–17. december. (= Stargazers and romanticists have good chances of seeing a shooting star when the swarm of shooting stars are active in the period between 4 and 17 December.)
- (3) Araberen, som den rødhårede pige red, var en typisk **stjernekygger**, som styrtede af sted med hovedet lige i vejret. (= The Arab on which the red-haired girl was riding was a typical stargazer, running to the front with its head lifted up.)
- (4) Den hvidkantede stjernekygger er svær at få øje på, når den har gravet sig ned i havbunden. Selv kan den nemt se de småfisk og skaldyr, der er så uheldige at svømme forbi dens vagtsomme blik. (The stargazer with white edges is difficult to find when it has buried itself in the sea sand. It can easily see small fish and seafood that has the bad luck to swim directly into its watchful sight.)

- (5) Stjernerne skaber opmærksomhed og tiltrækker en masse stjernekyggere, der håber at se en kendis på den røde løber. (The stars draw the attention and draw many stargazers to their close proximity, who hope to see a famous actor on the red carpet.)

The three dictionaries display the following polyseme selection for *stjernekygger* (= *stargazer*):

The NUD does not have this lemma.

The DDO includes this lemma and presents two polysemes:²

1. person that loves stars or likes to look at stars as a hobby, often with binoculars or a telescope.
2. child that is born with his face looking upwards and with the neck bent to the back.

The DDB has five polysemes:

1. person interested in stars and other celestial bodies and who observes them through appropriate binoculars; as well as professional astronomers and amateur astronomers.
2. baby that has turned at birth with his face to the top or not towards the womb of his mother that could cause difficulties during the birth because in this position the head cannot easily move through the pelvis of the mother.
3. horse that has a bad posture or has been trained incorrectly and has its head raised when being ridden.
4. person who goes to places where famous people are in order to see them and perhaps have contact with them.
5. family of fish with a typical appearance with the eyes on top of the head and with a big mouth directed upwards.

The more lemmas and the more polysemes a dictionary has, the better are the chances that a user with a text reception need will find the required assistance. NUD does not include the lemma *stjernekygger* at all and can therefore not be of real assistance. DDO has two polysemes, helping with examples (1) and (2) but not with the other three examples. The DDB includes the lemma and the polysemes that can explain all the different examples.

Already in this first example we can notice the tendency that is confirmed by the other examples: The bigger the lemma selection, the bigger the polyseme selection. This is interesting because one could have expected that dictionaries with fewer lemmas would present the most frequently used lemmas and their different meanings. Naturally, when saying this one has to negotiate the type of dictionary from which the lemmas have been selected but comparable dictionaries have been used for the selection of lemmas to be discussed in this section.

The following results can be obtained from the treatment of homonymy and polysemy with regard to the selected lemmas in the three dictionaries:

Lemma	NUD hom	NUD pol	DDO hom	DDO pol	DDB hom	DDB pol
sikkerhedslampe	0	0	0	0	0	4
stjernekipper	0	0	0	2	0	5
streamer	0	1	0	3	0	5
stringer	0	0	0	2	0	4
tabernakel	0	4	0	3	0	5
talent	2	3	2	4	2	3
Total	2	8	2	14	2	26

Sikkerhedslampe (safety lamp) only appears in one of the three dictionaries — the DDB that presents the following meanings:

1. lamp that burns with petroleum or another kind of inflammable liquid and of which the flame is protected in such a way that it cannot cause a fire, explosion or something similar in a mine.
2. lamp that hangs on a wall and lightens the environment when it registers a movement.
3. small lamp that a person carries or attaches somewhere, that flashes continuously or shines constantly so that other people can see an oncoming person or object, e.g. someone jogging in the dark.
4. lamp that is part of a security facility and that shines when the alarm is activated.

The word **sikkerhedslampe** (safety lamp) always refers to something regarding a lamp and security, but this is clear to anyone who speaks Danish. But he/she does not actually know exactly for what a security lamp is used in a specific instance and how it functions. This is only clear from a polyseme selection with different items giving the meaning as in the dictionary referred to above.

The lemma **streamer** (streamer) has been included in all three these dictionaries.

The NUD does not have different polysemes for the lemma but only the following single meaning:

a sticker that is used as advertisement for a company or a product and e.g. affixed to the back window of a car.

The DDO has three polysemes:

1. longish sticker with a radiant promotional text, often affixed to the inside of the back window of a car.
2. headline in a newspaper, written in capitals.
3. an artificial fly for fly fishing, often coloured silver or gold, with wings made of feathers or hair to imitate prey species of the fish.

The DDB has five polysemes:

1. longish poster to be affixed, especially on glass panes and car windows.
2. device that one can use, for different types of acoustic devices, e.g. cellular phone, computer or music device, so that the sound can be transferred directly without the use of a cable.
3. person or company that produces radio, music or films that are not transmitted by means of normal media but can be seen or heard in the Internet.
4. fly for fly fishing, consisting of a tiny hard core with a long tail of hair or feathers, to lure fish like trout by imitating small prey species when being moved through the water or on the surface of the water.
5. long cable pulled behind a boat that investigates the bottom of the sea, while powerful sound waves are transmitted that are reflected from the bottom of the sea and are registered by the cable when the reflecting wave reaches the surface of the water.

The lemma **stringer** (stringer) is not found in the NUD but has been included in both the other dictionaries.

The DDO has two polysemes:

1. journalist or other person providing information, often from abroad, without being fully employed, with the information used by one or more editorial desks as alternative source to their own correspondents.
2. long buttresses in the body of a boat or airplane.

The DDB has four polysemes:

- 1 beams on the long side of a ship or airplane to support the body and make it more stable.
- 2 journalist that does not work for a specific newspaper or medium but provides contributions for different media on a specific topic, often because he is abroad and reports on local matters.
- 3 cord through the mouth and gills of a fish that has been caught by a fisherman or diver, or must be held briefly before being set free again.
- 4 T-shirt or vest without sleeves but only thin straps over the shoulders that gives the user freedom of movement when used during training.

All three dictionaries include the lemma **tabernakel** (tabernacle).

The NUD has four polysemes but with a non-transparent numbering because two independent meanings do not have their own polysemy number but are only identified by means of a bullet. In spite of this unsystematic presentation all the meanings are given as separate polysemes below:

1. the transportable shrine of the Jews that especially played an important role during the period in the desert.
 - altar cupboard in a Catholic church where bread for the communion is kept.
 - a church of some Christian sects, e.g. the Mormons.
2. noise made by celebrating people.

The DDO has three polysemes:

1. sanctuary in the form of a tent that the Israelites took with them according to the Old Testament during their time in the desert.
2. altar cupboard in a Catholic church in which the Eucharistic bread is kept.
3. (place with) noise, festivity and commotion.

The DDB has five polysemes:

1. transportable shrine in a tent used by the Israelites during the desert trek from Egypt to Israel..
2. church of the Mormons.
3. small cupboard where the bread for the communion and the holy casks are kept in Catholic churches.
4. confusing variety of restless noisy people.
5. a fierce and hardly comprehensible discussion with input from different sides.

This example clearly shows that lemma selection is an overrated lexicographic problem. The selection of meaning is much more interesting in this regard as just as many meaning gaps can be detected as are the number of lemma gaps often discussed in dictionary criticism.

All three dictionaries have the lemma **talent** (talent).

NUD has two homonyms and in total three polysemes, with the exceptional way of giving a separate meaning for the first homonym albeit without an own number. We do not maintain this but rather present three meanings, i.e. also polysemes:

talent, Homonym 1

innate ability within a specific area

- a person who is talented

talent, Homonym 2

a weight and money unit used in ancient times

The DDO has two homonyms and in total four polysemes:

talent, Homonym 1

1. ability or gift within a specific area
2. person who possesses exceptional qualities

talent, Homonym 2

1. weight and money unit of different sizes that was used in ancient times in Babylon and Greece
2. new Greek weight unit, equal to 150 kg

The DDB has two homonyms and in total three polysemes:

talent, Homonym 1

- 1 innate ability to cope with a specific discipline or type of assignment with exceptional good results

2. person with superb abilities or qualities of whom one therefore has huge expectations
talent, Homonym 2
 old weight and money unit that was used in ancient times, among others in Babylon.

With regard to these dictionary articles with **talent** as lemma one can repeat the comment made with regard to **tabernakel**: Dictionaries do not differ that much when it comes to lemma selection; the more significant difference is on the level of polyseme selection.

4.2 Polysemy in six English dictionaries

For the purpose of this paper the following English dictionaries were used to look for the equivalents of the lemmas taken from the three Danish dictionaries:

- NODE (*The New Oxford Dictionary of English*) (1998) Roughly 86 000 lemmas.
- WIK (*Wiktionary*, Creative Commons Attribution/Share-Alike License) [Internet dictionary (Wordnik)] (2017) No lemma count.
- *WordNet 3.0* [Internet dictionary (Wordnik)] (2017) No lemma count.
- HID (*The Heritage Illustrated Dictionary of the English Language*) (1973). Roughly 68 500 lemmas.
- MWL (*Merriam-Webster's Advanced Learner's English Dictionary*) (2008) Roughly 38 500 lemmas.
- LSAD (*Longman South African School Dictionary*) (2007). Roughly 20 500 lemmas.

The following table shows a synopsis of the findings of polysemes:

LEMMA	NODE	WIK	WordNet 3.0	HID	MWL	LSAD
safety lamp	1	1	1	2	–	–
stargazer	3	2	3	2	1	–
streamer	2	6	4	4	2	1
stringer	3	8	4	5	1	–
tabernacle	4	9	3	5	3	–
talent	2	–	2	4	3	1
Total	15	26	17	22	10	2

Yet again, one has to negotiate the specific type of dictionary when comparing the number of polysemes presented in the different articles. As can be expected a school dictionary like the LSAD displays fewer polysemes because its cover-

age of both lexical items and their meanings is not as comprehensive as that of the other dictionaries. Whereas no lemma count is indicated for *WordNet* and *Wiktionary* a comparison of the NODE, HID and MWL shows, contrary to the results in paragraph 4.1, that the highest occurrence of polysemes does not prevail in the dictionary with the most lemmas (the NODE). This is not a significant issue in the current discussion, also because none of these dictionaries are monofunctional.

These dictionaries treat the selected lemmas as follows:

stargazer: With the exception of the LSAD all the other dictionaries have included a lemma *stargazer*. The meaning of this word is presented as follows:

The NODE has three polysemes:

1. an astronomer or astrologer
2. a horse that turns its head when galloping
3. a fish of warm seas that normally lies buried in the sand ...

WIK has two polysemes:

1. One who stargazes.
2. A perciform fish in the family Uranoscopidae.

WordNet gives the following three polysemes:

1. heavy-bodied marine bottom-lurkers with eyes on flattened top of the head
2. someone indifferent to the busy world
3. a physicist who studies astronomy

The HID has the following two polysemes:

1. an astronomer or astrologer
2. any of various marine bottom-dwelling fishes ...

MWL has a single meaning:

a person who looks at the stars; a person who studies astronomy or astrology

Yet again, the dictionaries have the same lemma but the polyseme selection differs considerably. A Google search on 16 March 2017 gives evidence of a significant number of occurrences of the word *stargazer* being used to refer to a certain fetal position. However, none of the dictionaries accounts for this meaning.

safety lamp: MWL and LSAD do not include this word as a lemma. It could well be because of a word bias that makes little provision for the inclusion of multiword lemmas.

The NODE, WIK and *WordNet* each has a single meaning for **safety lamp**:

NODE

a miner's portable lamp with a flame which is protected, typically by wire gauze ...

WIK

A miner's lamp designed to avoid explosion by enclosing the flame in fine wire gauze.

WordNet 3.0

an oil lamp that will not ignite flammable gases

The HID has two polysemes:

1. a miner's lamp with a protective wire gauze ...
2. any specially protected lamp

The same core meaning is given in all the dictionaries.

Streamer is the only word in the list that has been included all the selected dictionaries. It is treated as follows:

NODE has the following two polysemes for this word, albeit that the second "polyseme" actually gives another variant form:

1. long narrow strip of material used as a decoration or symbol
2. short for **TAPE STREAMER**

WIK gives the following six polysemes:

1. A long, narrow flag, or piece of material used or seen as a decoration.
2. A newspaper headline that runs across the entire page.
3. A data storage system, mainly used to produce backups, in which large quantities of data are transferred to a continuously moving tape.
4. In fly fishing, a variety of wet fly designed to mimic a minnow.
5. One who searches for stream tin.
6. A stream or column of light shooting upward from the horizon, constituting one of the forms of the aurora borealis.

WordNet also has four polysemes:

1. a newspaper headline that runs across the full page
2. long strip of cloth or paper used for decoration or advertising
3. a long flag; often tapering
4. light that streams

The HID has four polysemes:

1. a long narrow flag, banner or pendant
2. any long narrow pendant, strip of material
3. a shaft or ray of light extending upward from the horizon
4. a newspaper headline that runs across a full page

MWL displays two polysemes and the LSAD a single item giving the meaning:

MWL

1. A long, narrow piece of coloured paper or plastic that is used as a decoration
2. A long, narrow flag

LSAD

a long thin piece of coloured paper for decorating a place.

The different orderings of the same polysemes in the different dictionaries yet again confirm the often unsystematic treatment of polysemy and the difference in the polyseme selection, with senses that apparently have a high frequency of use not qualifying for inclusion in some dictionaries, cannot be explained.

Stringer occurs in all dictionaries but the LSAD.

It has three polysemes in the NODE:

1. a longitudinal structural piece in a framework, especially that of a ship or aircraft
2. *informal* a newspaper correspondent, not on the regular staff of a newspaper, ...
3. a stringboard [=a supporting timber or skirting in which the ends of the steps in a staircase are set]

WIK has no less than eight polysemes for this word:

1. Someone who threads something; one who makes or provides strings, especially for bows.
2. Someone who leads someone along.
3. A horizontal timber that supports upright posts, or supports the hull of a vessel.
4. A freelance correspondent not on the regular newspaper staff, especially one retained on a part-time basis to report on events in a particular place.
5. Wooden strip running lengthwise down the centre of a surfboard, for strength.
6. A hard-hit ball.
7. A cord or chain, sometimes with additional loops, that is threaded through the mouth and gills of caught fish.
8. A libertine; a wencher.

WordNet has the following four polysemes:

1. a worker who strings
2. a member of a squad on a team
3. brace consisting of a longitudinal member to strengthen a fuselage or hull
4. a long horizontal timber to connect uprights

HID has five polysemes with the second polyseme subdivided into two sub-senses:

1. a person or thing that strings
2. *architecture* a. A long heavy horizontal timber used for any of several connective or supportive purposes b. A stringboard

3. a lengthwise timber used to support rails
4. a member of a specified string or squad on a team
5. a part-time representative for a news publication ...

The MWL has but a single meaning:
a journalist who is not on the regular staff of a newspaper ...

The subdivision presented in the second polyseme given in the HID leads to a question regarding the selection of polysemes and the way in which polysemes are distinguished. Does the occurrence of different senses of the word *stringer* in the field of architecture imply that they are subsenses and not senses? In addition, the same comments made with regard to the lemma *streamer* also apply here. Although the dictionaries do not differ with regard to their lemma selection, there are vast differences regarding the polyseme selection. The inclusion of a word as lemma does not guarantee a successful consultation procedure for a user looking for a specific meaning. Linking lemmas to different meanings is not done in an exemplary way.

The learners' dictionary LSAD is the only one that does not include the lemma **tabernacle**.

The NODE has four polysemes:

1. (in biblical use) a fixed or movable habitation, typically of light construction
2. a meeting place for worship ...
3. an ornamented receptacle or cabinet ...
4. a partly open socket or double post on a sailing boat's deck ...

WIK has no less than nine polysemes — the highest number of polysemes for any one of the selected lemmas:

1. Any temporary dwelling, a hut, tent, booth.
2. The portable tent used before the construction of the temple, where the shekinah (presence of God) was believed to dwell.
3. Transferred to the Jewish Temple at Jerusalem as continuing the functions of the earlier tabernacle.
4. Any portable shrine used in heathen or idolatrous worship.
5. A sukkah, the booth or 'tabernacle' used during the Jewish Feast of Sukkot.
6. A small ornamented cupboard or box used for the reserved sacrament of the Eucharist, normally located in an especially prominent place in a Roman Catholic Church.
7. A temporary place of worship, especially a tent, for a tent meeting, as with a venue for revival meetings.
8. Of any abode or dwelling place, especially of the human body as the temporary dwelling place of the soul, or life.
9. A hinged device allowing for the easy folding of a mast 90 degrees from perpendicular, as for transporting the boat on a trailer, or passing under a bridge.

WordNet has three polysemes:

1. the Mormon temple
2. (Judaism) the place of worship for a Jewish congregation
3. (Judaism) a portable sanctuary in which the Jews carried the Ark of the Covenant on their exodus

HID has five polysemes and the MWL three:

HID

1. a. The portable sanctuary in which the Jews carried the Ark of the Covenant through the desert. b. The Jewish temple.
2. A case or box on a church altar containing the consecrated host and wine of the Eucharist.
3. A place of worship ...
4. A niche for a statue or relic.
5. A boxlike support in which the heel of a mast is stepped.

MWL

1. A place of worship ...
2. A box in which the holy bread and wine are kept
3. *The Tabernacle* a small moveable tent that was used as a place of worship by the ancient Israelites

WIK is the only dictionary that does not include the lemma **talent**.

The NODE has two polysemes:

1. natural aptitude or skill
2. a former weight and unit of currency ...

WordNet also has two polysemes:

1. a person who possesses unusual innate ability in some field or activity
2. natural abilities or qualities

HID has the most polysemes, no less than four, for this word:

1. A mental or physical aptitude ...
2. Natural endowment or ability of a superior quality.
3. Gifted people collectively.
4. A variable unit of weight and money ...

The MWL has three polysemes and the LSAD has one:

MWL

1. A special ability that allows someone to do something well
2. A person or group of people with a special ability to do something well
3. *Brit. Slang* People who are sexually attractive

LSAD

natural ability to do something well.

A look at the treatment of polysemy in these dictionaries emphasises the consistency in lemma selection but the lack of consistency in polyseme selection. The lesser occurrence of homonyms, in comparison with the Danish dictionaries, should also be noted.

5. Polyseme selection and lemma selection = article selection

In dictionaries that have to assist mother-tongue speakers as well as foreign language speakers with reception problems the most helpful, i.e. the most user-friendly, dictionary will be the one with the most lemmas and the most polysemes.³ With regard to mother-tongue speakers it especially applies to those with a good command of both written and spoken language that experience a text reception problem when confronted with a new word, a new meaning, a rare word or an infrequently used meaning. Including as many lemmas and as many polysemes as possible could lead to a space problem in printed dictionaries, and it should also be considered that comprehensive, especially multi-volume, dictionaries are not easy to handle and are very expensive. But the space problem does not prevail in e-dictionaries. Rapid access could be a problem but this problem can be solved.

A typical approach to polysemy is that the most frequently used words will have the most meanings. This is true when departing from the point of view that a single word has different meanings. One could then argue that a dictionary with few lemmas would usually have many more polysemes compared to a dictionary with many lemmas. The preceding analysis does not support this point of view — rather the opposite. With certain exceptions it was found that the more lemmas in a dictionary, the more polysemes are given for these lemmas. This may be helpful in different types of learners' dictionaries where only the high frequency words and the high frequency meanings are presented. In dictionaries for mother-tongue speakers and advanced learners an approach where a dictionary with many lemmas also has many polysemes is quite surprising and not as helpful to its target users as one would have expected.

Many dictionaries still maintain the traditional distinction between homonyms and polysemes. This tradition is not only dubious but it also does not concur with the understanding of a word. When one assumes that a lemma represents a lexical word or a lexicalised fixed expression and also assumes that a lexical word is a linguistic sign, it does not support the continuation of the current system. One can just look at the different meanings of *stjernerkigger* (*stargazer*). These different meanings have so little in common that one cannot work with the idea of a linguistic sign with solidarity between the expression and its contents.

Both the relevant discussions in linguistics and theoretical lexicography and the presentation and treatment of homonymy and polysemy in the lexicographic practice show ambiguity in the different perspectives on homonymy and polysemy as two forms of lexical ambiguity. From the field of theoretical lexicography Bergenholtz and Agerbo (2014c) offer their criticism whilst Lyons (1977) deals with uncertainties in linguistics regarding especially the problem of a distinction between homonymy and polysemy. Consequently one can argue in favour of the abolishment of the traditional distinction between homonymy and polysemy. The terms *lemma selection* and *polyseme selection* could rather be complemented by the more encompassing term *article selection*. In the lexicographic practice article selection could proceed in two subsequent steps: (1) lemma selection, that could also be referred to as the *selection of expressions* and (2) polyseme selection, also to be referred to as *selection of meaning*. Each article has but one lemma which is the name and guiding element of the article and it combines different variant forms of the expression with the same contents, e.g. all the inflexional forms of a specific lexical word/lexeme. The meaning selection determines what polyseme should be linked with the expression. This paper argues in favour of an alternative approach compared to the one prevailing in many current dictionaries.

6. Towards an alternative approach

Lyons (1977: 553-554) does not try to find a solution to the problem of distinguishing between homonymy and polysemy but rather finds ways of circumventing it. The two ways he proposes are "to maximise homonymy by associating a separate lexeme with every distinct meaning" — and this procedure he applies to occurrences of both homonymy and polysemy — and secondly to maximise polysemy. This implies "that distinctions of sense can be multiplied indefinitely" and will lead to an increased inclusion of all meanings and senses of a lexical item in a single dictionary article.

From the perspective of a dictionary user with a text reception problem the nature and theoretical status of the distinction between homonymy and polysemy is of little significance. Such a user wants to find the meaning of an unfamiliar word and in order to succeed in such a dictionary consultation this user has to find the specific expression — whether a word, a multiword unit or a lexical item smaller than a word. Whether the meaning is presented as one of the senses of a polysemous word, as a meaning of a homonym or as the only meaning allocated to a lemma that is one of a series of similar expressions does not determine the satisfaction of the user's need. Important is that the user finds the expression and finds the meaning.

The lexicographer has to present an article selection that allocates a specific meaning to a specific expression, presented as lemma of the article. The typical dictionary user is not supported or deterred by a system where lemmas are marked as homonyms — this is an insignificant lexicographic tradition that

can be abolished without diminishing the successful satisfaction of text reception needs. But also a maximisation of polysemy does not add value to the text reception solutions of a dictionary. Yet again the user is not interested in relations between senses of a given expression but rather in a specific meaning of a specific expression.

A new approach, relying on the groundwork done by Lyons (1977) in the field of linguistics and Bergenholtz and Agerbo (2014c) in the field of lexicography, could be to maximise article selection by linking each expression as lemma but to a single meaning — whether this meaning is a polyseme or a meaning not related to any other meaning. The suggestions by Lyons to maximise homonymy or polysemy are given within a linguistic frame; the idea to maximise article selection is directed at the lexicographic practice.

Allocating a single meaning to a lemma implies a much more simplified article. The major simplification lies therein that the new approach prevents a proliferation within an article into numerous senses. By restricting the items giving the meaning to a single one per article the user with a text reception need has no problem in quickly linking an expression to the relevant meaning — even though the maximisation of articles leads to the inclusion of different lemmas with the same form. These simplified articles could enhance the access of users to meaning.

Although the existence of the notion of polysemy is not contested, its lexicographic value is questioned. A new approach links expressions to single meanings without being hampered by questions regarding relatedness and non-relatedness between different meanings expressed by words with the same form. Introducing a simplified article structure by moving away from the nesting of polysemes in a single article creates the space and opportunity to present a slightly more comprehensive item giving the meaning. This could further enhance the text reception function of the given dictionary.

7. Some consequences

A solution consisting of a maximisation of articles and abolishing the distinction between polysemy and homonymy in the lexicographic practice will necessarily lead to an increase in the number of articles but not as massive an increase as one would have thought.

Den Danske Betydningsordbog (2017) contains "only" 222 instances where a single lemma has ten or more polysemes. This represents 0,19% of all articles in this dictionary. The spreading of polysemes is as follows:

Lemmas	lemmas with polysemes	polysemes
118.092	24.012 (20.33%)	66.248 (=2.76 polysemes (calculated for articles that do contain polysemes))

When looking at different part of speech classes it is interesting to note that a

high number of polysemes can be found in the articles of certain verbs:

verbs	103 (0.19% of all verb lemmas)
nouns	59 (0.07% of all noun lemmas)
adjectives	38 (0.24% of all adjective lemmas)
adverbs	15 (1.39% of all adverb lemmas)
prepositions	15 (9.38% of all preposition lemmas)
abbreviations	1 (0.06% of all abbreviation lemmas)

The relative occurrence of polysemes in the articles of verbs is even bigger if one looks at lemmas with a higher number of polysemes

	20 or more polysemes	30 or more polysemes	40 or more polysemes	50 or more polysemes	60 or more polysemes
all lemmas	47	24	14	8	4
verbs	31	19	10	7	4
nouns	2	0	0	0	0
adjectives	3	1	0	0	0
adverbs	10	3	4	1	0
prepositions	1	1	1	0	0
abbreviations	0	0	0	0	0

Because it is not possible to have an automatic count of the distribution of polysemes a non-representative count has been made by taking 20 x 100 lemmas from 20 different places in the alphabetical sorting. This shows that in the database for Danish monolingual dictionaries one finds more occurrences of polysemy for nouns than for other part of speech classes (the result for adverbs and prepositions gave less than 1% and is therefore not in the list):

	lemmas according to part of speech classes	lemmas with polysemes
nouns	71.84%	61%
proper names	3.10%	2%
verbs	8.15%	11%
adjectives	13.33%	16%
abbreviations	1.41%	10%

It is also clear from these numbers that there is a definite relation between the number of lemmas and their part of speech classes — with abbreviations as an exception. The number of polysemes for nouns is slightly less compared to the total number of nouns as lemmas. But these numbers do show that polysemy is especially relevant for nouns, whereas a huge number of polysemes occur at a small number of verbs. Of the 118.092 articles in *Den Danske Betydningsordbog* (2017) 25 012, i.e. 20.33%, display polysemes. The number of articles in a dictionary without homonymy and polysemy will therefore rise but not too much. Such an approach will also give the user a true indication of the extent of a dictionary because, as seen from the examples given in this article, two dictionaries could have selected a given number of lemmas but with completely different numbers of meanings for these lemmas.

Four of the English dictionaries used in this paper also show that for a random count of 100 lemmas the average number of polysemes per lemma where polysemy prevails is only 3,192 and only 31,25% of all lemmas display polysemes.

LEMMA	NODE	HID	MWL	LSAD	Total/average
Lemmas with polysemes (100 lemmas)	21	43	35	26	125
Polysemes	51	161	112	75	399
Average polysemes per polysemous lemma	2,4	3,7	3,2	2,8	3,192
Most polysemes	4	13	9	9	8,75

This yet again emphasises the fact that an abolishment of homonymy and polysemy in dictionaries in favour of a maximisation of articles will not increase the number of articles too drastically. In a printed dictionary the result will be manageable and in online dictionaries this approach will cause no problems at all.

The type of discussion presented in this paper with regard to polysemy could be expanded to other semantic relations, e.g. synonymy. Where *polysemy* and *polyseme* are terms that are used to reflect something of the relation between an expression and different meanings, the terms *synonymy* and *synonym* are used to reflect on the relation between a single meaning and different expressions. The suggested treatment of polysemy where each polyseme has its own article, coincides with the treatment of synonymy where the same meaning can be repeated in different articles introduced by different lemmas that are synonyms. In certain types of dictionaries synonyms can also be used as the item giving the meaning of another word. This topic will not be discussed in this paper. Bergenholtz and Gouws (2012) offer more in this regard.

8. In conclusion

In this paper the lexicographic presentation of homonymy and especially polysemy was discussed from the perspective of a dictionary for text reception purposes. It could be argued that a similar approach can be employed in dictionaries for the other communicative functions, i.e. text production and translation, but not in a dictionary with a cognitive function. There homonymy and polysemy may have a role to play, albeit that lexicographers should take cognizance of the suggestions made by Lyons (1977).

The approach suggested in this paper can be regarded as a theoretically based suggestion for the lexicographic practice. The needs and reference ease of the target user group should be a determining factor in each and every decision regarding the selection, presentation and treatment of items in a dictionary. This approach will lead to a noticeable increase in the number of articles this will not be a problem in online dictionaries, the default medium of modern-day lexicography. The simplification of dictionary articles with only a single item giving the meaning in each article shows that a text reception dictionary does not have to maintain the shaky distinction between homonymy and polysemy. The lack of such a distinction is no loss for lexicography if one considers the limited help the traditional lexicographic presentation of homonymy and polysemy has offered the user in need of text reception assistance.

Endnotes

1. In this paper the term *polyseme* refers to a linguistic expression that has more than one meaning and to each one of those meanings and the rest of the partial articles that has the same expression as lemma.
2. Henceforth only the English translations of the Danish polysemes will be given.
3. It may differ in a dictionary for text production purposes, especially for foreign language users, but this paper is only concerned with text reception.

Acknowledgement

This work is based on research supported in part by the National Research Foundation of South Africa (Grant specific unique reference number (UID) 85434). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

References

- Allen, R.E. (Ed.). 1990^s. *The Concise Oxford Dictionary*. Oxford: Clarendon Press.
- Bergenholtz, Henning et al. (Eds.). 2017. *Den Danske Betydningsordbog*. Odense: Ordbogen.com (www.ordbogen.com).

- Bergenholtz, Henning and Heidi Agerbo.** 2014a. Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries. *Hermes. Journal of Language and Communication in Business* 52: 125-139.
- Bergenholtz, Henning and Heidi Agerbo.** 2014b. Extraction, Selection and Distribution of Meaning Elements for Monolingual Information Tools. *Lexicographica* 30: 488-510.
- Bergenholtz, Henning and Heidi Agerbo.** 2014c. There is No Need for the Terms Polysemy and Homonymy in Lexicography. *Lexikos* 24: 27-35.
- Bergenholtz, Henning and Rufus H. Gouws.** 2012. Synonymy and Synonyms in Lexicography. *Lexicographica* 28: 309-335.
- Den Danske Ordbog. Moderne dansk sprog.* 2017. <http://ordnet.dk/ddo>. (Accessed February 2017.)
- Geeraerts, Dirk.** 1986. *Woordbetekenis: een overzicht van de lexicale semantiek*. Leuven/Amersfoort: Acco.
- Geeraerts, Dirk.** 2013. The Treatment of Meaning in Dictionaries and Prototype Theory. Gouws, Rufus H. et al. (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography*: 487-495. Berlin: Mouton de Gruyter.
- Gouws, Rufus H.** 1983. Polisemie as leksikografiese probleem. Sinclair, A.J.L. (Ed.). 1983. *G.S. Nienaber — 'n Huldeblyk*: 199-205. Bellville: UWK Drukkery.
- Gouws, Rufus H.** 1989. *Leksikografie*. Cape Town: Academica.
- Halas, Ana.** 2016. The Application of the Prototype Theory in Lexicographic Practice: A Proposal of a Model for Lexicographic Treatment of Polysemy. *Lexikos* 26: 124-144.
- Hanks, Patrick.** 2000. Do Word Meanings Exist? *Computers and the Humanities* 34: 205-215.
- Hanks, Patrick and Jane Bradbury.** 2013. Why Do We Need Pattern Dictionaries (and What is a Pattern Dictionary, Anyway)? *Kernerman Dictionary News* 21: 27-31.
- Jeffery, C.D.** 1982. An Application of Firthian Principles to the Analysis of Polysemous Lexical Items. Van Rensburg, M.C.J. (Ed.). 1982. *Kongresreferate: 18de Jaarlikse Kongres van die Linguistevereniging van Suider-Afrika, Rhodes Universiteit, Grahamstad*: 113-126. Bloemfontein: Universiteitsdrukkers UOVS.
- Leth, Jens.** 1800. *Dansk Glossarium. En Ordbog til Forklaring over det danske Sprogs gamle, nye og fremme Ord og Talemaader for unge Mennesker og for Ustuderede. Et Forsøg. Med en Fortale af Professor Rasmus Nyerup*. København: Trykt paa Hofboghandler Simon Poulsens Forlag hos Bogtrykker Morthorst's Enke & Comp.
- Louw, Johannes P.** 1995. How Many Meanings to a Word? Kachru, Braj and Henry Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary. Studies in Honor of Ladislav Zgusta*: 357-365. Tübingen: Niemeyer.
- Lyons, John.** 1977. *Semantics*. Cambridge: Cambridge University Press.
- Morris, William.** 1973. *The Heritage Illustrated Dictionary of the English Language*. Boston, Massachusetts: American Heritage Publishing Company.
- Panman, Otto.** 1982. Homonymy and Polysemy. *Lingua* 58: 105-136.
- Pearsall, J. (Ed.).** 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Perrault, Stephen, J. (Ed.).** 2008. *Merriam-Webster's Advanced Learner's English Dictionary*. Springfield, Massachusetts: Merriam-Webster.
- Politikens Nudansk Ordbog.* 2005. 19. udg. København: Politikens Forlag.

- Stock, Penelope F.** 1984. Polysemy. Hartmann, Reinhard R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 131-140. Tübingen: Max Niemeyer.
- Summers, Delia (Ed.)**. 2007. *Longman South African School Dictionary*. Harlow: Pearson Education.
- Wahrig, Gerhard**. 1973. *Deutsches Wörterbuch*. Gütersloh/Berlin/München/Wien: Bertelsmann Lexikon Verlag.
- Wiktionary*, Creative Commons Attribution/Share-Alike License. [Internet dictionary (Wordnik)] (2017).
- WordNet 3.0* [Internet dictionary (Wordnik)] (2017).
- Zgusta, Ladislav**. 1971. *Manual of Lexicography*. The Hague: Mouton.

Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated?*

Monika Biesaga, *The Institute of the Polish Language at the
Polish Academy of Sciences, Cracow, Poland (monika.biesaga@ijp.pan.pl)*

Abstract: The latest research conducted by De Jong (2014) indicated that only a few European academic dictionaries, among over one hundred and fifty, use pictorial illustrations. On the other hand, it has been mentioned several times in the literature (related to both printed and electronic reference works) that visual content makes a general dictionary more attractive for the user. Therefore a new methodology, enhancing graphical strategy in lexicography, would be proposed. Firstly, a thematic division of meanings will be applied to the authentic multimodal explanations taken from two general dictionaries (verbal definition and graphical facility). As a result, commonly illustrated thematic fields will be shown along with the ones used by lexicographers in a restrained way, as well as the fields left without illustrations for some reasons. The exemplary meanings from these latest, "orphan" groups will be confronted with the query results from the authentic multimodal corpora. Such a procedure will give us some more detailed information about the obstacles connected with illustrating this particular set of senses (for example multimodal denotation and connotation problems).

Keywords: MULTIMODALITY, GENERAL LEXICOGRAPHY, MONOLINGUAL DICTIONARIES, PICTORIAL ILLUSTRATION, THEMATIC FIELDS, IDEOGRAPHY, CONNOTATION, DENOTATION, ONOMASIOLOGY, MULTIMODAL LEXICOGRAPHY

Résumé: Tradition du Dictionnaire versus Corpora d'Images: Quels Champs Thématiques du Vocabulaire Doivent Être Illustrés? Les travaux de recherche les plus récents conduits par De Jong (2014) indiquent que seul très peu de dictionnaires académiques en Europe, parmi un peu plus de cent cinquante, utilisent des illustrations picturales. Cependant, il a été plusieurs fois mentionné dans la littérature lexicographique (s'agissant à la fois des ouvrages de référence sur papier comme des ouvrages de références électroniques) que le contenu visuel rend un dictionnaire général plus attrayant pour l'utilisateur. De ce fait, le présent article tente de proposer une nouvelle méthodologie en vue de renforcer une stratégie graphique en lexicographie. Premièrement, une catégorisation thématique des sens des mots s'appliquera aux explications multimodales authentiques extraits de deux dictionnaires généraux (définition verbale et provision

* This article was presented as a paper at the COST ENL WG3 meeting *Between Corpora and Dictionaries/Crowdsourcing and Gamification*, held in Budapest, Hungary, 24–25 February 2017.

graphique). Conséquemment, des champs thématiques habituellement illustrés seront exposés, de même que ceux utilisés de façon restreinte par les lexicographes et ceux souvent laissés sans illustration pour diverses raisons. Des exemples des sens de ce groupe "orphelin" seront confrontés aux résultats d'une investigation des corpora authentique multimodaux. Cette procédure permettra d'obtenir des informations plus détaillées sur les obstacles qui empêchent d'illustrer un ensemble particulier des sens lexicaux (par exemple les problèmes de dénotation et connotation multimodales).

Mots-clés: MULTIMODALITE, LEXICOGRAPHIE GENERALE, DICTIONNAIRES MONO-LINGUES, ILLUSTRATIONS PICTURALES, CHAMPS THEMATIQUES, IDEOGRAPHIE, CONNOTATION, DENOTATION, ONOMASIOLOGIE, LEXICOGRAPHIE MULTIMODALE

Introduction

In this paper various issues related to the inclusion of graphical illustration in general dictionaries will be discussed. Firstly, common lexicographical doubts concerning this matter would be characterized. A short overview of the most important theoretical achievements will be also given. Since I strongly believe that this visual content is still somehow underestimated the aim of this paper will be to present the ideographic groups of meanings which were subjected to visual presentation in dictionaries and confront them with the ideographic groups which were left behind. In order to recognize the reasons for such state-of-the-art in lexicography exemplary meaning-illustrating screenshots from the free media repositories will be displayed. As a result it will be indicated why some of the meaning groups should be illustrated (but they are not) and why the other ones are pictorially resistant (but could be subjected to graphical presentation, for example in social networks).

State-of-the-art

The latest research conducted by Gerbrich de Jong (2014) indicated that only a few European academic dictionaries, among over one hundred and fifty, use pictorial illustrations (drawings, photographs, schemes etc. incorporated in the entry in order to illustrate the meaning¹). On the other hand, it has been mentioned several times in the literature related to both printed and electronic reference works (for example Klosa 2016: 516, Kwaśnicka-Janowicz 2007: 163), that visual content makes a general dictionary more attractive for the user. This statement was even more accurate in the Internet era in which multimodal communication became an omnipresent standard. So why are we as lexicographers still so reluctant when it comes to this kind of content? Here one could point out a few more important factors.

When thinking about pictorial illustration, one should bear in mind that traditional lexicography distinguishes between two kinds of reference works:

dictionaries and encyclopedias, and therefore two kinds of knowledge that can be incorporated into the entry: linguistic knowledge and encyclopedic knowledge (Lara 1989, Burada and Sinu 2016: 61-62). It has been a great concern for lexicographers to exclude from dictionaries (especially definitions) additional and unnecessary scientific knowledge, not vital for meaning recognition. Pictorial illustrations were considered typical of encyclopedias and encyclopedic dictionaries (Hartmann and James 1998: 48-49, Hupka 1989: 704). Their function was not to help during the process of definition comprehension but to give additional information. That is why pictorial illustration, as a part of the scientific paradigm in lexicography, was excluded from many general dictionaries or limited, for example in learner's dictionaries, to a certain number of items that followed special criteria. What is more, technical and financial matters also influenced such state-of-the-art lexicography.

Apart from the lack of basic methodology in the general semiotic theory itself (how to analyze and describe multimodal communication) there are lexicographically-specific theoretical gaps that we have to face sooner or later. Without solving these problems, our everyday craft, connected with the inclusion of illustrations, will still be scientifically perceived as unstable and, to some extent, subjective.

During the last few years researchers have analyzed the problem of the semantic relationship between verbal information and pictorial illustration. That led them to two conclusions. Firstly, if we want to add a visual tool to our verbal definition (equivalent) a new coherent entity should come into existence. Kemmer (2014) calls it a symbiosis between verbal and pictorial facilities, Liu (2015) advocates a semantic ecology of multimodal definition, created by verbal and visual lexicographical tools.

Secondly, although particular typologies vary (for example Gangla 2001: 30-31, Liu 2015: 217) the basic distinction is based on mutual correspondence (complementarity) or the one-sided redundancy of knowledge given. This latest statement, though very general, should be seen as vaguer when one realizes that it was based on the analysis of randomly chosen multimodal dictionary explanations.

In the literature related to pictorial illustrations, a language meaning system is never perceived as a whole. What one can find are either random meanings, analyzed in a general way, either some groups of senses, indicated by dictionary editors (LDOCE analysis, see: Lew 2010, Stein 1991: 106) or as intuitively perceived by the author (Langridge 1998: 72-73). What is missing is a language semantic structure divided into the meaning groups that will approximately cover the whole lexical system (compare Casares 1959, Chapman 1977, Hallig and Von Wartburg 1963). In other words, it is not sufficient to indicate some thematic groups when one does not see which fields are missing and when one does not have the whole semantic picture.

In order to fill this gap, an already existing complete thematic division of meanings should be taken into account. Such classification was created by

B. Batko-Tokarz (2008) for the Great Dictionary of Polish project. Since 2007, this proposition has been used to divide meanings thematically from the contemporary Polish language (single words and discontinuous units). What is even more important, is that this classification was revised in 2013, after a few years of implementation of the (almost) entire vocabulary. Therefore it can be seen as reliable and apt for many languages, sharing the same cultural linguistic system.

The preliminary rule of this proposition is to distinguish thematic vocabulary and athematic lexicon. In the second, basic step thematic words and discontinuous units are divided into seven general thematic groups:

- a. Human as a Physical Being,
- b. Human as a Psychical Being,
- c. Everyday Life,
- d. Human in Society,
- h. Human and Technology,
- i. Human and Nature,
- j. Physical Categories.

Each of the main groups has its own subgroups. According to these semantic characteristics each meaning can be assigned to a more shallow or deeper thematic group. Also assignment to two different thematic groups is possible, obviously if a certain meaning requires this kind of procedure (amphora 'type of vessel' should be included in two different semantic fields: Closest Environment — subfield of Everyday Life; Tradition and Religion — subfield of Human in Society).

Scientific Procedure

In this text a new methodology, enhancing graphical strategy, will be proposed. What is important, it will be immediately tested on already existing two free-media corpora (Wikimedia Commons, Pixabay) that can be used in our lexicographical work².

For the purpose of this paper the thematic division mentioned in the introduction will be used in a slightly modified version (see Addendum 1). Some of the semantic fields have appeared, when confronted by authentic multimodal material, too detailed (too granular, itemized ideographic fields were unified). It does not automatically follow that this semantic classification in its original form is unsuitable for strictly verbal explanations.

Our thematic division of meanings will be applied to the authentic multimodal explanations taken from general dictionaries (verbal definition and graphical facility). As a result, commonly illustrated thematic fields will be shown along with the ones used by lexicographers in a restrained way, as well as the fields left without illustrations for some reasons.

The exemplary meanings from these latest, "orphan" groups will be confronted with the query results from the authentic multimodal corpora. Such a procedure should give us some more detailed information about the obstacles connected with illustrating this particular set of senses.

In order to make this comparison more reliable the illustrated word meanings from two different reference works were included:

- a. one printed dictionary: *Ilustrowany słownik języka polskiego (The Illustrated Dictionary of Polish)*, ed. by Elżbieta Sobol, 1999, hereafter referred to as IDP;
- b. one Internet dictionary: *Merriam-Webster Dictionary* (based on *Merriam-Webster's Collegiate Dictionary. Eleventh Edition*, published in 2003), hereafter referred to as MW.

What differs between these two dictionaries is a pictorial coverage of the lexicon. While Polish dictionary contains 40,000 entries and 1,800 illustrations (a medium-sized reference work, 4.5 percent of the entries with pictorial facilities), American dictionary contains 165,000 entries and 1,000 illustrations (a large-sized reference work, 0.6 percent of the entries with graphical facilities). To balance the numbers of analyzed entries in IDP (smaller dictionary) and MW (bigger dictionary), all of the entries starting with the letters A and B in the Polish dictionary and all of the entries starting with the A letter in the American reference work were taken into account (the MW website does not offer the possibility of displaying all illustrated entries, therefore the researcher must search for the word list manually).

Analyzing Dictionaries

To some extent, these dictionaries, reflecting contemporary lexicon taken from two different languages, are similar. They are monolingual and general in their nature. In both cases the publishing house also released an encyclopaedia (*Encyclopaedia Britannica* and *Encyklopedia PWN*). As we will later see, this encyclopaedic touch is visible in analyzed multimodal explanations. However, the nature of the exact relation between the illustrated entries in dictionaries and their equivalents in encyclopaedias should be seen as a topic for a different paper (see Biesaga 2017).

Both dictionaries offer multimodal explanations for different vocabulary levels. While IDP tends to illustrate meanings of words from the basic and advanced vocabulary levels, the MW explains mostly meanings connected with the language typical of the advanced level or specialized areas of interests (naturally there are some minor exceptions, for example: accordion, airplane).

Probably the most striking characteristic of multimodality in both reference works is the overrepresentation of two thematic fields: Plants and Animals. When discussing the meanings connected with the first group, we encounter for example such entries:

- a. IDP: akacja (acacia), arbuz (watermelon), baobab (baobab), brokuł (broccoli), bluszcz (ivy), bonsai (bonsai), bodziszek (geranium); bulwa (bulb):

baobab [...] bot. <<ogromne, długowieczne drzewo o białych kwiatach i dużych jadalnych owocach w kształcie ogórka, rosnące na sawannach podzwrotnikowej Afryki>> (IDP 1999: 58)

English translation: baobab — botanics: huge, long-living tree with white flowers and big, edible fruit in the shape of a cucumber, it grows on the African subtropical savannas.



Picture 1: An illustration from the entry *baobab* (IDP)

- b. MW: abelia, acorn, agave, almond, ash, asparagus, aster:

abelia — any of a genus (*Abelia*) of Asian or Mexican shrubs of the honeysuckle family having opposite leaves and white, red, or pink flowers (MW).

Illustration of ABELIA



Picture 2: An illustration from the entry *abelia* (MW)

Also the field related to animals should be perceived as surprisingly rich in content:

- a. IDP: alpaka (alpaca), amazonka (amazon parrot), anakonda (anaconda), biedronka (ladybird), bocian (stork), borsuk (badger), buldog (bulldog); akwarium (aquarium), barć (beehive);

biedronka [...] zool. <<mały chrząszcz czerwono-brunatny lub żółty, z czarnymi kropkami, żywi się mszycami; boża krówka>> (IDP 1999: 70).

English translation: ladybird – zoological: small beetle, red-brown or yellow with black spots, that eats greenflies.

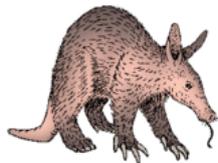


Picture 3: An illustration from the entry *biedronka* (ladybird) (IDP)

- b. MW: aardvark, aardwolf, addax, adder, agouti, amoeba, anaconda, angora goat, arctic fox, armadillo:

aardvark — a large burrowing nocturnal mammal (*Orycteropus afer*) of sub-Saharan Africa that has a long snout, extensible tongue, powerful claws, large ears, and heavy tail and feeds especially on termites and ants (MW).

Illustration of AARDVARK



Picture 4: An illustration from the entry *aardvark* (MW)

This rich content seems to have been inherited from a long-standing encyclopaedic tradition. It was typical of such reference works (as well as of thesauruses) to show to the user how more or less exotic animals or plants look like when characterizing their main features. This statement, however, demands further, more detailed analysis.

Both dictionaries tend to also activate (a medium or small level of presence) multimodal fields enumerated below:

- a. Artistic Activity:
— IDP: akant (acanthus), arabeska (arabesque); akordeon (accordion), altówka (viola), bałalajka (balalaika), banjo (banjo); baletnica (ballerina);
— MW: accordion, alpenhorn;
- b. Army and War:
— IDP: armata (cannon), bagnet (bayonet), broń palna (firearm), buzdygan (mace);
— MW: armor, arrow;

- c. Tradition and Religion:
 - IDP: aureola (aureole), bombka (bauble);
 - MW: amphora, ankh;
- d. Architecture:
 - IDP: absyda (apse), akwedukt (aqueduct), arkada (arcade), bazylika (basilica);
 - MW: alcazar, anta, arbor;
- e. Transport:
 - IDP: amfibia (amphibian), autobus (bus), balon (balloon), bryczka (chaise);
 - MW: anchor, airplane;
- f. Machines and Devices:
 - IDP: brona (harrow);
 - MW: abacus, anvil.

The compilers of the IDP also decided to graphically present the following fields:

- a. Closest Environment: abazur (lampshade), barylka (barrel), budzik (alarm clock);
- b. Clothing: beret (beret), biret (biretta), bransoleta (bracelet), burnus (burnous);
- c. Sport and Leisure Time: akrobacja (acrobatics), as (ace), atleta (athlete), bobslej (bobsleigh).

The compilers of the MW have chosen to illustrate meanings (at least two) from the thematic groups enumerated below:

- a. Body Parts and Body Functioning: acrosome, antibody, antigen, anus, aorta, artery, arteriole;
- b. Diseases and Treatment: amniocentesis, aneurysm.

The rest of the semantic fields in the analyzed dictionaries were either not activated graphically, or there was only one illustration. To explain the reason for such state-of-the art lexicography, exemplary, highly diverse noun meanings were chosen (two of them from each group, see Addendum 2). At the next stage the accuracy of the results displayed in the multimodal corpora was checked. This should help to formulate more general statements about typical obstacles, encountered during the illustrating process.

Analyzing Multimodal Corpora

What is interesting is the fact that most of the items from the "orphan" fields

were more or less difficult to illustrate. Therefore lexicographers were partially right to refrain from exploring these senses. Probably the only "orphan" groups of meanings that should be considered pictorially underestimated are: Food and Personal Care (meanings related to concrete items). As we can see below, exemplary meanings from this lexicon fields are quite easy to illustrate. Probably the only problem is to choose the most prototypical example.



Picture 5: Illustrations: *baguette* (Wikimedia Commons)



Picture 6: Illustrations: *razor* (Pixabay)

As for the other thematic groups, one could come up with two main reasons for the resulting lack of illustration. On the one hand, in the case of some thematic

fields, there are denotation issues. Luckily they can be quite easily solved when using pictorial methods of traditional lexicography. On the other hand, some meanings offer, when launching pictorial corpora, only connotation content⁴. In this particular situation, the compilers of a dictionary have to withdraw from the strictly lexical approach and decide to use, more or less successfully, connotative potential of a certain meaning. Naturally the typology shown below does not relate only to the analyzed group of meanings. The exemplary senses, taken into account, simply prove that the "orphan" groups are especially prone to the issues connected with pictorial denotation and connotation doubts.

Denotation Content

One of the most common problems with the denotation and culturally bonded graphics is the lack of one particular prototypical image. In other words, the information stored in the definition can be connected with different pictorial representatives that would duplicate this semantic content. In the case of such senses, in order to illustrate, one needs to choose a few graphics with different examples of the meaning content, for example: child (Course of Life), liquid (Matter), rain (Weather), schedule (Work), flag (State Functions and Administrative Division), confectionery (Industry).



Picture 7: Illustrations: *child* (Pixabay), MW definition *child* 2a: a young person especially between infancy and youth

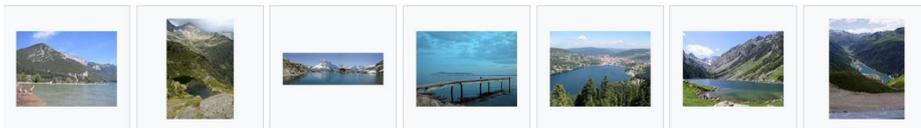


Picture 8: Illustrations: *liquid* (Wikimedia Commons), MW definition *liquid* 2: a fluid (such as water) that has no independent shape but has a definite volume and does not expand indefinitely and that is only slightly compressible

Another situation occurs when there is no pictorial prototype and the lexicographer has to choose an illustration containing elements from the context which might sometimes go beyond the definition, for example: barber (Personal Care, tool, object of action, typical scenery), teacher (Education, typical scenery), star (Sky and Universe, typical surrounding), lake (Earth, typical surrounding), perihelion (Sky and Universe, typical surrounding).



Picture 9: Illustrations: *barber* (Pixabay), MW definition *barber*: one whose business is cutting and dressing hair, shaving and trimming beards, and performing related services



Picture 10: Illustration: *lake* (Wikimedia Commons), MW definition *lake*: a considerable inland body of standing water; also: a pool of other liquid (such as lava, oil, or pitch)

The next group relates to the meanings connected with phases. In this case it is not sufficient to use one image, as one has to use a few of them to enable the process of understanding, for example: erosion (Earth), liposuction (Personal Care).



Picture 11: Illustrations: *erosion* (Wikimedia Commons), MW definition *erosion* 1a: the action or process of eroding (erode 1b: to wear away by the action of water, wind, or glacial ice)

All of the groups mentioned above are slightly pictorial-resistant. Such meanings should be included in dictionaries. They indeed stand on a thin line between dictionary and encyclopaedia, yet, fortunately, in the dictionary area.

Connotation Content

A completely different situation exists in the case of meanings with pictorial denotation content. If we look at the definition and the exemplary graphical facilities from multimodal corpora, we could see that culturally bonded visual code offers us only some connotative potential of meaning.

When there is no connection between the definition and the image content, a typical context connected with an abstract sense will be encountered in the picture. It is usually a conventional environment in which something occurs or exists, sometimes typical doers, symptoms or results, for example: school (Education, environment), factory (Industry, environment), crime (Law and Rules of Social Life, result or doer), friendship (Human Relations, symptoms), euphoria (Mental States, situation and doer), daydream (Mental States, situation and doer), kitsch (Judgement and Valuation, carrier).

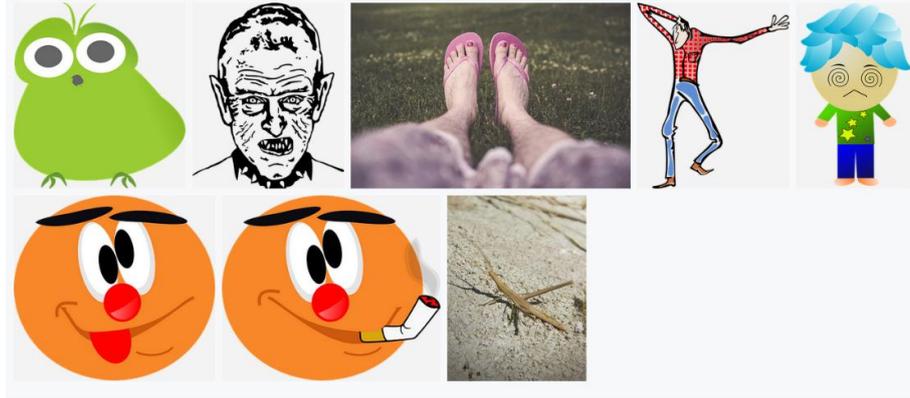


Picture 12: Illustrations: *friendship* (Wikimedia Commons), MW definition *friendship*: the state of being friends (friend: one attached to another by affection or esteem)



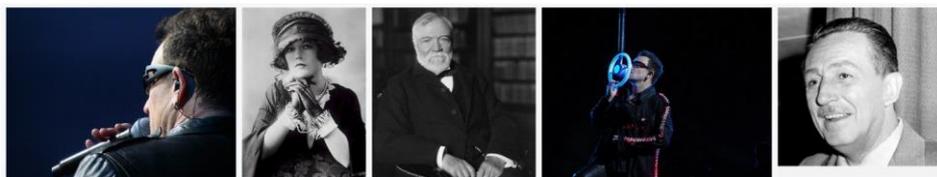
Picture 13: Illustrations: *daydream* (Pixabay), MW definition *daydream*: a pleasant visionary usually wishful creation of the imagination

Meanings connected with valuation, due to their subjectivity, also lead us to a very vast number of connotative images, for example: angel (Character), weirdo (Character), absurd (Valuation and Judgement).



Picture 14: Illustrations: *weirdo* (Pixabay), MW definition *weirdo*: a person who is extraordinarily strange or eccentric

Other very interesting denotation-typical groups occur when a meaning leads us to definite descriptions in multimodal corpora. For example, if we would like to search for images apt for illustrating senses such as: president (State Functions and Administrative Division), philanthropist (Human Relations) or celebrity (Mass Media) we will receive photographs of actual people. Naturally these graphics are prone to changing times and cultural environment.



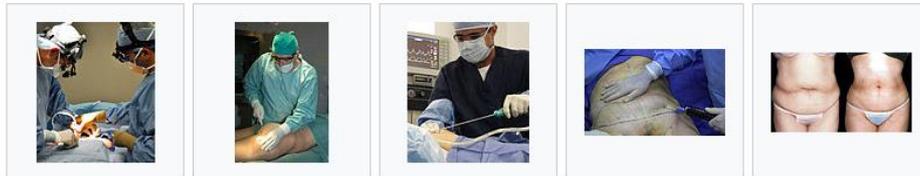
Picture 15: Illustrations: *philanthropist* (Pixabay), MW definition *philanthropist*: one who makes an active effort to promote human welfare: a person who practices philanthropy

Very interesting types of meanings leads us to symbolic images. In this case we will not encounter denotative elements of reality but culturally petrified symbols, for example: death (Course of Life), anarchy (Law and Rules of Social Life).



Picture 16: Illustrations: *anarchy* (Wikimedia Commons), MW definition *anarchy* 2a: absence or denial of any authority or established order

That leads us to the problem of the meaning taboo, existing in both denotative- and connotative-problematic groups. Some senses, especially connected with a human perceived as a physical being (diseases and their symptoms, medical treatments, death, sexual life), can be easily illustrated, however, they seem inadequate for a general reference work, for example: liposuction (Personal Care), nevus (Physicality Characteristics). When presenting them to readers, it is probably better to use drawings instead of photographs.



Picture 17: Illustrations: *liposuction* (Wikimedia Commons), MW definition *liposuction*: surgical removal of local fat deposits (as in the thighs) especially for cosmetic purposes

Conclusion

As can be seen, the problem of including images in dictionaries is partially related to the thematic division of vocabulary. If the authors of the analyzed dictionaries would take into account one of the complete ideographic classifications of language meanings their pictorial strategy would be more coherent and accurate. For now visual facility seems only an addition to the exact verbal explanation. In the future it should be a complementary tool, serving in the process of semantic description.

On the other hand it has been shown above that common lexicographical practice has been shaped not without a reason, while images from "orphan"

groups tend to more often create more or less serious pictorial problems.

Some thematic fields are highly underestimated in general lexicography (Food, Personal Care, Artistic Activity, Closest Environment, Clothing, Sport and Leisure Time, Work, Machines and Devices etc.). Their illustrative potential has not been used in a proper way and should be exploited in a more decisive manner. What is more, existing graphical practice lacks a sense of consequence since pictorially presented meanings are chosen randomly. It is also hard to indicate what is the purpose of presented illustrations (make reference work more attractive?, teach common or advanced vocabulary? help in meaning recognition and retention process?⁵).

The crucial problem in lexicography nowadays is to analyze more deeply the relation between verbal explanation and visual explanation. As we have seen above, it would be better, at least for now, to focus on the meaning types in which denotation can be shared by both definition and image. They should be divided in some, even very basic, coherent semantic groups. At the next step one could conduct experiments with connotative meanings. If pictorial strategy lead us to positive results with meaning perception, the solid fortress of traditional pictorial lexicography could be conquered and rebuilt in the future.

Notes

1. Different typologies of pictorial facilities in dictionaries as well as remarks concerning the usability of the particular presentation modes (drawings, photographs) can be found in: Hupka 1989, Stein 1999, Liu 2015, Biesaga 2016.
2. Before using the material from any free media corpora lexicographers should always ensure if all relevant copyright issues are cleared.
3. Due to the paper topic grammatical information from the IDP has been excluded.
4. Connotation is perceived in this paper as a set of socio-cultural associations connected with a certain meaning (Chandler 2007: 137-143).
5. On the usefulness of multimodal explanations in the process of meaning recognition and retention: Lew and Doroszewska 2009.

References

Dictionaries

- Casares, J. 1959. *Diccionario ideológico de la lengua española*. Second Edition. Barcelona: Editorial Gustavo Gili.
- IDP: Sobol, E. (Ed.). 1999. *Ilustrowany słownik języka polskiego* (English translation: *The Illustrated Dictionary of Polish*). Warszawa: Wydawnictwo Naukowe PWN.
- MW: *Merriam-Webster Dictionary*. Accessed at: www.merriam-webster.com (1-27 January 2017).

Multimodal Corpora

Pixabay. Accessed at: <http://pixabay.com/> (20-27 January 2017).

Wikimedia Commons. Accessed at: <http://commons.wikimedia.org> (20-27 January 2017).

Other Literature

- Batko-Tokarz, B.** 2008. Tematyczny podział słownictwa w Wielkim słowniku języka polskiego. Żmigrodzki, P. and R. Przybylska (Eds.). 2008. *Nowe studia leksykograficzne* 2: 31-48. Kraków: Lexis. Accessed at: <http://ijp.pan.pl/publikacje-elektroniczne/nowe-studia-leksykograficzne-t-2> (24 May 2017).
- Biesaga, M.** 2016. Pictorial Illustration in Dictionaries. The State of Theoretical Art. Margalitadze, T. and G. Meladze (Eds.). 2016. *Proceedings of the XVII EURALEX International Congress. Lexicography and Linguistic Diversity*: 99-108. Tbilisi: Ivane Javakhishvili Tbilisi State University. Accessed at: <https://euralex.org/publications/pictorial-illustration-in-dictionaries-the-state-of-theoretical-art/>.
- Biesaga, M.** 2017. Pictorial Illustrations in Encyclopaedias and in Dictionaries — A Comparison. Kosem I., C. Tiberius, M. Jakubiček, J. Kallas, S. Krek and V. Baisa (Eds.). *Electronic Lexicography in the 21st century. Proceedings of eLex 2017 Conference, Leiden, the Netherlands, 19–21 September 2017*: 221-236. Brno: Lexical Computing. Accessed at: <https://elex.link/elex2017/wp-content/uploads/2017/09/paper13.pdf> (20 September 2017).
- Burada, M. and R. Sinu.** 2016. *Research and Practice in Lexicography*. Braşov: Transilvania University of Braşov.
- Chandler, D.** 2007. *Semiotics. The Basics*. Second Edition. London/New York: Routledge.
- Chapman, R.L. (Ed.)**. 1977. *Roget's International Thesaurus*. Fourth Edition. New York: Thomas Y. Crowell.
- De Jong, G.** 2014. *An Inventory of European Scholarly Dictionaries*. Accessed at: <http://www.elexicography.eu/events/stsm/completed-stsms/> (14 December 2016).
- Gangla, L.A.** 2001. *Pictorial Illustrations in Dictionaries*. Unpublished M.A. Thesis. Pretoria: University of Pretoria. Accessed at: <http://www.repository.up.ac.za/dspace/handle/2263/22862> (24 May 2017).
- Hallig, R. and W. von Wartburg.** 1963. *Begriffssystem als Grundlage für die Lexikographie. Système raisonné des concepts pour servir de base à la lexicographie*. Berlin: Akademie-Verlag.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hupka, W.** 1989. Die Bebilderung und sonstige Formen der Veranschaulichung im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* 1: 704-726. Berlin/New York: Walter de Gruyter.
- Kemmer, K.** 2014. Rezeption der Illustration, jedoch Vernachlässigung der Paraphrase? Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*: 251-278. Berlin/Boston: De Gruyter.

- Klosa, A.** 2016. Illustrations in Dictionaries; Encyclopedic and Cultural Information in Dictionaries. Durkin, P. (Ed.). 2016. *The Oxford Handbook of Lexicography*: 515-531. Oxford: OUP.
- Kwaśnicka-Janowicz, A.** 2007. Problemy związane z wprowadzaniem ilustracji do haseł słownikowych. Żmigrodzki, P. and R. Przybylska (Eds.). 2007. *Nowe studia leksykograficzne* 1: 163-174. Kraków: Lexis. Accessed at: <http://ijp.pan.pl/publikacje-elektroniczne/nowe-studia-leksykograficzne> (24 May 2017).
- Langridge, S.** 1998. The Genesis and Development of Illustration in the English Dictionary. De Beaugrande, R., M. Grosman and B. Seidlhofer (Eds.). 1998. *Language Policy and Language Education in Emerging Nations: Focus on Slovenia and Croatia and with Contributions from Britain, Austria, Spain and Italy*: 69-76. Stamford, CT/London: Ablex.
- Lara, L.F.** 1989. Dictionnaire de langue, encyclopédie et dictionnaire encyclopédique: le sens de leur distinction. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* 1: 280-287. Berlin/New York: Walter de Gruyter.
- Lew, R.** 2010. New Ways of Indicating Meaning in Electronic Dictionaries: Hope or Hype? Zhang, Y. (Ed.). 2010. *Learner's Lexicography and Second Language Teaching*: 387-404. Shanghai: Shanghai Foreign Language Education Press.
- Lew, R. and J. Doroszevska.** 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.
- Liu, X.** 2015. Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos* 25: 210-232.
- Stein, G.** 1991. Illustrations in Dictionaries. *International Journal of Lexicography* 4(2): 99-127.

Addendum 1

Thematic division of meanings

1. Human as Physical Being
 - 1a. Body Parts and Body Functioning
 - 1b. Course of Life
 - 1c. Personal Care
 - 1d. Diseases and Treatment
 - 1d. Physical Characteristics
2. Human as Psychical Being
 - 2a. Mental States
 - 2b. Character
 - 2c. Judgement and Valuation
 - 2d. Intellectual Activity
 - 2e. Artistic Activity
 - 2f. Human Relations
3. Everyday Life
 - 3a. Family
 - 3b. Closest Environment
 - 3c. Clothing
 - 3d. Food
 - 3e. Sport and Leisure Time
 - 3f. Work
4. Human in Society
 - 4a. State Functions and Administrative Division
 - 4b. Army and War
 - 4c. Law and Rules of Social Life
 - 4d. Language
 - 4e. Tradition and Religion
 - 4f. Education
 - 4g. Mass Media
 - 4h. Business and Finances
5. Human and Technology
 - 5a. Architecture
 - 5b. Transport
 - 5c. Machines and Devices
 - 5d. Industry
6. Human and Nature

6a. Sky and Universe

6b. Weather

6c. Earth

6d. Plants

6e. Animals

7. Physical Categories

7a. Matter

7b. Space

7c. Time

7d. Quantity and Numbers.

Addendum 2

Exemplary meanings from the following semantic fields

1. Human as Physical Being:
 - Course of Life: child, death;
 - Personal Care: barber, liposuction, razor;
2. Human as Psychical Being
 - Physicality Characteristics: nevus, hunk;
 - Mental States: euphoria, phantasy;
 - Character: angel, weirdo;
 - Judgement and Valuation: absurd, kitsch;
 - Intellectual Activity: aerodynamics, illiterate;
 - Human Relations: philanthropist, friendship;
3. Everyday Life
 - Family: daughter, adoption;
 - Food: baguette, blender;
 - Work: schedule, lawyer;
4. Human in Society
 - State Functions and Administrative Division: president, flag;
 - Law and Rules of Social Life: crime, anarchy;
 - Language: noun, proverb;
 - Education: school, teacher;
 - Mass Media: casting, celebrity;
 - Business and Finances: manager, money;
5. Human and Technology
 - Industry: factory, confectionery;
6. Human and Nature
 - Sky and Universe: star, perihelion;
 - Weather: rain, frost;
 - Earth: lake, erosion;
7. Physical Categories:
 - Matter: liquid, aroma;
 - Space: height, corner;
 - Time: minute, era;
 - Quantity and Numbers: crowd, kilogram.

A Computational Approach to Zulu Verb Morphology within the Context of Lexical Semantics

Sonja E. Bosch, *Department of African Languages, University of South Africa, Pretoria, South Africa (boschse@unisa.ac.za),*
and

Laurette Pretorius, *School of Interdisciplinary Research and Graduate Studies, University of South Africa, Pretoria, South Africa (pretol@unisa.ac.za)*

Abstract: The central research question that is addressed in this article is: How can ZulMorph, a finite state morphological analyser for Zulu, be employed to add value to Zulu lexical semantics with specific reference to Zulu verbs? The verb is the most complex word category in Zulu. Due to the agglutinative nature of Zulu morphology, limited information can be computationally extracted from running Zulu text without the support of sufficiently reliable computational morphological analysis by means of which the essential meanings of, amongst others, verbs can be exposed. In this article we describe a corpus-based approach to adding the English meaning to Zulu extended verb roots, thereby enhancing ZulMorph as a lexical knowledge base.

Keywords: ZULU VERB MORPHOLOGY, VERB EXTENSIONS, LEXICAL SEMANTICS, COMPUTATIONAL MORPHOLOGICAL ANALYSIS, ZULMORPH, ZULU LEXICAL KNOWLEDGE BASE, BITEXT

Opsomming: 'n Rekenaarmatige benadering tot Zoeloe werkwoordmorfologie binne die konteks van leksikale semantiek. Die sentrale navorsingsvraag wat in hierdie artikel onder die loep kom, is: Hoe kan ZulMorph, 'n eindige toestand morfologiese ontleder vir Zoeloe, gebruik word om waarde toe te voeg tot Zoeloe leksikale semantiek, met spesifieke verwysing na Zoeloe werkwoorde? Die werkwoord is die mees komplekse woordkategorie in Zoeloe. As gevolg van die agglutinerende aard van Zoeloe morfologie kan net beperkte inligting sonder die ondersteuning van voldoende betroubare rekenaarmatige morfologiese analise uit lopende Zoeloe teks onttrek word. Die morfologiese inligting stel essensiële betekenis van, onder andere, werkwoorde, bloot. In hierdie artikel beskryf ons 'n korpusgebaseerde benadering om die Engelse betekenis aan uitgebreide werkwoordwortels van Zoeloe toe te ken en sodoende ZulMorph as leksikale kennisbasis uit te brei.

Slutelwoorde: ZOELOE WERKWOORDMORFOLOGIE, WERKWOORDUITBREIDINGS, LEKSIKALE SEMANTIEK, REKENAARMATIGE MORFOLOGIESE ANALISE, ZULMORPH, ZOELOE LEKSIKALE KENNISBASIS, BITEKS

1. Introduction

The integral role of the Internet and the world-wide web in facilitating the production and consumption of enormous amounts of information in digital space depends on the ability of computers to perform a wide variety of tasks involving human language. This requires, amongst others, computational approaches to representing and understanding world knowledge on the one hand, and knowledge about human language in machine-processable form on the other hand. Central to this endeavour is the notion of meaning or semantics, and more specifically lexical semantics, generally defined as the linguistic study of the meaning of individual words, and the meaning-related connections between words. Moreover, contemporary research in lexical semantics as such also relies on natural language processing (NLP) for a wide range of computational approaches and on large electronic corpora that have "revolutionized the possibilities of investigating usage patterns in real language across genres and cultures and further develop probabilistic usage-based ideas." (Paradis 2012).

Typical computational lexical semantics tasks include word sense disambiguation in context, computing word similarity and word relatedness, as well as other relations between words, and semantic role labelling (Jurafsky and Martin 2009). In turn, NLP applications such as machine translation, question answering, information retrieval, information extraction, text classification and multilingual conversational agents, to name but a few, rely on these basic tasks in realising a digital space in which the users of diverse languages can participate in cross-lingual knowledge production and consumption. Performing computational lexical semantics tasks across languages brings the added complexity of requiring access to NLP support in multiple languages. For under-resourced languages it has become common practice to use a well-resourced language such as English as a type of pivot language for providing word meaning and cross-lingual lexical semantics.

Lexical semantic knowledge has up to now been captured mainly through two approaches. "The first is the knowledge-based approach, in which human linguistic knowledge is encoded directly in a structured form, resulting in various types of lexical knowledge bases. The second is the corpus-based approach, in which lexical semantic knowledge is learnt from corpora and then represented in either explicit or implicit manners." (Gurevych et al. 2016: xiii). Broadly speaking, lexical knowledge bases are knowledge bases that provide lexical information about words of a particular language.

In this article the focus is on Zulu, an official language of South Africa, which is, amongst others, characterised by its rich agglutinative morphology in which the verb is the most complex word category. In spite of its official status, Zulu is considered an under-resourced language. When dealing with under-resourced languages, it is common practice to use as much of the available language data and resources as possible. For this reason, both kinds of approaches to lexical semantic knowledge are employed: hand-crafted expert linguistic/lexical knowledge in machine-processable form as well as growing volumes of

electronic Zulu text corpora. There is also a deeper linguistic justification for employing these two complementary approaches: The first exploits the regularity of linguistic structure — in our case the basic morphological structure and the so-called predictable meanings associated with morphemes, in this case the verb extensions. The second caters for the irregularities, the idiosyncrasies that occur in all languages, and for the "unpredictable" lexicalised meaning of extended verb roots. More specifically, *we show how ZulMorph, a comprehensive hand-crafted finite state morphological analyser for Zulu, and the South African Constitution (SAC), a small electronically available parallel English–Zulu corpus which is an official document of the highest order, translated into all official languages, can contribute to Zulu lexical semantics with English as pivot language.*

2. Basic approach

A lexical knowledge base (LKB) is a digital knowledge base "that provides lexical information about words" (Gurevych et al. 2016). Conceptually the most basic unit or entry in a lexical knowledge base is the so-called (*lemma*¹, *meaning*) pair². While our ultimate aim is to construct such pairs for all the words of Zulu, *nouns* and *verbs* are specifically important since they play a central role in knowledge representation — nouns usually name concepts about which information is represented and verbs often express relationships between concepts. Moreover, verbs are the morphologically most complex word category in Zulu. The verb in Bantu languages, in general, incorporates a great deal of information, to the extent that it may even stand alone as a sentence. It is for this reason that we focus on the latter word category in this article.

We propose a computational approach based on ZulMorph. As a comprehensive hand-crafted finite state morphological analyser, ZulMorph not only contains lemmas of most Zulu words, based on various paper dictionaries, other language resources and text books for Zulu (Pretorius and Bosch 2003; Bosch and Pretorius 2006), but it is also arguably the most complete model of the morphological structure of Zulu words. So, when presented with a valid Zulu word, it provides the lemma as part of the full morphological analysis of the word. What ZulMorph does not yet provide, is the meaning of the lemma.

Representing the meaning, also often referred to as the sense, of a lemma is well-known to be hard (see, for example, Faruqui 2016) and has been studied extensively for a language such as English, generally considered a well-studied and digitally well-resourced language. Jurafsky and Martin (2009) provide an excellent introduction to and overview of computational approaches to the representation of word meaning and word sense in English. Therefore, since computational word meaning representation approaches and resources (Lazaridou et al. 2013) for Zulu are not readily available, *we propose a cross-lingual approach with English as pivot language for providing the meaning of a Zulu lemma.* More specifically, we enhance ZulMorph to output a lemma, as well as its English translation equivalent as the meaning of the lemma. Endowed with this added

capability, we then propose that ZulMorph, as basic Zulu LKB, would enable the user to rely on the rich computational infrastructure of English word meaning representation in further processing and applications.

The structure of the article is as follows: Section 2 outlines the approach followed to address the stated problem. Section 3 provides a brief overview of Zulu verb morphology with specific reference to verb extensions, their complexity, their predictability of meaning and related lexicalisation issues. We specifically emphasise morphological (lemma) and semantic (meaning) challenges. In Section 4 ZulMorph is presented as an approach to lemmatisation. As before, the focus is on verbs, their roots and their extensions. In Section 5 the hand-crafting of a basic Zulu LKB from existing paper dictionaries and grammar texts as a "snapshot" of Zulu lexical semantic information is presented. In section 6 the focus is on a corpus-based approach to semi-automatically extracting new verb roots, new extensions and new lexicalised meanings³ for the possible addition to the ZulMorph-based LKB. Section 7 concludes the article and provides suggestions for future work.

3. Zulu verb morphology

The morphological composition of the verb is considerably more complex than that of any other word category in Zulu. A number of slots, preceding and also following the verb root, may contain numerous morphemes with functions such as derivation, inflection for tense-aspect and marking of nominal arguments. Examples are cross-reference of the subject and object by means of class- (or person-/number-) specific markers, locative affixes, morphemes distinguishing verb forms in clause-final and non-final position, negation morphemes and so forth. In this article we concentrate on the so-called verb extension morphemes (Poulos and Msimang 1998: 183-207). As is the case with most Bantu languages, the complex verb morphology of Zulu is characterised by the use of so-called verb extensions to extend or adapt the meaning of a particular verb. By means of a verb extension or a combination of extensions "definite variations of meaning are derived, variations which in English can only be made by the use of auxiliary verbs, adverbs or prepositions." (Doke 1973: 135).

In the inflectional morphology of Zulu the basic meaning of a verb root in Zulu may therefore be modified by suffixing one or more extension morphemes to the verb root⁴, e.g.

- | | | |
|------|---|--|
| (1a) | <i>-phind-a</i>
-verb.root-terminative | 'repeat, do again, return, go back, fold,
make double, duplicate' |
| (1b) | <i>-phind-an-a</i>
-verb.root-reciproc.ext-terminative | 'fold one into the other /coil together' |
| (1c) | <i>-phind-el-a</i>
-verb.root-appl.ext-terminative | 'repeat for, fold for; return' |

-
- | | | |
|------|---|---|
| (1d) | <i>-phind-el-el-a</i>
-verb.root-appl.ext-appl.ext-terminative | 'repeat again and again; return again and again' |
| (1e) | <i>-phind-is-a</i>
-verb.root-caus.ext-terminative | 'cause to repeat, return, fold, send back; retaliate, take vengeance, avenge oneself' |
| (1f) | <i>-phind-is-el-a</i>
-verb.root-caus.ext-appl.ext-terminative | 'send back for, send back to; retaliate against, repay vengeance, revenge oneself upon' |

It is significant that the verb root *-phind-* may use 22 different combinations of verb extensions of which 6 feature as headwords in the Zulu–English Dictionary (ZED) (1964: 662–663). In the outer matter (ZED 1964: ix), it is indicated that separate entries have been made for "verbal derivatives" (extended verb stems) that "convey some meaning or idiomatic usage not deducible from the inherent significance of the derivative form", e.g.

- | | | |
|------|-------------------|--------------------------------|
| (2a) | <i>-hamb-a</i> | 'travel, move along' |
| (2b) | <i>-hamb-el-a</i> | 'visit, be on good terms with' |

In other cases, where the "inherent significance of the derivative form" is easily deducible from the basic verb stem, the derivative forms are listed in brackets after the entry of the basic form, e.g.

- | | | |
|-----|---|-------------------------|
| (3) | <i>-pikiz-a</i>
(pass. <i>-pikizwa</i> ; ap. <i>-pikizela</i> ; caus. <i>-pikizisa</i>) | 'wriggle about, waggle' |
|-----|---|-------------------------|

According to Wilkes (1971: 261) there is theoretically no limit to the number of verb extensions that may be suffixed to a verb root. However, the database of over 6000 examples collected for his study (Wilkes *op. cit.*) contained very few examples with more than three verb extensions being used simultaneously.

In summary, verb extensions are a key feature of Zulu verbs and their meanings and have to be accounted for in a LKB for Zulu, both in terms of the easily deducible meanings and also the lexicalised and idiomatic usage.

3.1 Morphological challenges

Within a rule-based approach to morphology, the following are examples of morphological challenges (morphotactics and morphophonological alternation rules) that are encountered with regard to verb extensions:

- | | |
|-----|---|
| (a) | Some basic verb roots resemble extended verb roots, e.g. the verb root <i>-hlangan-</i> 'come together; unite; connect' in which the morpheme <i>-an-</i> |
|-----|---|

resembles the reciprocal extension. In this case it is not an extension but part of the verb root.

- (b) Rule-based palatalisation occurs in the formation of passives when the final syllable of a verb root begins with a bilabial consonant, also when such a verb root is separated from the passive extension *-w-* by another extension, e.g.

- (4a) *-boph-a* 'tie, fasten, button up'
-verb.root-terminative
-bosh-w-a 'be tied, fastened, buttoned up'
-verb.root-pass.ext-terminative
- (4b) *-boph-el-a* 'tie for, imprison for'
-verb.root-appl.ext-terminative
-bosh-el-w-a 'be tied for, be imprisoned for'
-verb.root-appl.ext-pass.ext-terminative

Occasionally however, idiosyncrasies occur when bilabials appearing elsewhere in the verb root are palatalised, e.g.

- (5) *ezisetshenziswa* 'that are used'
-sebenz-is-w-a
-verb.root-caus.ext-pass.ext-terminative
-setshenz-is-w-a
(not *-sebenziswa** as expected)

- (c) The order of extensions is not always fixed. For instance the passive extension usually follows other extensions, e.g.

- (6a) *-akh-el-w-a* 'be built for'
-verb.root-appl.ext-pass.ext-terminative

In some cases, the reciprocal may, however, follow the passive extension, e.g.

- (6b) *-akh-el-w-an-a* 'be built for each other'
-verb.root-appl.ext-pass.ext-recip.ext-terminative
(cf. Van Eeden 1956: 657)

It should be noted that we do not deal separately with verb roots that end in *-k-* and *-l-* and which are subject to varying modifications in the formation of the causative (e.g. *-vuk-is-a* > *vu-s-a*; *-vel-is-a* > *-ve-z-a*). The reason is that such extended roots are lemmatised as such in most dictionaries, e.g. Dent and

Nyembezi (1969: 506-507) contains the entries *-vuka* (v) 'wake up; rise up' and *-vusa* (v) 'awaken; rouse up; warn against danger; lift up'.

3.2 Semantic challenges

Whereas the basic meaning of verb roots is easily accessible from existing dictionaries, the semantic challenge lies in the extended or lexicalised meanings that come about when the verb root is extended by means of a variety and combination of verb extensions. In most grammatical descriptions of the Bantu languages, verb extensions are considered to be inflectional suffixes since "they do not change the word category to which a word belongs, but add a regular, predictable meaning to the word" (Kosch 2006: 109). The predictable meanings of extended verb roots can be summarised as in Table 1⁵:

Type of extension	Extension	Predictable meaning
passive	<i>-w-</i> or <i>-iw-</i>	be, being
applicative	<i>-el-</i>	for, on behalf of, to
intensive	<i>-isis-</i>	expresses intensity
causative	<i>-is-</i>	cause to, help
neuter	<i>-ek-</i>	cause or assist to perform an action
reciprocal	<i>-an-</i>	each other
completive	<i>-e!el-</i>	action carried out to perfection or completion

Table 1: Predictable meanings of Zulu verb extensions

Not all verb roots may take all extensions arbitrarily since there are restrictions on the combinations of certain meanings (Poulos and Msimang 1998: 183). The following examples are ungrammatical (*) because the neuter extension is incompatible with the meaning of the two verbs and therefore signifies a semantic restriction:

- (7a) *-ephuk-a* 'get broken; die suddenly' > *-ephuk-ek-a**
 (7b) *-shon-a* 'sink, go down, die etc.' > *-shon-ek-a**

Exceptions occur when the meaning of an extended verb root is lexicalised, and therefore becomes unpredictable to a large extent. Kosch (2006: 106) singles out an extension such as the causative which is prone to lexicalisation in combination with certain verb roots. The result is an unpredictable meaning and a display of derivational properties, e.g.

- (8a) *bon-a* 'see'
 -verb.root-terminative

	<i>-bon-is-a</i> -verb.root-caus.ext-terminative	'show'
(8b)	<i>-lum-a</i> -verb.root-terminative	'bite, suffer sharp pain, itch'
	<i>-lum-is-a</i> -verb.root-caus.ext-terminative	'cause to bite/itch; give a bite of food to/share with'

The applicative extension is also used to indicate "in a direction" when followed by a noun indicating location, e.g.

(9)	<i>-gijim-el-a ezintabeni</i> -verb.root-appl.ext-terminative	'seek shelter in the mountains'
-----	--	---------------------------------

An interesting case is found with the meanings of the verbs *-khohla* 'escape from the memory, slip from the memory' and *-khola* 'satisfy, have confidence in', in the sense that they are unexpectedly used as transitive verbs in the passive, e.g.

(10)	<i>-khohl-w-a</i> <i>-khol-w-a</i>	'forget, overlook' 'be satisfied, believe in'
------	---------------------------------------	--

The predictable versus lexicalised meaning phenomenon has been considered from various perspectives that are important for our computational approach to the lexical semantics of Zulu verbs.

On the one hand, the predictable nature of meaning has been documented and provides justification for us to include such regularity in our computational model of Zulu verbal lexical semantics through the "standard" (rule-based) semantic annotation of verb extensions in ZulMorph. According to Wilkes (1971: 50-51) the adding of a verb extension in Zulu does not imply a radical modification of the lexical-semantic aspect of a verb since this remains basically the same. The modification that takes place is that of the manner in which a process progresses or is executed, while the nature of the process remains unchanged. In cases of combinations of verb extensions being suffixed to a verb, it is only the first suffix after the basic verb root that modifies it. Each of the following extensions in turn modifies the foregoing modification (extended root). This modification process is demonstrated in Figure 1. We return to this sequencing of extensions and their "composite" meanings in Section 5.1.

On the other hand, Chabata (1998: 146) points out that verb extensions in the Bantu language Shona are considered to be derivational morphemes and not inflectional morphemes, one of the reasons being that "they usually change the meanings of the verb roots in question in highly significant ways". This suggests that there is good reason to also make provision in our Zulu LKB for

verb extensions to have "unpredictable" lexicalised meanings. These are not systematic and cannot be captured by means of rules. They have to be found individually mainly through corpus-based approaches and added to the LKB as part of its maintenance and continued enhancement.

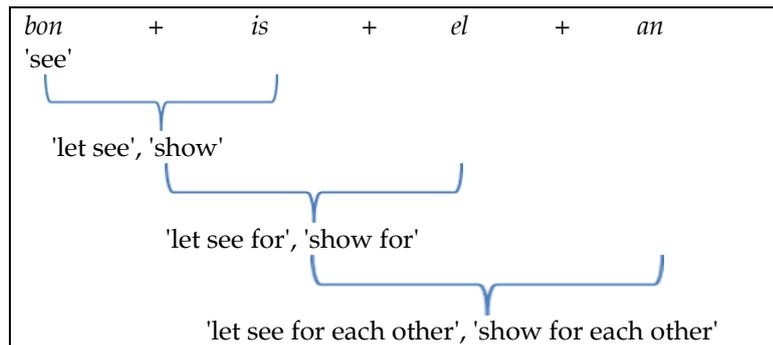


Figure 1: Left-associativity of the compositional meaning of the extended verb root *-boniselan-*, with 'let see' lexicalised as 'show'.

4. Computational Zulu verb morphology and lemmatisation

Before providing the essential details of ZulMorph as the basis for a basic Zulu LKB, we develop the core notion of *Zulu word sense pair*, in this case for verbs.

4.1 What is the lemma and word sense pair of a Zulu verb?

We start by illustrating by means of an example what a *word sense pair* — a (lemma, meaning) pair in English is. We then use this to explicate the notion of *Zulu word sense pair* — a word sense pair in which the lemma is in Zulu and its meaning is the English translation equivalent⁶.

Example 1: 'travels' is a word in the English sentence 'He travels to Johannesburg.' The appropriate meaning of 'travels', according to the Princeton WordNet⁷, is "undertake a journey or trip". The lemma of 'travels' is 'travel'⁸ and therefore the English word sense pair is (travel, undertake a journey or trip).

But what constitutes the lemma and the Zulu word sense pair of a Zulu verb? Four important aspects have to be addressed:

- (i) Lemmatisation via morphological analysis: a standard approach to lemmatisation is through computational morphological analysis (Jurafsky and Martin 2009: 645). For Zulu, the complex agglutinative morphologi-

cal structure of a Zulu verb includes, amongst others, the verb root and its verb extensions. For the purposes of Zulu verbal lexical semantics, the verb root together with its extensions, i.e. the extended verb root, constitutes the lemma of the Zulu verb. This decision is based on the insight that the lexical semantics of the Zulu verb is determined by the verb root AND its extensions since, as we have seen in Section 3, the extensions are meaning changing suffixes to the root. This aspect is addressed in Section 4.3;

- (ii) Assigning a meaning in the form of its English translation equivalent to the verb root. This aspect is addressed in Section 4.4;
- (iii) Assigning English meaning(s) to the verb extensions so that they can be combined (composed) with the meaning of the verb root. This aspect is addressed in Section 4.4;
- (iv) Combining the information in (ii) and (iii) to yield a lemma and a word sense pair for any given Zulu verb in which the meaning is provided as the English translation equivalent of the Zulu lemma, English being our pivot language.

But how is this composite meaning of the Zulu lemma, as defined in (i), obtained? In Table 1 of Section 3.2 the predictable meanings of the respective verb extensions are given and the question now arises as to how a sequence of meanings is combined into one meaning for the extended verb root as Zulu lemma. To answer this question we rely on the left-associative compositional nature of the meaning of the verb root and its sequence of extensions, as already documented by Wilkes (1971) (see Section 3.2 and Figure 1). We illustrate this by means of Example 3. Although we primarily base our modelling of the "composite" meaning on the predictable meaning of extensions, we also attend to lexicalised meaning where relevant.

Example 2: Consider the word *uyahamba* 'he travels'. Through morphological analysis (see example (2a)) we obtain the lemma *hamb*⁹. The appropriate word sense pair is (*hamb*, travel). In further applications¹⁰ the lemma *hamb* may then be linked to the original Princeton WordNet sense "undertake a journey or trip" via the English translation equivalent 'travel'. For the words *uyahambisa* and *uyahambela* the word sense pairs are (*hambis*, cause to travel) and (*hambel*, travel on behalf of/travel towards) or (*hambel*, visit) if the lexicalised meaning of (2b) is used. Similarly linking to the Princeton WordNet could further yield (*hambis*, cause to undertake a journey or trip) and (*hambel*, undertake a journey or trip on behalf of) or (*hambel*, go to certain places as for sightseeing).

Example 3: For the word *uyahambelisa* the lemma is *hambelis*. Its meaning is obtained by composing the respective meanings from the left, as shown by means of the bracketed representation of the lemma (*((hamb)el)is*) : 'cause to'

(meaning of *hambel*) => 'cause to travel on behalf of', 'travel towards' or 'visit' if the lexicalised meaning is used. This yields the Zulu word sense pair (*hambelis*, cause to travel on behalf of/travel towards) or (*hambelis*, cause to visit). As before, this may then be further expanded via the Princeton WordNet to (*hambelis*, cause to undertake a journey or trip on behalf of) or (*hambelis*, cause to go to certain places as for sightseeing).

For any verb in Zulu, we are now able to conceptually provide its Zulu word sense pair. In subsequent sections we show how this lexical semantic information is computationally obtained and encoded in ZulMorph as basic Zulu LKB.

4.2 ZulMorph

ZulMorph was developed with the Xerox finite state toolkit (Beesley and Karttunen 2003) and has also been successfully compiled with Foma (Hulden 2009), a free, open source finite state toolkit. The two central problems of morphology, viz. *morphotactics* (rules for morpheme sequencing) and *morphophonological alternation rules* (rules for spelling and sound changes) are computationally modelled by and implemented as finite state transducers, which are then composed to form one single transducer, which constitutes the morphological analyser. For modelling the morphotactics the *lexc* programming language with its cascading continuation classes of morpheme lexicons (Beesley and Karttunen 2003: 210) is provided and for the alternations rules, *xfst*, a language for using the extensive Xerox finite state calculus, is used¹¹.

It is well-known that the coverage of a finite state morphological analyser such as ZulMorph is determined by (i) the accurate and complete modelling of the morphological structure of the language, and (ii) the comprehensiveness of the noun stem and verb root lexicons. Only valid Zulu words, of which the noun stems or verb roots are present in the respective lexicons, can be analysed correctly. For such a morphological analyser to be maximally useful, these stem and root lexicons need to be maintained and extended as new words enter the language. This remains ongoing work.

In principle, the cascading continuation classes of morpheme lexicons model the filling of slots in the morphological structure of the verb. However, the slots that we are interested in here are those for the verb root and its extensions, since together these constitute the lemma. While the order of the verbal prefixes is fixed (cf. Poulos and Msimang 1998: 305), this is not the case for the extensions. There is no fixed order or number since these are semantically determined. Indeed, as discussed in Section 2, the various verb extensions are not compatible with all verb roots, and there are no hard and fast rules that determine the possible combinations, i.e. roots with extensions, as well as extensions with one another. Comprehensive information on these combinations is not available — not even paper dictionaries provide complete information on combinations and sequences for all verb roots. The inclusion of such

"idiosyncratic" information about verb roots and their (semantically) valid extensions in ZulMorph further emphasises its role as one of the most comprehensive computational models yet of Zulu morphology.

4.3 Modelling the Zulu verb lemma

Before explaining the computational modelling of the Zulu verb lemma, we return to the morphological challenges of Section 3.1 and how we address them. Challenge (a) concerns the common ambiguity of human language for which no real solution exists except to deal with it through semantic context-based disambiguation at a later stage of processing — at the morphological level such limited over-generation will thus occur. Challenge (b) is non-rule-based and is met by hand-crafting the analyser to accurately model all the individual known cases.

Challenge (c) is closely related to aspect (i) in Section 4.1 and is the core of this section. In modelling verbs and their lemmas in ZulMorph, we make provision for different possibilities: a known basic root with no extensions, a known basic root with its own attested sequence(s) of extensions and a known basic root with an as yet unattested (i.e. new) sequence of extensions. Verbs based on basic roots that are not included in ZulMorph will not be analysed. As we shall see in this section, the distinction between morphology and the root lexicon becomes somewhat fuzzy in the case of the Zulu verb and its extensions in that the attested extension sequences of any specific basic verb root should be marked on the relevant basic root and thereby become part of the "lexicon".

In order to describe the modelling of the Zulu verb lemma and its meaning, we briefly explain the notion of `Lexicon` in `lexc`, as well as the technical use of so-called flag diacritics in both the Xerox toolkit and Foma. We show how they are used to record information about the verb lemma in `lexc`.

4.3.1 The verb root lexicon, extension sequences and flag diacritics

In order to keep explanations short, an example is used instead of trying to explain the technical details in a more general setting. The example `lexc` script for the root `-hamb-` is given in Appendix A. As a code fragment for explanatory purposes, it does not, for example, show how verbal prefixes are modelled. It consists of broadly four sections: the preamble in which certain so-called multi-character symbols are declared, the verb root lexicon that typically contains thousands of roots, but for the example contains only the entry `-hamb-`, the modelling of the verb extensions and finally the morpheme lexicon containing the verb terminative `-a`. Each section is briefly discussed.

In the preamble two tags, `[ATT]` and `[NEW]`, are declared for distinguishing between attested extension sequences for `-hamb-` and possibly newly discov-

ered ones in the output produced by ZulMorph, as well as a number of so-called flag diacritics¹² that are used to mark the attested extension sequences of any particular verb root in the verb root lexicon (LEXICON VRoot) in the second section. This lexicon contains various entries for the verb root *-hamb-* each annotated with a P flag diacritic that encodes the specific attested extension sequence. It also shows the next continuation class (Lexicon VExt) containing the morpheme lexicon from which the next morpheme in the input verb should be matched. The third section shows the morpheme lexicons of next morpheme(s) (extensions) that may follow the basic verb root in accordance with the structure of the verb. As mentioned before, we distinguish between the basic root with no extensions, attested and new extensions. This is modelled by LEXICON VExt and its continuation classes VerbTerm, VExtAttested and VExtNew. In lexicon VExtAttested the R flag diacritic is used to match precisely the attested extension sequence that was marked by the corresponding P flag diacritic in the verb root lexicon entry. The lexicon VExtNew and the cyclic lexicon VExtNew2 model any new extension sequence of arbitrary length. The fourth section shows the last continuation class, LEXICON VerbTerm, which models the final verb terminative morpheme, here *-a*, followed by # to indicate that no further (input) morphemes may follow.

While the attested extension sequences are precise and correct, the cyclic modelling of the recognition of new as yet unattested sequences may cause over-generation in that any arbitrary (finite) sequence of extensions, even sequences that are semantically not plausible, will be recognised. This implementation is specifically useful for the purposes of mining new sequences of extensions from a corpus with the understanding that any new sequence will be subjected to human elicitation before inclusion in ZulMorph as an attested sequence.

4.3.2 Coverage

By way of example, the verb root *-hamb-* in Appendix A has eight different extension sequences. In ZulMorph *-hamb-* has eleven attested sequences: *-w-*, *-ek-*, *-el-*, *-is-*, *-isis-*, *-elw-*, *-elex-*, *-elan-*, *-isw-*, *-isel-* and *-isan-*.

ZulMorph contains 8 031 basic roots and 28 477 (extended) verb roots with attested extension sequences, bringing the number of entries in the verb root lexicon of ZulMorph to approximately 36 000. From the extensive data harvested from available paper dictionaries, grammar textbooks and other paper resources, 113 different extension sequences were identified, with the first 20 most frequent sequences (see Appendix B) representing more than 97% of all attested extensions. Statistics about the number of extensions per basic verb root are provided in Appendix C. We note that 22 of them allow between 20 and 30 combinations of one or more verb extensions. The number of lexicalised headwords, as recorded by Doke and Vilakazi (1964), is given in brackets. For example, the basic verb root in ZulMorph with the largest number of extensions,

viz. 30, is *-fan-* ('resemble'). The basic root *-bon-* ('see') has 28 extension sequences. Moreover, ZulMorph contains 6 153 basic verb roots that have at least one attested extension and 1 878 that have no attested extensions. In Appendix D we list the basic verb roots that have the longest attested extension sequences, as recorded in ZulMorph, for example:

- (11) *-ling-an-is-el-a* 'equalise for/make equal for'

The extensive coverage of both Zulu morphology and its verb roots, basic and extended, in ZulMorph provides the basis for the LKB of the next section.

5. Hand-crafting a basic LKB for Zulu

Hand-crafting a basic LKB for Zulu consists of a systematic and comprehensive usage of the expert knowledge that has been published and made available for Zulu. Three kinds of information need to be encoded — firstly the morphology, secondly the Zulu lemmas and thirdly their meanings. Since ZulMorph is an accurate model of Zulu morphology and its comprehensive coverage of Zulu verb lemmas was addressed in the previous section, we now turn our attention to the acquisition and inclusion of their meanings using a so-called expert knowledge-based approach, as already alluded to in Section 4.1 (ii)–(iv). More specifically, a meaning in the form of an English translation equivalent is assigned to each verb root and its extensions. While our main focus is on predictable meaning as a first step, lexicalised meaning is also considered.

5.1 Representing the meaning of the lemma

The first step in adding meaning to each basic verb root in ZulMorph is including the English translation equivalent to each basic verb root in the `VRoot` lexicon and the predictable meaning to each extension in the `Attested`, `VExtNew` and `VExtNew2` lexicons. For example, the code fragments

```
hamb(travel)@P.Basic.ON@: hamb@P.Basic.ON@ VExt;
hamb(travel)@P.ExtEL.ON@: hamb@P.ExtEL.ON@ VExt;
```

and

```
an(each_other)[RecipExt]: an VExtNew2;
el(for)[ApplExt]: el VExtNew2;
is(cause_to)[CausExt]: is VExtNew2;
el(for)[ApplExt]@R.ExtEL.ON@: el@R.ExtEL.ON@ VerbTerm;
```

yield the following analyses¹³:

```
uyahambela:
u[SC][1]ya[LongPres]hamb(travel)[VRoot]el(for)[ApplExt][ATT]a[VT]
u[SC][2ps]ya[LongPres]hamb(travel)[VRoot]el(for)[ApplExt][ATT]a[VT]
u[SC][3]ya[LongPres]hamb(travel)[VRoot]el(for)[ApplExt][ATT]a[VT]
```

```
bayahambelisana:  
ba[SC][2]ya[LongPres]hamb(travel)[VRoot]el(for)[ApplExt]is(cause_  
to)[CausExt]an(each_other)[RecipExt][NEW]a[VT]
```

The respective word sense pairs are (*hambel*, travel for/travel towards) and (*hambelisan*, cause to travel for/towards each other). Note the composite meaning in the latter pair.

By adding basic meanings to the 8 031 basic verb roots and by including the predictable meanings of the various extensions (7 in total) we are able to provide not only a first approximation of the meaning of each of the ~36 000 entries in the verb root lexicon, but also produce word sense pairs for all the Zulu verbs that are based on these basic roots. Keeping in mind that the extensive Princeton WordNet for English has 11 529 verbs, the ZulMorph coverage of the Zulu extended verb root semantics is significant and can already be used in applications, as alluded to in Section 1.

Adding lexicalised meaning is the most resource intensive part of endowing ZulMorph verb analyses with accurate lexical semantics since it has to be added manually for each verb root individually. For each basic verb root and a particular extension sequence for which a lexicalised meaning is available, the meaning of the *basic* root is replaced by the lexical meaning of the *extended* root while the meaning of the extension that caused the lexicalisation is no longer explicit. The tag [LEX] shows that lexicalisation has occurred. As before, the predictable meanings of any subsequent extensions, if present, are still shown. By way of example we consider the extended root *-hambel-*, which also has the lexicalised meaning of 'visit'. Therefore, the verb root lexicon entry is as follows:

```
hamb(visit)[VRoot]el[ApplExt][LEX]@P.Lex.ON@ @P.Basic.ON@: hambel@  
P.Lex.ON@ @P.Basic.ON@ VExt;
```

and yields the analyses

```
uyahambela:  
u[SC][1]ya[LongPres]hamb(visit)[VRoot]el[ApplExt][LEX]a[VT]  
u[SC][2ps]ya[LongPres]hamb(visit)[VRoot]el[ApplExt][LEX]a[VT]  
u[SC][3]ya[LongPres]hamb(visit)[VRoot]el[ApplExt][LEX]a[VT]
```

The resulting word sense pair is (*hambel*, visit).

In summary, by annotating each entry in the verb root lexicon with its meaning (either predictable or lexicalised) and by providing the meanings of the 113 extension sequences, the morphological analysis of any Zulu verb will contain sufficient semantic information to support a basic notion of semantic linking or interoperability — a possibility that did not exist before.

6. Enhancing the Zulu LKB through a corpus-based approach

Improving and updating an electronic LKB to keep it current and maximally

useful, specifically for an under-resourced language such as Zulu, is essential for its digital (web) presence, as discussed in Section 1. Having exploited available paper resources such as dictionaries, grammar textbooks, wordlists and terminologies etc., the obvious next step is to "mine" electronically available language corpora for new lexical information to add to ZulMorph. Such lexical information includes new verb roots, new extension sequences, and new (as yet unrecorded) lexicalised meanings of extended roots as they occur in authentic language use. For this purpose we propose in this section a semi-automated corpus-based approach to the extraction of new lexical information about verbs.

By way of example, the SAC (parallel English and Zulu versions) that has been sentence-aligned is used. It was chosen for mainly four reasons: firstly it is publicly available in all the official South African languages, secondly it is assumed to have been professionally quality assured, thirdly it is by its very nature well-structured and lends itself to accurate sentence alignment, and fourthly it uses contemporary formal language. The idea is that this process should be continued as new parallel corpora become available in due course.

The extraction of bilingual lexical information from bitexts¹⁴ has a long tradition. Tiedemann (2011) provides an overview of techniques that may be applied for this purpose. Although he focuses on statistical approaches to word alignment, he also briefly discusses a number of non-statistical techniques for lexicon extraction from bitexts (Tiedemann 2011: 100-102). While automatic word alignment "is just too noisy to be useful for qualitative investigations", these non-statistical techniques "focus on the extraction of reliable translation equivalents", usually emphasising high precision links between words and multi-word units.

The approach that we follow in this article may also be seen as such a non-statistical technique aimed at high precision.

New basic verb roots lead to morphological analysis failures. Through human elicitation and by individually considering these failures, new basic roots are identified and added to ZulMorph, together with their English translation equivalents. Alternatively, we could apply the guesser variant of ZulMorph to the failures and in this way obtain new verb root candidates. These also need to be subjected to human linguistic scrutiny before adding them to ZulMorph. The occurrence of a *new extension sequence* is tagged in the morphological analyses of a verb as [NEW]. Such a sequence is then manually checked and added to ZulMorph, as shown in Section 5. For *additional attested sequences for specific basic verb roots* basically the same procedure is followed.

For the extraction of *new (lexicalised) meanings and (extended) roots as they occur in authentic language use* we employ bitexts — it is here that the sentence aligned parallel corpus plays a central role. For each sentence may we proceed as follows:

1. Perform part of speech (POS) tagging of the English sentence. For this purpose we used TreeTagger¹⁵.

2. Perform a morphological analysis of the Zulu sentence, using ZulMorph.
3. Isolate the verbs in the English sentence using the POS tags, and the verb roots and their extensions in the Zulu sentence using the morphological analysis tags for the verb root and its verb extension, and align these (the POS tags and morphological tags). This directly links the English lemma¹⁶, i.e. the new (lexicalised) meaning, which is our translation equivalent for the new Zulu word sense pair, and the Zulu (extended) verb root, the Zulu lemma in our new word sense pair.
4. Add the information to ZulMorph so that it includes the new Zulu word sense pair.

In this semi-automated process steps 1 and 2 are automated while steps 3 and 4 as yet require manual intervention.

Specific examples that have been extracted in this way are shown in Tables 2–6¹⁷.

In Table 2 we demonstrate how a new lexicalised meaning 'impart' has been detected for *-dlulis-* in the verb alignment process. In sentence <s103>¹⁸ the English verb 'impart' links up with the extended Zulu verb root *-dlulis-*, forming a new lexicalised addition to those already listed for *-dlulisa* in the ZED (1964: 162), namely 'cause to pass; carry past, send past ...'. Verb alignment between a new lexicalized meaning 'impart' and the Zulu lemma *dlulis* therefore results in the new word sense pair (*dlulis*, impart).

English sentence <s103> fragment	Zulu sentence <s103> fragment
... freedom to receive or impart information inkululeko yokuthola noma ukudlulisa imininingwane ...
English — Word, POS and lemma	Zulu — morphologically analysed with ZulMorph, disambiguated manually
<s103> freedom NN freedom to TO to receive VB receive or CC or impart VB impart information NN information	<s103> inkululeko: i [NPrePre] [9] n [BPre] [9] khululeko.9-10 [NStem] yokuthola: ya [PC] [9] u [NPrePre] [15] ku [BPre] [15] thol [VRoot] a [VT] noma: noma [Conj] ukudlulisa: u [NPrePre] [15] ku [BPre] [15] dlul [VRoot] is [CausExt] [ATT] a [VT] imininingwane: i [NPrePre] [4] mi [BPre] [4] niningwane.3-4 [NStem]

Table 2: New lexicalisation of Zulu lemma *dlulis*

Table 3 demonstrates verb alignment between the new lexicalized meaning 'limit' and the Zulu lemma *nciphis* to form a new word sense pair (*nciphis*, limit). The English verb 'limit' in sentence <s286> links up with the extended Zulu verb root *-nciphis-* and produces a new lexicalised supplement to those already

listed for *-nciphisa* in the ZED (1964: 532): 'diminish; make small, less; minimize'.

English sentence <s286> fragment	Zulu sentence <s286> fragment
... no law may limit any right awukho umthetho onganciphisa noma yiliphi ilungelo ...
English — Word, POS and lemma	Zulu — morphologically analysed with ZulMorph, disambiguated manually
<s286> no DT no law NN law may MD may limit VB limit any DT any right NN right	<s286> awukho: a[NegPre]u[SC][3]khona[Adv] umthetho: u[NPrePre][3]mu[BPre][3]thetho.3-4[NStem] onganciphisa: o[RC][3]nga[Pot]nciph[VRoot]is[CausExt][ATT]a[VT] noma: noma[Conj] yiliphi: yi[CopPre]li[EC][5]phi[EnumStem] ilungelo: i[NPrePre][5]li[BPre][5]lungelo.5-6[NStem]

Table 3: New lexicalisation of Zulu lemma *nciphis*

Verb alignment between the novel lexicalised meaning 'register' and the Zulu extended verb root *-bhalis-* is shown in Table 4. A new word sense pair (*bhalis*, register) is created for possible inclusion in dictionaries (e.g. ZED, and isiZulu.net) where *-bhalisa* has not yet been listed as headword. It should be noted however, that the SZD (1969: 309) lists *-bhalisa* as headword with the meaning 'put name on waiting list', while the OZSD (2010: 18) does in fact list *-bhalisa* with the meaning 'register'.

English sentence <s2471> fragment	Zulu sentence <s2471> fragment
... to register that immovable property wokubhalisa leyo mpahla engagudluki ...
English — Word, POS and lemma	Zulu — morphologically analysed with ZulMorph, disambiguated manually
<s2471> to TO to register VB register that DT that immovable JJ immovable property NN property	<s2471> wokubhalisa: wa[FC][3]u[NPrePre][15]ku[BPre][15]bhal[VRoot]is[CausExt][ATT]a[VT] leyo: leyo[Dem][9][Pos2] mpahla: n[BPre][9]pahla.9-10[NStem] engagudluki: e[RC][9]nga[NegPre]gudluk[VRoot]i[VTNeg]

Table 4: New lexicalisation of Zulu lemma *bhalis*

In Table 5 it becomes clear how a new lexicalised meaning 'affirm' has been identified for the Zulu lemma *qinisekisa* in the verb alignment process. In sentence <s51> the English verb 'affirm' links up with the extended Zulu verb root *-qinisekisa-*, forming a new lexicalised addition to those already listed for *-qinisekisa* in the OZSD (2010: 198): 'make sure; make certain'. Verb alignment between a new lexicalised meaning 'affirm' and the Zulu lemma *qinisekisa* therefore results in the new word sense pair (*qinisekisa*, affirm), which could also qualify for inclusion in a dictionary such as ZED, where *-qinisekisa* has not yet been listed as headword. The same procedure applies to the occurrence of the extended Zulu verb root *-qinisekisa* as occurs in <s45> and <s157> respectively, resulting in two further new word sense pairs (*qinisekisa*, ensure) and (*qinisekisa*, secure).

English sentence <s51> fragment	ZUL SAC text <s51>
... affirms the democratic values of human dignity uqinisekisa amagugu entando yeningi yokwazisa isithunzi somuntu ...
ENG — Word, POS and lemma	Zulu — morphologically analysed with ZulMorph, disambiguated manually
<s51> affirms VBZ affirm the DT the democratic JJ democratic values NNS value of IN of human JJ human dignity NN dignity	<s51> uqinisekisa: u[SC][3]qin[VRoot]is[CausExt]ek[NeutExt]is[CausExt][ATT]a[VT] amagugu: a[NPrePre][6]ma[BPre][6]gugu.5-6[NStem] entando: a[PC][6]i[NPrePre][9]n[BPre][9]thando.9-10[NStem] yeningi: ya[PC][9]i[NPrePre][9]n[BPre][9]ningi.9[NStem] yokwazisa: ya[PC][9]u[NPrePre][15]ku[BPre][15]az[VRoot]is[CausExt][ATT]a[VT] ya[PC][9]u[NPrePre][15]ku[BPre][15]az[VRoot]is[CausExt]a[VT] isithunzi: i[NPrePre][7]si[BPre][7]thunzi.7-8[NStem] somuntu: sa[PC][7]u[NPrePre][1]mu[BPre][1]ntu.1-2[NStem]

Table 5: New lexicalisation of Zulu lemma *qinisekisa*

The example in Table 6 reflects verb alignment between the new lexicalised meaning 'refer back' and the Zulu lemma *buyisel* resulting in a new word sense pair (*buyisel*, refer back). The English verb 'refer' together with its RB (adverb) 'back' in sentence <s364> links up with the extended Zulu verb root *-buyisela-* and produces a new lexicalised supplement to those already listed for *-buyisela* in the ZED (1964: 96), namely 'restore to; return to ... make amends to; replace for or by ... retaliate ... fill up again (as river)'. The same concerns the OZSD (2010: 22) where *-buyisela* is listed with the meanings 'return to; bring back to ... restore (to)'. Another example that confirms this new word sense pair occurs in <s766> of the SAC, namely:

... referring a Bill back to the National Assembly ...
 ... ukubuyisela uMthethosivinywa emuva ePhalamende ...

English sentence <s364> fragment	ZUL SAC text <s364>
... it may refer a dispute back to the organs of state involved ingayibuyisela leyo ngxabano kulezo zingxeny e zombuso ezithintekile ...
ENG — Word, POS and lemma	Zulu — morphologically analysed with ZulMorph, disambiguated manually
<s364> it PP it may MD may refer VB refer a DT a dispute NN dispute back RB back to TO to the DT the organs NNS organ of IN of state NN state involved VBN involve	<s364> ingayibuyisela: i[SC][9]nga[Det]yi[OC][9]buy[VRoot]is[CausExt]el[ApplExt][ATT]a[VT] leyo: leyo[Dem][9][Pos2] ngxabano: n[BPre][9]ngxabano.9-10[NStem] kulezo: ku[LocPre]lezo[Dem][10][Pos2] zingxeny e: zin[BPre][10]ngxeny e.9-10[NStem] zombuso: za[PC][10]u[NPrePre][3]mu[BPre][3]buso.3-4[NStem] ezithintekile: ezi[RC][10]thint[VRoot]ek[NeutExt][ATT]ile[VTPerf]

Table 6: New lexicalisation of Zulu lemma *buyisel*

Finally, the discussion of how the chosen bitext and the semi-automated process were used to uncover new lexical information is concluded by considering further examples in Tables 7 (new root), 8 (new extension sequences) and 9 (new lexicalisations).

Extended verb root and its new English lexicalisation from SAC bitext	-chibiyelw- 'amended by'
Verb root	-chibiyel-
Extension	-w-
Examples from SAC bitext	si-chibiyel-w-e 'amended by' <s511>, <s869>, <s1863> i-chibiyel-w-e 'amended by' <s1200> li-chibiyel-w-e 'amended by' <s2467>
Comments	An example of a new verb root and its extension that does not as yet occur in the ZulMorph embedded verb root lexicon: -chibiyel-w- the meaning of which is 'amended by' as harvested from the SAC. The fact that -chibiyela does occur in the monolingual ISZ (2006: 143), serves as affirmation of the validity of the verb stem.

Table 7: New root -chibiyel- identified from parallel bilingual SAC corpus

Verb and its new English new lexicalisation from SAC bitext	-xoxisan- 'negotiate'
Verb root	-xox-
Extensions	-is-an-
Example from SAC bitext	uku-xox-is-an-a 'negotiating' <s1973>

Comments	<i>-xox-is-an-</i> is not listed as headword in the ZED (1964), nor is the extension string <i>-is-an-</i> listed under the entry <i>-xoxa</i> 'narrate, tell, give account, hold conversation, converse, chat' (ZED 1964: 868). The extension string <i>-is-an-</i> for the verb root <i>-xox-</i> is also not an attested combination in ZulMorph.
----------	--

Table 8a: New extension sequence *-is-an-* for *-xox-* identified from parallel bilingual SAC corpus

Verb and its new English new lexicalisation from SAC bitext	<i>-hlinzekelw-</i> 'be provided for'
Verb root	<i>-hlinz-</i>
Extensions	<i>-ek-el-w-</i>
Examples from SAC bitext	<i>ku-hlinz-ek-el-w-a</i> 'be provided for' <s174> <i>zi-hlinz-ek-el-w-e</i> 'be provided for' <s1434> <i>ku-nga-hlinz-ek-el-w-a</i> 'may be provided for' <s1629>
Comments	<i>-hlinz-ek-el-w-</i> is not listed as headword in the ZED (1964), nor is the extension string <i>-ek-el-w-</i> listed under the entry <i>-hlinzeka</i> "get skinned, murdered, operated upon ... prepare food for expected visitor" (ZED 1964: 329). The extension sequence <i>ek-el-w-</i> in combination with the verb stem <i>-hlinza</i> does not occur in the monolingual ISZ (2006: 486), and it is also not an attested combination in ZulMorph.

Table 8b: New extension sequence *-ek-el-w-* for *-hlinz-* identified from bilingual SAC corpus

Extended verb root and its new English lexicalisation from SAC bitext	<i>-shicilel-w-</i> 'published by'
Verb root	<i>-shicilel-</i>
Extension	<i>-w-</i>
Examples from SAC bitext	<i>u-shicilel-w-e</i> 'is published' <s307> <i>u-shicilel-w-e</i> 'be published' <s1258>, <s1334>
Comments	An example of a new lexicalisation of a verb root and its extension: <i>-shicilel-w-</i> the meaning of which is 'published by' as harvested from the SAC. The verb stem <i>-shicilela</i> is not listed in the OZSD (2010), although it does occur in the SZD (1969: 479) with the meaning 'print; make an impression', in isiZulu.net with the meaning 'print, publish' as well as in the monolingual ISZ (2006: 1104), which is a confirmation of the validity of the verb stem. Furthermore, it is interesting to note that isiZulu.net does not recognise the passive extension in combination with <i>-shicilela</i> . It should be noted that the verb root <i>-shicilel-</i> does not as yet occur in the ZulMorph embedded verb root lexicon.

Table 9a: New lexicalisation of *-shicilel-* identified from parallel bilingual SAC corpus

Extended verb root and its new English lexicalisation from SAC bitext	<i>-khankasel-</i> 'campaign, lobby'
Verb root	<i>-khankas-</i>
Extension	<i>-el-</i>
Examples from SAC bitext	<i>nelo-ku-khankas-el-a</i> 'campaign' <s118> <i>o-ku-khankas-el-a</i> 'campaign' <s1571> <i>uku-khankas-el-a</i> 'lobby' <s1582>
Comments	An example with a "new" meaning that differs from the basic (original) meaning of the verb stem: <i>-khankasa</i> according to the ZED (1964: 380) is "move in horseshoe formation, with a view to intercepting or outmanoeuvring" or according to the SZD (1969: 387) the meaning is "press on (as army), be on the war-path". The new lexicalisation of the extended verb root <i>-khankasel-</i> as identified from parallel bilingual SAC corpus is 'to campaign for', 'to lobby'. The verb stem <i>-khankasa</i> is listed neither in the OZSD (2010) nor in isiZulu.net, although it does occur (with its extension <i>-el-</i>) in the monolingual ISZ (2006: 552), which is a confirmation of the validity of the verb stem.

Table 9b: New lexicalisation of *-khankas-* identified from parallel bilingual SAC corpus

7. Conclusion and future work

We have shown how ZulMorph, a comprehensive hand-crafted finite state morphological analyser for Zulu, and a small electronically available parallel English–Zulu corpus, namely the South African Constitution (SAC), which is an official document of the highest order, translated into all official languages, can enrich Zulu lexical semantics with English as pivot language.

While our approach to enhancing ZulMorph to produce Zulu word sense pairs applies to all word categories, our focus was on the verb as the morphologically most complex word category in Zulu. This complexity arises mainly from (sequences of) verb extensions that are suffixed to the basic verb root to produce modified or new verb meanings. We noted that although a morphological analyser may provide accurate morphological analyses of Zulu verb constructions, these analyses do not offer much information in terms of the meaning of the verb. This constitutes a major impediment to a computational understanding of what a Zulu verb means, and therefore also to applications such as, for example, information extraction from Zulu text, question answering in and from Zulu, machine translation between Zulu and any other language and Zulu natural language generation. In this article we presented a Zulu LKB that uses the well-resourced English language as pivot language towards addressing this challenge.

It is important to note that for a language such as Zulu (morphologically complex and under-resourced) statistical and machine learning approaches

have not yet yielded sufficiently accurate results for the applications mentioned above. Recent experience has shown that building the necessary high-quality, sufficiently large electronic corpora for Zulu has proven more difficult and expensive than handcrafting ZulMorph. This is clear from the fact that ZulMorph actually exists while no corpus-driven statistical approach to Zulu computational (verb) morphology has, as yet, yielded results that are comparable to those of ZulMorph. It is our view that the Zulu LKB that we have reported on in this article has the potential to serve as an important and novel component in future hybrid systems (robust combinations of handcrafted, rule-based, statistical and data-driven machine learning approaches) for Zulu lexical semantics.

Our core contribution is twofold:

- the enhancement of ZulMorph to constitute a large basic LKB for Zulu that, for any input verb, produces a word sense pair consisting of the Zulu lemma of the verb (here the extended root) and its meaning (here its English translation equivalent). The meaning is computationally composed from the meaning of the root and the predictable meaning of its verb extensions;
- a proposed semi-automated corpus-based approach in which existing NLP tools, viz. TreeTagger and ZulMorph, and a bitext in the form of the electronically available sentence-aligned English–Zulu parallel corpus, are used to expose new verb roots, new extension sequences and new lexicalisations of existing verbs and their extensions for addition to the Zulu LKB.

Future work may include increasing the automation of the process while also extending the process to other word categories to offer a more comprehensive Zulu LKB. We also envisage using further parallel English–Zulu corpora across a variety of domains as they become available to extend ZulMorph and the Zulu LKB, and eventually experimenting with the use of the Zulu LKB in some of the mentioned applications. In the longer term we may consider developing LKBs for other languages for which finite state morphological analysers are available.

Endnotes

1. The canonical or so-called citation form of a surface word form. For example, *write* is the lemma of the surface forms *writes*, *wrote* and *written* (cf. Section 3).
2. See, for example, Gurevych et al. (2016: 1)
3. Lexicalisation is also discussed in detail in subsequent sections.
4. For the sake of convenience a verb root followed by one or more extensions, is called an extended root in this article.
5. Also cf. De Schryver (2010: 178).

6. A *word* is taken to be a surface word form as found in a sentence or an utterance; a *lemma* is a specific grammatical form of a word, often also referred to as citation form or canonical base form; *lemmatisation* is the process of mapping a word to a lemma; *meaning* is the denotation, referent, or idea associated with a word; and a *translation equivalent* is a corresponding word or expression in another language (see, for example, Jurafsky and Martin 2009: 645; Gurevych et al. 2016: 1).
7. Princeton University "About WordNet." WordNet. Princeton University. 2010. <<http://wordnet.princeton.edu>>
8. In English the canonical base form of the verb (travel, travels, travelling, travelled) is 'travel'.
9. While we consistently use the hyphen (-) to indicate morpheme boundaries, we view the lemma as an entity that can stand on its own in the context of a word sense pair and therefore the notion of morpheme boundary is not important and therefore not indicated.
10. A discussion of such applications falls outside the scope of this article.
11. The detailed explanation of the **lexc** and **xfst** languages falls outside the scope of the article. The interested reader is referred to Beesley and Karttunen (2003).
12. Flag diacritics provide a light-weight approach to feature-setting and feature-unification operations for enhancing modelling accuracy and runtime efficiency. Specific uses are to enforce separated dependencies and mark idiosyncratic morphotactic behaviour (see Beesley and Karttunen 2002) for a comprehensive exposition). In **lexc** and **xfst** flag diacritics are so-called multicharacter symbols with a distinctive spelling: @operator.feature.value@ and @operator.feature@ where the operators are P (positive (re)setting), N (negative (re)setting), R (require test), D (disallow test), C (clear feature) and U (unification test). The features and values are specified by the user. In ZulMorph flag diacritics are used extensively to, amongst others, model the Zulu noun class system (Bosch and Pretorius 2002; Pretorius and Bosch 2003), long distance dependencies (Pretorius and Bosch 2008), part of speech information and a wide variety of other morphotactic constraints that apply in Zulu. In this article the focus is on their use for annotating each basic verb root with its valid and attested extension sequences.
13. The morphological tags, enclosed in [and], are listed in Appendix E.
14. The term bitext originally referred to "documents along with their translations into other languages to be used in translation studies. Since then, bitexts have attracted a lot of interest in a larger community with many other applications in mind. Therefore, it is now common to let the term bitext refer to a wider range of parallel resources, not only original documents and their direct translations. ... An important characteristic of a bitext is the property that there is some kind of correspondence between the two texts coupled together, for example, translational equivalence." (Tiedemann 2011: 1)
15. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
16. TreeTagger terminology
17. The English POS tags are at <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/Penn-Treebank-Tagset.pdf> and the Zulu tags are provided in Appendix E.
18. For example, the number <s103> refers to sentence 103 in the bitext.
19. The number in brackets after the root provides the number of lexicalised headwords, as recorded by Doke and Vilakazi (1964).

20. Class numbers are added to the tags as [c] where c is the class number. For example, [BPre] [5] denotes the basic prefix of class 5.

References

Dictionaries and Corpora

- ISZ = **Mbatha, M.O. (Ed.)**. 2006. *Isichazamazwi SesiZulu*. Pietermaritzburg: New Dawn Publishers.
- OZSD = **De Schryver, G.-M. (Ed.)**. 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi*. Cape Town: Oxford University Press Southern Africa.
- SAC = *South African Constitution*. Available at: <http://www.justice.gov.za/legislation/constitution/SACConstitution-web-eng.pdf> and <http://www.constitutionalcourt.org.za/site/theconstitution/thetext.htm>.
- SZD = **Dent, G.R. and C.L.S. Nyembezi**. 1969. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.
- ZED = **Doke, C.M. and B.W. Vilakazi**. 1964. *Zulu-English Dictionary*. Second Edition. Johannesburg: Witwatersrand University Press.
- isiZulu.net*. <http://isizulu.net>

Other literature

- Beesley, K.R. and L. Karttunen**. 2003. *Finite State Morphology*. Stanford: CSLI Publications.
- Bosch, S.E. and L. Pretorius**. 2002. The Significance of Computational Morphological Analysis for Zulu Lexicography. *South African Journal of African Languages* 22(1): 11-20.
- Bosch, S.E. and L. Pretorius**. 2006. A Finite-state Approach to Linguistic Constraints in Zulu Morphological Analysis. *Studia Orientalia* 103: 205-227.
- Chabata, E**. 1998. Using the Predictability Criterion for Selecting Extended Verbs for Shona Dictionaries. *Lexikos* 8: 140-153.
- De Schryver, G.-M**. 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.
- Doke, C.M**. 1973. *Textbook of Zulu Grammar*. Sixth Edition. Cape Town/Johannesburg: Longman Southern Africa.
- Faruqi, M**. 2016. *Diverse Context for Learning Word Representations*. Ph.D. Thesis. Pittsburgh, PA: Language Technologies Institute, School of Computer Science, Carnegie Mellon University.
- Gurevych, I., J. Eckle-Kohler and M. Matuschek**. 2016. *Linked Lexical Knowledge Bases: Foundations and Applications*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool.
- Hulden, M**. 2009. Foma: A Finite-state Compiler and Library. *Proceedings of the EACL 2009, Demonstrations Session: 29-32*, Athens: Association for Computational Linguistics.
- Jurafsky, D. and J.H. Martin**. 2009. *Speech and Language Processing*. Second Edition. New Jersey, USA: Pearson Education.
- Kosch, I.M**. 2006. *Topics in Morphology in the African Language Context*. Pretoria: Unisa Press.

- Lazaridou, A., M. Marelli, R. Zamparelli and M. Baroni.** 2013. Compositionally Derived Representations of Morphologically Complex Words in Distributional Semantics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 4–9 2013: 1517-1526. Association for Computational Linguistics.
- Paradis, C.** 2012. Lexical Semantics. Chappelle, C.A. (Ed.). 2012. *The Encyclopedia of Applied Linguistics*: 3357-3356. Oxford: Wiley-Blackwell.
- Poulos, G. and C.T. Msimang.** 1998. *A Linguistic Analysis of Zulu*. Cape Town: Via Afrika.
- Pretorius, L. and S.E. Bosch.** 2003. Finite-State Computational Morphology: An Analyzer Prototype for Zulu. *Machine Translation. Special Issue on Finite-state Language Resources and Language Processing* 18: 195-216.
- Pretorius, L. and S. Bosch.** 2008. Containing Overgeneration in Zulu Computational Morphology. *Southern African Linguistics and Applied Language Studies* 26(2): 209-216.
- Tiedemann, J.** 2011. *Bitext Alignment*. Synthesis Lectures on Human Language Technologies, May 2011 (doi: 10.2200/S00367ED1V01Y201106HLT014).
- Van Eeden, B.I.C.** 1956. *Zoeloe-grammatika*. Stellenbosch/Grahamstown: Die Universiteitsuitgewers en -Boekhandelaars.
- Wilkes, A.** 1971. *Agtervoegsels van die werkwoord in Zulu*. Unpublished Doctoral Thesis. Johannesburg: Rand Afrikaans University.

Addendum

Appendix A: Example lexc script for the verb root *-hamb-* 'walk, travel'

```
Multichar_Symbols
@P.Basic.ON@ @R.Basic.ON@ @D.Basic@
@P.ExtEK.ON@ @R.ExtEK.ON@
@P.ExtELAN.ON@ @R.ExtELAN.ON@
@P.ExtELEL.ON@ @R.ExtELEL.ON@
@P.ExtEL.ON@ @R.ExtEL.ON@
@P.ExtISAN.ON@ @R.ExtISAN.ON@
@P.ExtISEL.ON@ @R.ExtISEL.ON@
@P.ExtISIS.ON@ @R.ExtISIS.ON@
@P.ExtIS.ON@ @R.ExtIS.ON@
[ATT] [NEW]

...
LEXICON VRoot
hamb@P.Basic.ON@ VExt;
hamb@P.ExtEK.ON@ VExt;
hamb@P.ExtEL.ON@ VExt;
hamb@P.ExtIS.ON@ VExt;
hamb@P.ExtISIS.ON@ VExt;
hamb@P.ExtELEL.ON@ VExt;
hamb@P.ExtELAN.ON@ VExt;
hamb@P.ExtISEL.ON@ VExt;
hamb@P.ExtISAN.ON@ VExt;

LEXICON VExt
@R.Basic.ON@: 0@R.Basic.ON@ VerbTerm;
[NEW]@R.Basic.ON@: 0@R.Basic.ON@ VExtNew;
[ATT]@D.Basic@: 0@D.Basic@ VExtAttested;

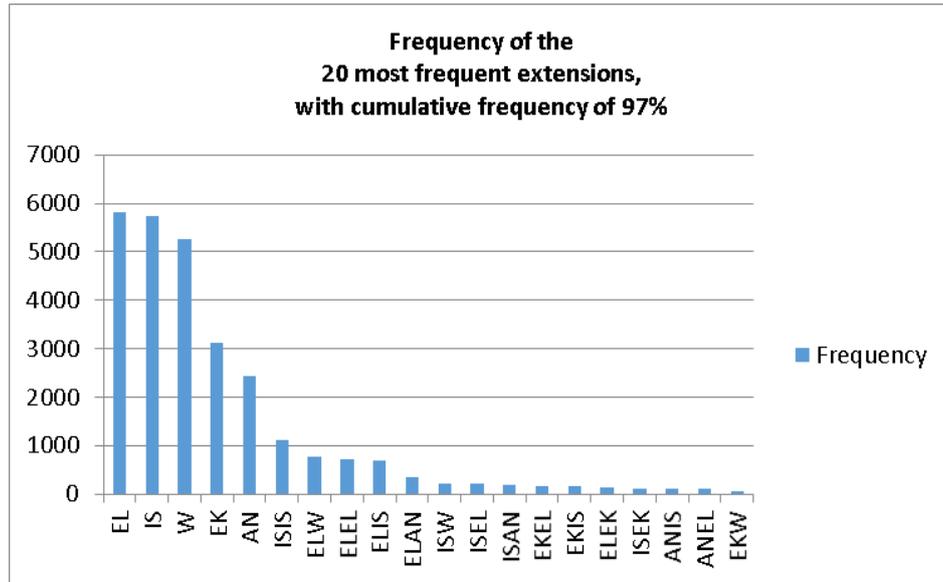
LEXICON VExtNew
an[RecipExt]: an VExtNew2;
ek[NeutExt]: ek VExtNew2;
el[ApplExt]: el VExtNew2;
is[CausExt]: is VExtNew2;

LEXICON VExtNew2
! Recursion to cater for unknown order and arbitrary number of extensions
an[RecipExt]: an VExtNew2;
ek[NeutExt]: ek VExtNew2;
el[ApplExt]: el VExtNew2;
is[CausExt]: is VExtNew2;
VerbTerm;

LEXICON VExtAttested
ek[NeutExt]@R.ExtEK.ON@: ek@R.ExtEK.ON@ VerbTerm;
el[ApplExt]@R.ExtEL.ON@: el@R.ExtEL.ON@ VerbTerm;
is[CausExt]@R.ExtIS.ON@: is@R.ExtIS.ON@ VerbTerm;
el[ApplExt]an[RecipExt]@R.ExtELAN.ON@: elan@R.ExtELAN.ON@ VerbTerm;
el[ApplExt]el[RecipExt]@R.ExtELEL.ON@: el[RecipExt]@R.ExtELEL.ON@ VerbTerm;
is[CausExt]an[RecipExt]@R.ExtISAN.ON@: isan@R.ExtISAN.ON@ VerbTerm;
is[CausExt]el[ApplExt]@R.ExtISEL.ON@: isel@R.ExtISEL.ON@ VerbTerm;
isis[IntensExt]@R.ExtISIS.ON@: isis@R.ExtISIS.ON@ VerbTerm;

LEXICON VerbTerm
a #;
```

Appendix B: Number and frequency of extension sequences in the 28 477 lexical entries in ZulMorph



Appendix C: Basic verb roots with the most extension sequences

No. of extension sequences	Verb root
30	<i>fan</i> (6) ¹⁹
29	<i>enz</i> (8)
28	<i>bon</i> (7), <i>fund</i> (6)
27	<i>az</i> (7), <i>buth</i> (6), <i>buy</i> (6)
26	<i>hlab</i> (8), <i>muk</i> (5)
24	<i>qal</i> (4), <i>sh</i>
23	<i>ling</i> (4)
22	<i>emuk</i> (5), <i>photh</i> (5)
21	<i>futh</i> (6), <i>phik</i> (4), <i>phind</i> (5)
20	<i>nbuk</i> , <i>khub</i> , <i>phath</i> , <i>thath</i> , <i>zal</i>
19	<i>bek</i> , <i>dl</i>
18	<i>al</i> , <i>bhac</i> , <i>bhek</i> , <i>boph</i> , <i>cob</i> , <i>gan</i> , <i>mangal</i> , <i>qin</i> , <i>shay</i>
17	<i>akh</i> , <i>beth</i> , <i>chith</i> , <i>fik</i> , <i>gab</i> , <i>phis</i> , <i>song</i> , <i>thol</i> , <i>vimb</i>
16	<i>fun</i> , <i>khohl</i> , <i>khol</i> , <i>phamb</i> , <i>qand</i> , <i>thel</i> , <i>yek</i> , <i>zw</i>
15	<i>amuk</i> , <i>band</i> , <i>bind</i> , <i>chach</i> , <i>cim</i> , <i>cin</i> , <i>dlul</i> , <i>eq</i> , <i>f</i> , <i>fic</i> , <i>hlom</i> , <i>lung</i> , <i>phish</i> , <i>qond</i> , <i>vul</i>
14	<i>bang</i> , <i>bung</i> , <i>chath</i> , <i>ehl</i> , <i>elam</i> , <i>esab</i> , <i>hlal</i> , <i>hlol</i> , <i>hlum</i> , <i>hol</i> , <i>kham</i> , <i>khaph</i> , <i>khul</i> , <i>mel</i> , <i>ngen</i> , <i>nik</i> , <i>nqum</i> , <i>phons</i> , <i>qed</i> , <i>swel</i> , <i>theng</i> , <i>theth</i> , <i>val</i>
13	<i>bhumbuth</i> , <i>cash</i> , <i>chaz</i> , <i>ding</i> , <i>dlal</i> , <i>dluny</i> , <i>ehluk</i> , <i>emul</i> , <i>encik</i> , <i>gudl</i> , <i>hlinz</i> , <i>lov</i>

Appendix D: Basic verb roots with the longest extension sequences

Extension sequence	Verb root
<i>anisanel</i>	<i>ling</i>
<i>aniselan</i>	<i>ahluk, cin, ehluke, futh, ling</i>
<i>aniselel</i>	<i>futh, ling, phamb</i>
<i>elekelis</i>	<i>phish</i>
<i>elelanel</i>	<i>buth, photh</i>
<i>elelanis</i>	<i>bek</i>
<i>elelisan</i>	<i>hlab</i>
<i>elelisel</i>	<i>hlab</i>
<i>isanisis</i>	<i>jojoz</i>
<i>isekelan</i>	<i>sh</i>
<i>isekelis</i>	<i>sh</i>
<i>aniselw</i>	<i>ahluk, cin, ehluke, futh, ling, phamb</i>
<i>aniswan</i>	<i>ling</i>
<i>elekelw</i>	<i>phish</i>
<i>elelisw</i>	<i>hlab</i>
<i>elelwan</i>	<i>buth, photh</i>
<i>eliselw</i>	<i>balek</i>
<i>isekelw</i>	<i>sh</i>
<i>iselwan</i>	<i>dlal, gan</i>

Appendix E: ZulMorph morpheme tag set (only tags that occur in the examples are provided)

Tag	Description
"Class ²⁰ , person and/or number dependent tags"	
BPre	Basic prefix
Dem	Demonstrative pronoun
EC	Enumerative concord
NPrePre	Noun preprefix
OC	Object concord
PC	Possessive concord
SC	Subject concord
RC	Relative concord
"Tags independent of class, person and/or number"	
Adv	Adverb
ApplExt	Applied extension
CausExt	Causative extension
Conj	Conjunction
CopPre	Copulative prefix
EnumStem	Enumerative stem
IntensExt	Intensive extension
LocPre	Locative prefix
LongPres	Long present tense
NegPre	Negative prefix
NeutExt	Neuter extension
NStem	Noun stem
Pot	Potential
RecipExt	Reciprocal extension
VT	Verb terminative
VTNeg	Verb terminative negative
VTPerf	Verb terminative perfect
VRoot	Verb root
ATT	Attested verb extension sequence
LEX	Lexicalisation
NEW	New verb extension sequence

The Effectiveness of Using a Bilingualized Dictionary for Determining Noun Countability and Article Selection

Alice Yin Wa Chan, *Department of English, City
University of Hong Kong, Hong Kong (enalice@cityu.edu.hk)*

Abstract: This article discusses the use of a bilingualized dictionary, namely *Oxford Advanced Learner's English-Chinese Dictionary 8* (OALECD8), by advanced Hong Kong Cantonese ESL learners in the determination of noun countability and associated article use. A homogenous group of 30 English majors in a local university participated in the study, which consisted of a noun countability and article selection task without and with the use of the dictionary. The results show that although bilingualized dictionaries are useful in helping learners determine noun countability and associated article use, learners often misinterpret dictionary information and model on inappropriate structures, resulting in article errors and/or wrong countability judgments. Chinese translations are also sometimes sources of errors. The results of the study provide lexicographers with signposts to the selection of noun information to be included in a learner's dictionary. More explicit information about noun countability and related article use should be provided in a more user-friendly arrangement. ESL teachers are also advised to engage advanced learners in analyses uncovering the different syntactic requirements of equivalent vocabulary items in the target and source languages when using a bilingualized dictionary.

Keywords: ENGLISH NOUN COUNTABILITY, ENGLISH ARTICLE SYSTEM, ARTICLE SELECTION, BILINGUALIZED DICTIONARIES, LEARNERS' DICTIONARIES, CANTONESE ESL LEARNERS, SECOND LANGUAGE ACQUISITION, EFFECTIVENESS OF DICTIONARIES

Opsomming: Die effektiwiteit van die gebruik van verklarende woordeboeke met 'n tweetalige dimensie om substantiewe se telbaarheid en lidwoordseleksie te bepaal. Hierdie artikel bespreek die gebruik van 'n verklarende woordeboek met vertalings ("bilingualized dictionary"), naamlik die *Oxford Advanced Learner's English-Chinese Dictionary 8* (OALECD8), deur gevorderde Hongkongse Kantonese EVT-leerders vir die bepaling van die telbaarheid van substantiewe en die gepaardgaande lidwoordgebruik. 'n Homogene groep van 30 studente met Engels as hoofvak by 'n plaaslike universiteit het aan die studie deelgeneem, wat uit 'n taak bestaan het waar die telbaarheid van substantiewe bepaal en lidwoorde geselekteer moes word sonder en met die gebruik van 'n woordeboek. Die resultate toon dat, hoewel 'n verklarende woordeboek met vertalings nuttig is om leerders te help om die telbaarheid van substantiewe en die gepaardgaande lidwoorde te bepaal, leerders dikwels woordeboek-inligting verkeerd interpreteer en op onvanpaste strukture baseer, wat tot lidwoordfoute en/of

verkeerde besluite ten opsigte van telbaarheid lei. Chinese vertalings is ook soms die oorsaak van foute. Die resultate van die studie verskaf aanwysings aan leksikograwe vir die seleksie van inligting oor substantiewe wat in 'n aanleerderswoordeboek ingesluit behoort te word. Meer eksplisiete inligting oor die telbaarheid van substantiewe en die verwante gebruik van lidwoorde behoort in 'n meer gebruikersvriendelike ontwerp voorsien te word. EVT-onderwysers word ook aangeraai om gevorderde leerders analyses te laat doen wat die verskillende sintaktiese vereistes van ekwivalente woordeskatitems in die doel- en die brontale ontbloom wanneer 'n verklarende woordeboek met vertalings gebruik word.

Sleutelwoorde: TELBAARHEID VAN ENGELSE SUBSTANTIEWE, LIDWOORDSTELSEL IN ENGELS, ARTIKELSELEKSIE, VERKLARENDE WOORDEBOEKE MET VERTALINGS, AANLEERDERSWOORDEBOEKE, KANTONNESE EVT-LEERDERS, TWEETAALVERWERWING, EFFEKTIVITEIT VAN WOORDEBOEKE

Introduction

The acquisition of English articles by second language learners has often been found to be a difficult process (Huebner 1983; Master 1987; Parrish 1987; Pica 1985; Thomas 1989; Murphy 1997; Robertson 2000; Zobl 1980), especially for learners whose native languages lack articles (Ionin, Zubizarreta and Maldonado 2008; Snape 2008), like Chinese. Article use with abstract nouns is even more difficult than with concrete nouns, resulting in more article errors (Hua and Lee 2005; Ogawa 2008). The intuitive judgment that learners make on the countability of a noun is important in affecting their use of articles with that noun, especially *a/an* vs. *ZERO*, and a direct link between learners' use of *ZERO* and their judgment of non-count nouns has been found in Yoon (1993). Learners' noun countability judgment, however, often deviates from native speaker norms. Many learners have a tendency to consider countability as a fixed or static property and assume that nouns are either countable or uncountable irrespective of context (Butler 2002), and there are learners who think that abstract nouns must be uncountable (Amuzie and Spinner 2013). However, Amuzie and Spinner (2013), following the claim of Allan (1980) that there are different levels of countability, argue that countability should better be treated as a non-binary concept. Whether the perceptual system of noun countability that native English speakers use is "describable, explainable or acquirable by second language learners" (Yoon 1993: 284) is, thus, questioned. Deviant countability judgments often result in learner errors, such as the pluralization of uncountable nouns (e.g. *advices, equipments*), which has been argued in the literature as fairly common (Schneider 2011). Although there exist conflicting findings that the countable use of uncountable nouns is highly infrequent when compared with usage that matches native speaker norms (Hall, Schmidtke and Vickers 2013), ESL/EFL learners' difficulties with the use of English articles as a result of indeterminate noun countability are well-attested (Xue 2010).

Noun Countability in English

Noun countability is one area in English which is most problematic, as the countability of an English noun is not so clear-cut: The countable/uncountable distinction is even described as "primarily arbitrary, unprincipled, or idiosyncratic" (Wisniewski, Lamb and Middleton 2003: 585). Although most English nouns are clearly countable (e.g. *boy, girl*) or uncountable (e.g. *furniture, equipment*), there are many nouns which may be countable or uncountable depending on their meaning (e.g. *light, difficulty*) (Greenbaum and Nelson 2009). It has even been argued in the literature that there is no absolute constraint which will prevent any nouns from functioning uncountably (Allan 1980; cited in Dziemianko 2012). Uncountable nouns are often described as nouns which can only be accompanied by determiners which do not refer to a distinction in number, such as possessives (e.g. *their*) and demonstratives (e.g. *that*) (Greenbaum and Nelson 2009), yet there are uncountable nouns which can be used with the indefinite article *a/an* (e.g. *knowledge in a good knowledge of music, awareness in an awareness of the importance*). ESL learners often find it difficult to understand why a certain English noun is mass instead of count, or vice versa (Lock 1996), as many nouns which are often perceived as countable by ESL learners (e.g. *furniture*) are in fact uncountable in English, and nouns which are often perceived as uncountable can be countable (e.g. *water*). Popular grammar books on the market, including those targeting university students, only give examples of clearly countable and uncountable nouns and explain the conditions, often in very general terms and with a few examples, under which uncountable nouns can be used as countable or vice versa. Other less common usages, such as the use of the indefinite article with an uncountable noun, is seldom, if at all, mentioned. As a result, other learning resources such as learners' dictionaries become important information providers especially when the usage of individual nouns is in question. As Lock (1996) argues, "learners need to have access to a good dictionary in which the different uses of nouns associated with any differences in their countability status are clearly marked" (p. 24), and with more use of learners' dictionaries, learners could identify the countability of a noun more correctly (Miller 2005). A good dictionary should show, therefore, the countable and uncountable usages of a noun clearly and be able to help learners differentiate the differences between the two usages, as it is the uses of a noun which should be classified into countable and uncountable rather than the noun itself (Tang 2006).

Noun Countability Information in Learners' Dictionaries

Although noun countability information is one important piece of information about nouns in learners' dictionaries (Chan and Loong 1999) and has been argued as a ready-made tool which can help learners "acquire one of the hardest grammatical features of the English language" (Miller 2006: 435), there are

doubts about the usefulness of such information in learners' dictionaries. Xue (2010), for example, has observed that the lack of indication of what articles or quantifiers should be used before a noun, as well as discrepancies between the labels (e.g. uncountable) and usage examples (e.g. *an education*), limits the effectiveness of dictionary information in helping learners overcome their problems in noun countability.

A preliminary survey of popular learners' dictionaries used in Hong Kong confirms Xue (2010)'s observations. Different notations are used in different dictionaries for giving countability information. Some dictionaries (e.g. *Oxford Advanced Learner's English-Chinese Dictionary 8* (OALECD8)) use symbols such as [C], [U], [sing.] and [pl.] for countable, uncountable, singular and plural nouns, respectively. Others (e.g. COBUILD4) use more elaborate notations such as *N-count*, *N-count usu. sing*, *N-sing*, *N-Var* etc. While these notations are largely self-explanatory, the countability information given for a certain noun is sometimes equivocal. For example, the countability information given for the first sense of the word *explanation* in OALECD8 is [C, U] with about 10 examples, some of which use the target noun as a countable noun while others use it as uncountable with no indication which is which. It may be argued that the user's guide can help clarify the equivocal information using a general rule of thumb, yet a thorough reading of the explanatory notes and/or user's guides is not something that dictionary users will normally do (Chan 2005). What is more, the information and examples given in the guides may just be clear exemplars and cannot capture all possible scenarios. For example, OALECD8 only specifies, in the references section, that [*sing*] refers to nouns that are always singular and have no plural forms and such nouns are often used with a particular determiner. The example given as illustration is *fillip*, with the pattern (e.g. **a fillip to/for sth**) (Hornby 2013: R17). This may not be useful in helping learners distinguish the different usages of other words, such as *choice*, the second usage of which was specified as [*U, sing*]. Given that the only information provided in the sub-entry is the noun being uncountable or singular without any specifications about the use of a determiner, it is not clear if learners will know whether *choice* should be used with or without an article. The presentation of nouns as both countable and uncountable in learners' dictionaries may, therefore, result in confusion rather than assurance if there are no further specifications about the usage of the nouns in different contexts (Xue 2010). Whether learners can successfully retrieve the required information of a noun from a learner's dictionary, including noun countability and related article choices, is worth investigating.

Objectives

Given that Chinese does not have articles and the countability of an English noun is often different from that of its Chinese equivalent, it is interesting to investigate if the corresponding information provided in a bilingualized dic-

tionary¹ is useful enough for learners' judgment of English noun countability and related article use. The main objective of the study was to investigate how effective a bilingualized dictionary was in helping ESL learners determine the countability of English nouns and associated article use. In so doing, learners' problems in this aspect of language use, as well as the usefulness of Chinese translations in bilingualized dictionaries, would be revealed. As the most popular bilingualized dictionary used by advanced Hong Kong ESL learners is OALECD (Chan 2005), the latest edition available on the market at the time of the study, namely OALECD8 (Hornby 2013), was used for consultation.

Methodology

Participants

A homogenous group of 30 Cantonese university English majors from a local university were invited to participate in the study. Participant selection was based on convenience sampling. These included 8 males and 23 females. Their ages were from 19 years to 35 years at the time of the study. All of them were taking English as their majors. Three of them had learnt English for 14–16 years, and seventeen for 17–19 years, and the rest for 20 years or more. Twenty-four had received a C or above in the Hong Kong Advanced Level Use of English (UE)² exam, 7.5 or above in the International English Language Testing System (IELTS) test, or 5 or above in the Hong Kong Diploma of Secondary Education (HKDSE)³, and the rest had received a D in HKALE, 7 in IELTS, 4 in HKDSE, or C in the Hong Kong Certificate of Education Examination (HKCEE). One had received a C in the General Certificate of Education (GCE) A-level examination. In view of their linguistic backgrounds, the participants could all be regarded as advanced ESL learners. Having many more female students than male students was inevitable, as there are many more female English majors than male English majors in Hong Kong (and probably worldwide).

Two male Cantonese university English graduates from a local university and two female Cantonese university English majors from another local university were invited to serve as a pilot group before the implementation of the actual study. Their ages were from 23 to 30 years of age. All of them had learnt English for 20 years or more. One had received a C, two had received a D, and one had received an E in HKALE respectively. Their linguistics backgrounds were deemed comparable to those of the participants.

Target Nouns

The present study targeted nouns which have varied uses in terms of countability and can be countable/singular/plural or uncountable depending on the context and/or the meaning of the word in that context. A research assistant,

who graduated from an English programme in a local university and was doing a Master's degree in English in another local university, was responsible for selecting potential target nouns and sentence contexts from different online and paper dictionaries. These included *Oxford Dictionaries Online* (<http://www.oxforddictionaries.com/>), *Cambridge Dictionaries Online* (<http://dictionary.cambridge.org/>), *Collins COBUILD Advanced Learner's English Dictionary*, 4th edition (Sinclair 2003), *Oxford Collocations Dictionary for Students of English* (Deuter, Greenan, Noble and Phillips 2002), and *English-Chinese Learner's Thesaurus* (Manser 2009). A total of 50 nouns (with different sentence contexts for each) which satisfied the criteria for inclusion in the study, were identified. From these, the researcher chose 14 for the pilot study based on her experience of ESL learners' use of the words as well as the research assistant's own report, as an English graduate, of his perception of the levels of difficulty of the use of the words with and without the help of a dictionary.

The characteristic of the target nouns can be seen by the word *thought*, which is uncountable when used in sentences such as "*She doesn't give any **thought** to her appearance*" (Cambridge Online) but plural in "*If he wasn't there physically, he was always in her **thoughts***" (Sinclair 2003: 1508). Another example is the word *ability*, which is uncountable as used in "*She utilizes her resources to the best of her **ability***" (Oxford Online), but countable as used in "*She is a woman of considerable **abilities***" (Cambridge Online). The countability of these words in a certain context, whether an article should/can be used with these words in a certain context, and whether these words should be used in the plural form, if any, in a certain context, are all difficult to determine.

Procedures

The research assistant was responsible for implementing the task, which included a noun countability judgement task without and with the use of a dictionary and an introspective questionnaire for each target noun (see *Noun Countability Task* and *Introspective Questionnaires*), and giving initial instructions to the participants. A dictionary packet consisting of a scanned version of the introductory guide and the entries of the target nouns from OALECD8 was provided to each participant for consultation.

Before the implementation of the actual task, a draft version of the task sheet consisting of 14 nouns was piloted with the pilot group (see *Participants*). A total of 14 nouns (with three sentences each) were included in the pilot task. The whole pilot group found that the level of difficulty of the sentences was suitable for English majors and that the instructions were clear. Only some minor changes to the wording of the instructions of the task sheet and the introspective questionnaire were made to facilitate participants' understanding of the requirements. The nouns which were found to be relatively easy to the pilot group (e.g. with one or more sentences accurately completed by most or all of the pilot group participants without the use of a dictionary) were deleted,

because it was thought that the use of a dictionary for such words was not very necessary. As a result, only seven nouns, including *awareness, behavior, feeling, knowledge, reason, thought, understanding*, were chosen for the actual task, and the majority of the participants spent about 1 to 1.5 hours on the completion of the task and the questionnaires.

Noun Countability Task

For each of the seven chosen nouns, three sentence contexts were given, at least one of which was a context in which the noun was used as an uncountable noun, and one in which the noun was used as a singular noun, a plural noun, or a countable noun. The third sentence context could be either. For each sentence context, a word group of which the target noun formed a part was deleted. The participants had to complete each sentence by (i) choosing the correct word group from among three given options and (ii) deciding whether the target noun was used as a countable noun or an uncountable noun in the sentence context. The first option was the target noun in a singular form without an article (e.g. *problem behavior*), the second option was the target noun in plural form without an article (e.g. *problem behaviors*), and the last option was the target noun in singular form with *a/an* (e.g. *a problem behavior*). The task consisted of two parts. The first part of the task was to be completed without the use of a dictionary. This tested the participants' performance without the help of any learning resources and their performance could reflect their initial response of the usage of the target nouns. The second part of the task was done with the use of OALECD8. The participants were required to use the relevant information given in the dictionary entries to help them make decisions about noun countability and related article use for each sentence context. The participants' performance in this part could reflect the effectiveness of the dictionary information they relied on.

Care was taken to ensure that the sentences included in the task were not identical to the examples given in OALECD8, so that the participants had to use their own discretion when choosing the most suitable option for a sentence. However, there were still some sentences or phrases which were very similar to those in OALECD8, such as *There is a feeling of dissatisfaction with the government* (Cambridge Online; cf. 3B in Appendix I) and *a feeling of hunger/excitement/sadness* in OALECD8. Care was also taken to ensure that the context would not give hints to the countability of a target noun, so such items (e.g. a singular verb) were included as part of the options to be chosen. For example, for the sentence *We had not set a date for marriage but there _____* (Sinclair 2003: 1579; cf. 7B in Appendix I), the options included in the task were *was understanding, were understandings* and *was an understanding*, with the verbs included as part of the options. As such, no hints about the number of the target noun would be given, and the participants could not use any obvious cues (such as the incompatibil-

ity of a mass noun with a plural verb or the incompatibility of a singular verb with a plural noun) to assist their selection.

Because the three options given for each sentence followed the same pattern (i.e. in singular form without an article, in plural form without an article, and in singular form with *a/an*), it was inevitable that some options were ungrammatical (e.g. *intense awareneses*). The occasional inclusion of an ungrammatical option was deemed acceptable, given that the purpose of the task was to require participants to decide on the correct form of a noun (and its associated article use) in a certain context.

Introspective Questionnaires

Following a similar methodology employed in the literature (e.g. Chan 2012a, 2012b), after the completion of all the three sentences for each noun, the participants had to complete an introspective questionnaire. The objective of the introspective questionnaires was to gather information about how the information given in the dictionary helped them make their decisions. There were forced-choice questions asking participants from which part of the dictionary entry they found the information they wanted to look for, including English definitions, English examples, Chinese translations of definitions, Chinese translations of examples, codes/abbreviations, and special features such as bold or italics. There were also questions asking participants how they felt about their decisions on ending a search (i.e. they were sure that their decisions were correct, they were not sure if their decisions were correct, and they didn't think they got the relevant information from the dictionary). For participants who were sure that their decisions were correct, they were required to write out the definitions, examples etc. which led them to their decisions and describe how such definitions, examples, etc. showed them that their decisions should be correct. For those who were not sure that their decisions were correct, they were also asked to write out the definitions, examples etc. which led them to their decisions and say why they were doubtful about their decisions. For those who did not think that they could find the relevant information from the dictionary entries, they were asked to describe the difficulties they encountered during consultation and how they made their final decisions. All the participants were also asked to do six ranking questions asking them to rank the overall usefulness of the different kinds of information in doing a search, including English definitions, English examples, Chinese translations of definitions, Chinese translations of examples, codes/abbreviations, and special features (see Appendix II).

Although there were three sentences for each noun and participants may have relied on different pieces of information from the dictionary entries when making a decision about each sentence, to avoid the task being too time-consuming and draining, it was decided that only one questionnaire was used for

each noun, so each participant had to do only a total of seven questionnaires and report seven dictionary consultations, resulting in a total of 210 (i.e. 30 participants x 7 nouns) questionnaires done by for the whole study. Clear instructions were given in the questionnaires requiring participants to specify clearly which definitions, etc. were for which sentence if different definitions, examples, etc. for the three sentence contexts were relied on.

Data Analysis

Language Task: Article Selection and Countability Judgement

The participants' choice of articles for each noun and their corresponding judgment of countability were recorded separately, so even if a correct article was chosen for the completion of a sentence yet the judgement of countability of the target noun was inaccurate, or vice versa, the performance could be reflected independently. This way of processing the language task resulted in a total of six verdicts for each noun (three for article selection and three for grammaticality judgment) for each participant.

- For both article selection and noun countability judgment, the number of correct choices made by the participants on each noun (all three uses), as well as that for the seven nouns together, was counted, and the corresponding accuracy rates were calculated.
- For the three article options (i.e. *ZERO* + singular, *ZERO* + plural (or *a/an* + singular)⁴, and *a/an* + singular), the overall accuracy rate of each was obtained by calculating the total number of correct selections made for that article option out of the total number of target instances of that article option.
- For the two noun countability options (i.e. U or C), the overall accuracy rate of each was obtained in a similar fashion (i.e. calculating the total number of correct judgements made for that countability use out of the total number of target instances of that countability use).

Introspective Questionnaires

Because the introspective questionnaires consisted of both forced-choice questions and open-ended questions, both quantitative and qualitative analyses were performed accordingly.

- For the forced-choice question about participants' feeling on ending a search, the number of participants choosing a certain option (e.g. *sure that the decision was correct*) was counted and the percentage was calculated out of the total number of participants (i.e. 30).

- For the questions about the overall usefulness of each kind of dictionary information, the number of instances that a certain option (e.g. *very useful*) was chosen was counted, and the corresponding percentage was calculated out of 210 (i.e. 30 participants x 7 nouns).
- Answers to the open-ended questions were tabulated and emerging patterns about the participants' problems in using dictionary information and/or strategies employed for problem solving, etc. were identified. The examples, definitions and/or other dictionary information which the participants based their decisions on were referred to when necessary. No quantitative analyses of the open-ended questions were done, as there was no attempt to compare learner problems and/or strategies.

Statistical Analyses

Proportion Z-tests using Excel 2013 were conducted to determine the significance of the differences in the participants' noun countability judgments and article selections without and with the use of a dictionary. A proportion Z-test is a test of the significance of the difference between two proportions from independent samples (Davis 1982). Assuming that the samples are normally distributed, if Z (Z-value) > 1.96 , then there is a significant difference between the two proportions at the 0.05 significance level. Otherwise, the difference can be attributed to sampling errors. Given that the results on countability judgments and article selections were calculated as correctness percentages, Proportion Z-tests were considered the most appropriate statistical analysis for comparison.

Quantitative Results (Language Task and Introspective Questionnaires)

Language Task: Performance on Article Selection without and with the Use of a Dictionary

The participants' overall performance on article selection showed a significant improvement from an overall accuracy rate of 47.6% without the use of a dictionary to an overall accuracy rate of 60.6% with the use of a dictionary ($Z= 4.63$). However, when individual nouns were taken into consideration, only the improvements for *awareness*, *behavior*, *knowledge* and *reason* were statistically significant ($Z= 2.24, 4.37, 2.71, 3.22$ respectively). Although there were some numerical rises or drops for the other three nouns, the differences were not statistically significant (see Table 1).

The article selection accuracy rate rose from 44.3% to 51.3% when the target article was *ZERO* + singular and from 39% to 70% when the target article was *a/an* (+ singular) with the use of a dictionary. There was a numerical drop from 70.8% to 67.5% with the use of a dictionary for *ZERO* + plural. However,

proportion Z-tests showed that only the improvement for *a/an* + singular was statistically significant ($Z= 6.38$) (see Table 1).

Table 1: Language Task: Participants' performance on article selection with-out and with the use of a dictionary

<i>Noun</i>	<i>Article Selection Without a Dictionary</i>		<i>Article Selection With a Dictionary</i>	
	<i>Correct</i>	<i>Incorrect</i>	<i>Correct</i>	<i>Incorrect</i>
awareness	43/90 (47.8%)	47/90 (52.2%)	58/90 (64.4%)	32/90 (35.6%)
behavior	49/90 (54.4%)	41/90 (45.6%)	76/90 (84.4%)	14/90 (15.6%)
feeling	56/90 (62.2%)	34/90 (37.8%)	55/90 (61.1%)	35/90 (38.9%)
knowledge	42/90 (46.7%)	48/90 (53.3%)	60/90 (66.7%)	30/90 (33.3%)
reason	18/90 (20%)	72/90 (80%)	38/90 (42.2%)	52/90 (57.8%)
thought	39/90 (43.3%)	51/90 (56.7%)	51/90 (56.7%)	39/90 (43.3%)
understanding	53/90 (58.9%)	37/90 (41.1%)	44/90 (48.9%)	46/90 (51.1%)
Total	300/630 (47.6%)	330/630 (52.4%)	382/630 (60.6%)	248/630 (39.4%)
<hr/>				
<i>Target Article</i>	<i>Article Selection Without a Dictionary</i>		<i>Article Selection With a Dictionary</i>	
	<i>Correct</i>	<i>Incorrect</i>	<i>Correct</i>	<i>Incorrect</i>
ZERO + singular	133/300 (44.3%)	167/300 (55.7%)	154/300 (51.3%)	146/300 (48.7%)
ZERO + plural (or <i>a/an</i> + singular)	85/120 (70.8%)	35/120 (29.2%)	81/120 (67.5%)	39/120 (32.5%)
<i>a/an</i> + singular	82/210 (39%)	128/210 (61%)	147/210 (70%)	63/210 (30%)

Language Task: Performance on Countability Judgment without and with the Use of a Dictionary

The participants' performance on countability judgment showed an improvement from an overall accuracy rate of 57.1% without the use of a dictionary to an overall accuracy rate of 70.3% with the use of a dictionary. The difference was statistically significant ($Z= 4.87$). When individual nouns were taken into consideration, although improvement was shown for most nouns, again only the improvements for three words, namely, *awareness*, *behavior* and *knowledge*, were statistically significant ($Z=4.54$, 4.5, 3.95 respectively). When the target countability was uncountable, the participants' judgement rose significantly from an overall accuracy rate of 52.9% to 75.1% with the use of a dictionary ($Z= 6.94$), whereas when the target countability was countable, the participants' judgement with the use of a dictionary did not show a statistically significant difference (see Table 2).

Table 2: Language Task: Participants' performance on countability judgment without and with the use of a dictionary

<i>Noun</i>	<i>Countability Judgment Without a Dictionary</i>		<i>Countability Judgment With a Dictionary</i>	
	<i>Correct</i>	<i>Incorrect</i>	<i>Correct</i>	<i>Incorrect</i>
awareness	67/90 (74.4%)	23/90 (25.6%)	88/90 (97.8%)	2/90 (2.2%)
behavior	48/90 (53.3%)	42/90 (46.7%)	76/90 (84.4%)	14/90 (15.6%)
feeling	54/90 (60%)	36/90 (40%)	55/90 (61.1%)	35/90 (38.9%)
knowledge	71/90 (78.9%)	19/90 (21.1%)	88/90 (97.8%)	2/90 (2.2%)
reason	31/90 (34.4%)	59/90 (65.6%)	39/90 (43.3%)	51/90 (56.7%)
thought	42/90 (46.7%)	48/90 (53.3%)	51/90 (56.7%)	39/90 (43.3%)
understanding	47/90 (52.2%)	43/90 (47.8%)	46/90 (51.1%)	44/90 (48.9%)
Total	360/630 (57.1%)	270/630 (42.9%)	443/630 (70.3%)	187/630 (29.7%)
<hr/>				
<i>Target Countability</i>	<i>Countability Judgment Without a Dictionary</i>		<i>Countability Judgment With a Dictionary</i>	
	<i>Correct</i>	<i>Incorrect</i>	<i>Correct</i>	<i>Incorrect</i>
uncountable	238/450 (52.9%)	212/450 (47.1%)	338/450 (75.1%)	112/450 (24.9%)
countable	122/180 (67.8%)	58/180 (32.2%)	105/180 (58.3%)	75/180 (41.7%)

Introspective Questionnaires: Participants' Feelings after a Search

The results of the introspective questionnaire showed that the use of a dictionary did not seem to be very helpful in participants' article selection and countability judgment processes. For five out of the seven target nouns, there were many more dictionary consultations which ended with participants not being sure whether their decisions were correct than the number of dictionary consultations with participants feeling sure that their decisions were correct. In total, only after 38.1% of the consultations (n= 210) were participants sure of their decisions, whereas over 58.1% of the dictionary consultations ended with participants feeling unsure about the correctness of their decisions, and the rest (3.8%) ended with participants feeling that they didn't get the relevant information from the dictionary consulted. The difference between these two feelings was statistically significant (Z= 10.35) (see Table 3).

Table 3: Introspective Questionnaires: Participants' feelings on ending a search

<i>Noun</i>	<i>A. Sure that my decision was correct</i>	<i>B. Not sure whether my decision was correct</i>	<i>C. I don't think I got the relevant info from the dictionary</i>
awareness	9/30 (30%)	19/30 (63.3%)	2/30 (6.7%)
behavior	17/30 (56.7%)	12/30 (40%)	1/30 (3.3%)
feeling	5/30 (16.7%)	22/30 (73.3%)	3/30 (10%)

knowledge	17/30 (56.7%)	13/30 (43.3%)	0 (0%)
reason	11/30 (36.7%)	17/30 (56.7%)	2/30 (6.7%)
thought	7/30 (23.3%)	23/30 (76.7%)	0 (0%)
understanding	14/30 (46.7%)	16/30 (53.3%)	0 (0%)
Total	80/210 (38.1%)	122/210 (58.1%)	8/210 (3.8%)

Introspective Questionnaires: Usefulness of Dictionary Information on Article Selection and Countability Judgment

With regard to the kinds of information that participants found most useful, consonant with what has been found in the literature (Chan 2012a), examples came first, which were found to be the most useful in 46.7% of the total number of consultations. Unlike the results of previous literature such as Chan (2011), codes and abbreviations came next. They were regarded as very useful in 43.8% of total number of consultations, just slightly lower than that for examples and much higher than that for English definitions (30%). Chinese examples, on the other hand, were mostly thought to be of no use (35.7%) (see Table 4).

Table 4: Introspective Questionnaires: The usefulness of different kinds of dictionary information

Information	very useful	useful	of little use	of no use
English definitions	63/210 (30%)	111/210 (52.9%)	30/210 (14.3%)	6/210 (2.9%)
English examples	98/210(46.7%)	97/210 (46.2%)	14/210 (6.7%)	1/210 (0.5%)
Chinese definitions	38/210 (18.1%)	85/210 (40.5%)	52/210 (24.8%)	35/210 (16.7%)
Chinese examples	13/210 (6.2%)	50/210 (23.8%)	72/210 (34.3%)	75/210 (35.7%)
Codes & abbreviations	92/210 (43.8%)	79/210 (37.6%)	32/210 (15.2%)	7/210 (3.3%)
Special features	52/210 (24.8%)	89/210 (42.4%)	49/210 (23.3%)	20/210 (9.5%)

A careful comparison between participants' performance on the noun countability task and their responses on the introspective questionnaires showed that for cases where they were sure that their answers were correctly made with the use of a dictionary, many instances of their article selection and/or countability judgments were actually incorrect. A total of 80 dictionary consultations (out of 210; 38.1%) ended with participants being sure that their answers were correct, yet 59 of those (73.8%) had one or more incorrect decisions made (there were three questions for each of the seven target nouns, with each question requesting two decisions to be made, one on article selection and one on countability judgment; see *Noun Countability Task*). If individual decisions were taken in consideration, a total of 148 decisions (out of 480; 30.8%) were actually wrong

despite participants being sure about their decisions. The words *reason* and *understanding* showed the poorest accuracy-confidence match, with all the decisions for all the questions being wrong despite participants' feeling sure about their decisions after dictionary consultation⁵.

Qualitative Results (Introspective Questionnaires)

Problems Reported and Strategies Employed

DISTINCTIONS BETWEEN VARIOUS USES OF A TARGET NOUN: A number of problems with dictionary information were reported in the introspective questionnaires about participants' use of a bilingualized dictionary for article selection and countability judgement. One prevalent problem was difficulties in discerning the subtle differences between the countable, uncountable and singular uses of a target noun even after reading relevant dictionary information, especially when the definitions/examples were quite similar to each other⁶:

After consulting the dictionary, I thought treating "awareness" as uncountable or singular was actually the same. I was wondering why the dictionary doesn't specify the two usages in two sections if they do make a difference. So, I encountered a problem about differentiating the countability of the word (Student 2);

Especially usage 6, the code at the beginning tells the user that "feeling" is uncountable with that specified meaning, but then after the first example sentence, there suddenly appears a prepositional phrase structure "for sb/sth" followed by an example demonstrating a countable usage. This really made me confused (Student 2).

ACCEPTABILITY OF AN ARTICLE: Another major problem was difficulties in deciding whether an article was acceptable or needed in a certain context. For nouns that were marked as singular in the dictionary, such a problem was more widespread.

It's hard to determine when to put an article before the noun (Student 15);

"Awareness" is not something tangible, so I feel awkward when it can go with "an" together (Student 24; after reading a dictionary example showing "an awareness");

Even though the dictionary says that it should be singular, I'm not sure if the article "a" should be included or not (Student 21);

ADEQUACY OF DICTIONARY EXAMPLES: There were concerns about the adequacy of examples when participants could not find enough examples for them to base their decisions on, yet occasionally the presence of too many definitions/examples, or too many usages of the same noun, also resulted in confusion:

There isn't an example in the dictionary that fully explains both [U] and [C], like when do we use [U]/[C] (Student 16);

I didn't know why for usage 2, the dictionary doesn't show any examples of the plural form of "understanding" to show clearly when the word is treated as a countable noun instead of [usually singular] (Student 2);

There are too many definitions, as well as examples, which confuse me (Student 26);

Hard to distinguish between the three English definitions of "knowledge" (Student 27);

Not sure if it is countable or uncountable because the dictionary shows both for that definition (Student 22).

STRATEGIES EMPLOYED IN MAKING DECISIONS: When participants thought that they could not find the relevant information from the dictionary or were doubtful about the usefulness/appropriateness of dictionary information, strategies such as guessing, or using intuition and/or elimination were employed.

I used my intuition (Student 29);

By reading the definitions again and guessing (Student 26).

Sometimes they based their decision on their own understanding of the word and their previous encounters instead of on the dictionary entry or by guessing.

Based on my own understanding of the word and my observation. I don't think I have seen the plural form of "behavior" (Student 24);

I think there are many different problems and good behaviors of a child. There could be a "s" after behavior (Student 19).

Occasionally Chinese was used as the basis of decisions. This strategy was employed when participants thought that the difference in meaning resulted in different countability based on the countability of the corresponding Chinese nouns.

The Chinese definition of the first (感覺) and the third (情感) clarified the difference (i.e. the former more physical and senses while the latter focusses on how the heart feels) (Student 23);

All I knew is just that if "feeling" means "感覺", then it is countable. However, I was still not sure whether I could say "feelings of dissatisfactory" (Student 2).

Unnoticed Problems

The previous section revealed the problems that participants reported that they had encountered, leading to their uncertainty of the accuracy of their choices. Those were the problems that they themselves were aware of. However, careful scrutiny of the introspective reports showed that some participants had misun-

derstood, misread or been misled by certain dictionary information, resulting in inaccurate countability judgement and/or article selection they were not aware of. This section will examine these unnoticed problems⁷.

CO-OCCURRENCE OF A/AN WITH AN UNCOUNTABLE NOUN: One prevalent problem was participants' unawareness of the acceptability of the co-occurrence of *a/an* with an uncountable noun in certain contexts. This misconception was often reinforced by the [U] code found in the dictionary entry, despite the fact that [sing] was sometimes also given in the same dictionary entry to show another use of the noun. The following reports on the word *knowledge*, where participants chose *limited knowledge* instead of *a limited knowledge* in *He has a limited knowledge of French* (Cambridge Online) and/or *intimate knowledge* instead of *an intimate knowledge* in *She has an intimate knowledge of the Asian market* (Deuter et al 2002), demonstrate this.

It must be [U] (from the dictionary). There are no countable nouns, so I can eliminate the others (Student 12);

It shows that "knowledge" is uncountable (Student 14);

No matter which meaning it carries, there's no way for it to be countable (Student 24).

Interestingly, there were also cases when participants were misled that *a/an* could always be used for uncountable nouns which could also be used as singular nouns. An example can be seen from the word *awareness*, which was shown in the dictionary as [U, sing]. There were a number of examples showing the use of the word with *an* (e.g. *an awareness of the importance of eating a healthy diet; develop an awareness of how the Internet can be used.; an increasing awareness of sth; a general awareness that this is not the solution* (Hornby 2013: 121)). Therefore, many participants, after reading the examples with the co-occurrence of *an* with the target noun, thought that the indefinite article was needed in all contexts without being aware of the examples which suggested otherwise (e.g. *complete lack of awareness of the issues involved; environmental awareness; energy awareness week* (ibid: 121)).

When I look at those examples on the dictionary as noun phrases, it always has an article (a/an) in front of how awareness is described (Student 8);

Now I realize the nature of the word, awareness, that it is an uncountable singular noun. Option C could be eliminated and an article is needed (Student 11);

Because in the dictionary entries, there is usually an article before awareness. Except for "lack of awareness", the rest in those examples is followed by an article at the beginning (Student 20).

RELATED WORDS/PHRASES IN DICTIONARY EXAMPLES AND TARGET SENTENCE: A related word in a dictionary example or definition sometimes triggered an

inappropriate comparison and misled participants into thinking that the contexts of the target sentence (in the noun countability task) and the dictionary sentence(s) were the same and the latter could be appropriately used as a model. A case in point was the phrase "with reason" in the dictionary example *She complained with reason* (= rightly) (ibid: 1708), which was taken as a model for validating participants' inappropriate selection of ZERO for the phrase *without justified reason in the sentence *Their goal is simply to cause terror without a justified reason* (Oxford Online; see Appendix I) on the basis of the similarity between *with reason* and *without (a) justified reason*. The difference in meaning, and thus difference in article use, between the two phrases escaped participants' notice.

Another example of inappropriate correspondence made between two similar phrases was *feeling*, where *a nasty feeling* in the example *I had a nasty feeling that we were lost* (Hornby 2013: 761) was used as a model to confirm participants' choice of *a* for **a genuine feeling* in the sentence *It all feels so lacking in genuine feeling* (Oxford Online; see Appendix I).

Sentence 5C resembles the example "She complained with reason (= rightly), so it should be "justified reason" (Student 11);

Genuine is an adjective while nasty is also adjective so same rule should be applied (Student 20).

INFLUENCE OF CHINESE TRANSLATIONS: Chinese translations of dictionary definitions/examples were also sometimes sources of problems. Participants who thought that a target English phrase should best be translated into a certain Chinese equivalent were inclined to model their construction on corresponding English dictionary examples/definitions. Student 12, who (mistakenly) thought that the Chinese phrase 理性 (*the power of the mind to think in a logical way, to understand and have opinions, etc.* (Hornby 2013: 1708)) should best represent the meaning of *reason* in the target sentence *Their goal is simply to cause terror without a justified reason* (Oxford Online; see Appendix I), chose ZERO instead of *a* for *a justified reason* on the basis of the dictionary example (i.e. *Only human beings are capable of reason* (Hornby 2013: 1708)) given for that definition of the target noun. Similarly, Student 8 chose *thoughts* (*a person's mind and all the ideas that they have in it when they are thinking* (ibid: 2176)) instead of *thought* for the sentence *She doesn't give any thought to her appearance* (Cambridge Online; see Appendix I) based on the English example *My thoughts turned to home* (Hornby 2013: 2176) of the Chinese definition 心思; 思想. A similar problem was seen with the article selection and countability judgment of the word *understanding* by Student 26 (details omitted).

The definition of 理性 make me choose it (Student 12);

It is clear that Question A is '心思; 思想' (Student 8);

As I know, from the Chinese translations of definitions, the agreement "協議" is suitable in this sentence, so option c is correct (Student 26).

Discussion

The results of the study show that although learners may be aware of the importance of contexts on word usage and that different senses of the same word may be associated with different linguistic structures (Chan 2012a), they may not possess the ability to identify the correct sense of a target noun in a certain context so as to determine the countability of the noun and/or its associated article use. The results that no significant improvements were found in the participants' noun countability and/or related article use judgements of many target nouns after dictionary consultation also confirm the researcher's speculation that noun countability information in learners' dictionaries may not be useful. Though learners are sometimes aware of the inadequacy of dictionary entries and their difficulties in interpreting relevant dictionary information, many of their interpretation problems go unnoticed or even reinforced by their dictionary consultation processes. The different structures associated with the countable and uncountable uses of an English noun, as well as why and how they are different, are not always well explained or apparent in learners' dictionaries. A detailed look at some dictionary entries may help explain this.

In the middle of the dictionary entry for the target noun *awareness* after the syntactic specifications (*[U, sing], ~ (of sth), ~ (that...)*) and English and Chinese definitions, some example phrases/sentences (and their Chinese translations) are given, including

- (1) *an awareness of the importance of eating a healthy diet^s,*
- (2) *There was an almost complete lack of awareness of the issues,*
- (3) *It is important that students develop an awareness of how the Internet can be used*
- (4) *to raise/heighten/increase public awareness of sth,*
- (5) *a greater/a growing/an increasing awareness of sth* (Hornby 2013: 121).

It can be seen that some examples use the target noun as a singular noun with *an* (example 1), and others as an uncountable noun with *ZERO* (example 2), yet all examples follow the syntactic specifications of having a post-modifier with *of*, showing that the presence of a post-modifier does not impact on its countability. It is not clear whether a pre-modifier will have any effects on the countability of the noun, as there are both examples of singular and uncountable uses without a pre-modifier (examples 1–3), as well as examples of singular and uncountable uses with a pre-modifier (examples 4–5). Learners will be left to wonder whether a pre-modifier like *public* (example 4) is different from *growing/increasing* (example 5), and what the difference between them is, if any. It is also not clear whether the target noun used as a direct object of a transitive

verb (e.g. *to raise public awareness*; example 4) is always uncountable. The meaning of the target noun does not seem to have an impact on countability either, as all the examples are grouped under the same sense, and only one single sense of the word is given in the dictionary entry. It is, thus, not difficult to understand how and why learners may be led astray by noun countability information in dictionary entries, resulting in misinterpretations and encoding problems.

As for the usefulness of Chinese translations, although there are no articles in Chinese and the concepts of countability in English and Chinese differ, Chinese translations in a bilingualized dictionary do play a role in Chinese ESL learners' noun countability and/or article selection determination. Their role is manifested in learners' modelling on the syntactic structure of an English definition/example with a Chinese translation which is thought to best represent the meaning of a target noun in a certain context. The strategy can be proceduralized as follows:

For a certain sentence context (C) which motivates dictionary consultation, learners will

- (i) identify the most suitable Chinese translation for the target English word,
- (ii) locate an English example/definition given for that chosen Chinese translation,
- (iii) model on the syntactic structure of the English example/definition;
- (iv) insert the target English word with the associated syntactic structure in the dictionary example/definition into the sentence context (C).

While a translation in a bilingualized dictionary (e.g. Chinese) is preferably insertable in a target sentence of the same language (i.e. Chinese) (Gauton 2008), such a strategy of inserting a corresponding English phrase (based on the Chinese translation) into an English sentence is apparently not an appropriate dictionary consultation strategy. Very few equivalent words in two languages have precisely the same meanings. Two different languages may also have different syntactic requirements (e.g. verb transitivity) for corresponding vocabulary items or different syntactic patterns "to determine certain aspects of experience" (Gauton 2008:112), so the context for which a Chinese word is appropriate may not be a suitable context for the corresponding English phrase, or vice versa. These meaning and usage discrepancies are not usually reflected in the Chinese translations in an English–Chinese bilingualized dictionary, so learners who rely on a Chinese translation equivalent for the syntactic requirements of a target English word will likely encounter problems. Despite the usefulness of translations for decoding purposes (Cowie 1999), our results show that they are not equally useful for encoding. Where the source and target languages differ in a certain aspect of grammar, such as noun countability, translations in another language may even undesirably affect learners'

encoding performance, providing illusive confirmation for learners' incorrect language choices like what have been observed in the present study.

Implications

The problems identified in the present study have important pedagogical and lexicographical implications. They inform ESL teachers and lexicographers of learners' dictionary consultation problems not just in the area of noun countability and/or article selection but also of the possible negative impacts of translations in bilingualized dictionaries on encoding. Given that article use is often regarded by ESL teachers as their students' number one difficulty (Covitt 1976; cited in Celce-Murcia and Larsen-Freeman 1983), it is important for ESL teachers to design suitable teaching programmes to help remedy the problems. Although it has been argued that learners may be able to learn the English article system better via exposure to the input (Lightbown and Spada 2013), explicit linguistic analyses using ample authentic data and relevant metalinguage should also be useful (Chan 2016). Exercises similar to the noun countability task used in the present study can be designed for an advanced ESL classroom to alert learners to the variability of noun countability and related article use as well as the importance of the context in making relevant judgments. Teachers can then engage learners in awareness raising discussions to explore the differences in the meanings of a target noun resulting from the use of a different article (*ZERO* inclusive) and/or different countability. It is also advisable for ESL teachers to alert learners to the different (or similar) syntactic requirements of equivalent vocabulary items in two languages when using a bilingualized dictionary. While both exposure to the input and explicit teaching are needed, the use of learners' dictionaries is also essential for noun countability and article use judgments, so it is advisable that ESL teachers encourage learners to make full use of such self-learning resources and introduce relevant dictionary skills training. However, they should also be made aware of the possible traps that they may fall into. Dictionary skills training programmes for advanced ESL learners can incorporate a metalinguistic analysis component where learners are engaged in activities in discovering such differences (or similarities) and the varied functions of dictionary information for encoding and for decoding.

Learners' dictionaries, being the most important self-learning resources, are indispensable in such teaching programmes. However, the information in learners' dictionaries about the syntactic environments and/or semantic specifications which impact on the countability of a noun and its associated article use is not user-friendly enough and needs refinement, as dictionary information has to result in correct language production if it is to be user-friendly for encoding (Dziemianko 2006). Lexicographers are recommended to improve the arrangement of relevant dictionary entries and include user-friendly information to facilitate learners' understanding and advance their language produc-

tion. For dictionaries which use notations such as [C, U] or [U, sing] in the same sub-entry of a noun (e.g. OALECD), the subtle differences between the countable and uncountable uses of the target noun have been made indistinguishable. Such notations are better replaced by separate countability notations. At least one example sentence should be given for a context which triggers a different countability and/or related article use. The contexts under which a usually uncountable noun (e.g. *knowledge*) can be used as a singular noun with an article (e.g. *a knowledge of*) should better be specified clearly using guiding information. For example, while OALECD8 does give an example sentence (*He has a wide knowledge of painting and music* (Hornby 2013: 1157)) with *a/an* used with the word *knowledge* after the guiding phrase *knowledge of/about sth*, there is no clear indication when *a/an* can be used with the target word. With a clear guiding note such as "can be used with *a/an* in patterns like *a/an (adj) knowledge of/about*", then the contexts in which an article is allowed will become more salient and apparent.

Conclusion

In this paper, I have reported on the results of a noun countability determination and article selection task without and with the use of a bilingualized dictionary. It is found that noun countability is indeed a problematic area even for advanced ESL learners. Although noun countability information and examples of associated article use are included in a bilingualized dictionary, the amount of information included is not necessarily adequate in capturing the subtlety of this aspect of English grammar. The distinction between the meaning of a noun in a certain context and its meaning in another context is often not clearly presented, so the subtle differences between the countable and uncountable uses of many nouns are difficult for learners to discern, and misinterpretation or misapplication of dictionary information often results. The dictionary consultation problems identified in the present study can be grave, as revealed by participants' unawareness of their article mis-selections and/or countability misjudgments and their confidence in the "accuracy" of their decisions. Given that learners' dictionaries are widely accepted as useful self-learning resources, errors resulting from misreading/misinterpretation of dictionary information will often go unnoticed and may even be firmly ingrained in learners' minds, especially when there are superficially equivalent or similar structures mistakenly taken as models for production. Our results provide lexicographers with signposts to the selection and arrangement of noun information to be included in a learner's dictionary. It is important for lexicographers to anticipate the potential problems that learners have in interpreting noun countability judgment and/or related article use, so more explicit information should be provided in a more user-friendly manner to enable learners to unlock the variations in the structures related to the different countability of a noun.

Acknowledgements

The work described in this article was fully supported by the Hong Kong Research Grants Council (Project Number: CityU 11400614). The support of the council is acknowledged. I would also like to thank all the participants for their participation, my research assistants for their administrative help, and the consultants in the City University Statistical Consulting Unit for their expert advice on statistical analyses.

Endnotes

1. A bilingual dictionary relates the vocabularies of two languages (e.g. English and Chinese) together simply by means of translation equivalents (Hartmann and James 1998). A bilingualized dictionary (e.g. OALECD) is a dictionary whose entries have been translated in full or in part into another language. Definitions and examples in both the target and source languages are included (see also Hartmann 1994; James 1994; Marelllo 1998).
2. The Hong Kong Advanced Level Use of English (UE) examination aimed to test students' ability to understand and use English at a level required for tertiary education and/or for future employment. (http://www.hkeaa.edu.hk/DocLibrary/HKALE/Subject_and_Syllabuses/2013/2013as-e-ue.pdf). It was normally taken by F.7 students in Hong Kong who had completed their two-year matriculation studies. UE Grade E was regarded as equivalent to Grade E in the GCE A level examinations.
3. The Hong Kong Diploma of Secondary Education examination aims to measure the attainment of students upon their completion of six years of secondary education and has been the only public examination in the new 3-3-4 education system in Hong Kong since 2012. 5** is the highest grade that students can attain for a certain subject, followed by 5* and 5.
4. For target nouns which were used countably, sometimes more than one answer (both ZERO + plural and *a/an* + singular) was acceptable. For example, it is acceptable to say "*There are feelings of dissatisfaction*" or "*There is a feeling of dissatisfaction.....*".
5. The purpose of this paragraph is to identify possible problems with dictionary information, so corresponding data when participants were not sure whether their decisions were correct and those when they thought they could not find the relevant information from the dictionary are not reported.
6. All the introspective reports included in this section were given by participants who were not sure whether their decisions were correct after dictionary consultation, or those who didn't think they got the relevant information from the dictionary.
7. All the introspective reports included in this section were given by participants who were sure that their decisions were correct after dictionary consultation but who had made inappropriate article choices and/or countability judgments.
8. The numbers before the example sentences/phrases are not given in the OALECD8. They are just given in this paper for clarity purposes.
9. The sources of the sentences are given in the appendix for acknowledgement. They were not included in the task sheets given to the participants. The items in bold represent the versions given by the cited dictionaries, but alternative answers may be acceptable for some sentences.

References

Dictionaries

- Deuter, M., J. Greenan, J. Noble and J. Phillips.** 2002. *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.
- Hornby, A.S.** 2013. *Oxford Advanced Learner's English-Chinese Dictionary*. Eighth edition. Hong Kong: Oxford University Press.
- Manser, M.H.** 2009. *English-Chinese Learner's Thesaurus*. Hong Kong: Commercial Press.
- Sinclair, J.** 2003. *Collins COBUILD Advanced Learner's English Dictionary*. Fourth edition. Glasgow: HarperCollins.
- <http://dictionary.cambridge.org/>. Retrieved on 17/10/2016.
- <http://www.oxforddictionaries.com/>. Retrieved on 17/10/2016.

Other Literature

- Allan, K.** 1980. Nouns and Countability. *Language* 56(3): 541-567.
- Amuzie, G.L. and P. Spinner.** 2013. Korean EFL Learners' Indefinite Article Use with Four Types of Abstract Nouns. *Applied Linguistics* 34(4): 415-434.
- Butler, Y.G.** 2002. Second Language Learners' Theories on the Use of English Articles: An Analysis of the Metalinguistic Knowledge Used by Japanese Students in Acquiring the English Article System. *Studies in Second Language Acquisition* 24: 451-480.
- Celce-Murcia, M. and D. Larsen-Freeman.** 1983. *The Grammar Book. An ESL/EFL Teacher's Course*. Rowley, MA: Newbury House.
- Chan, A.Y.W.** 2005. Tactics Employed and Problems Encountered by University English Majors in Hong Kong in Using a Dictionary. *Applied Language Learning* 15(1-2): 1-28.
- Chan, A.Y.W.** 2011. Bilingualized or Monolingual Dictionaries? Preferences and Practices of Advanced ESL Learners in Hong Kong. *Language, Culture and Curriculum* 24(1): 1-21.
- Chan, A.Y.W.** 2012a. The Use of a Monolingual Dictionary for Meaning Determination by Advanced Cantonese ESL Learners in Hong Kong. *Applied Linguistics* 33(2): 115-140.
- Chan, A.Y.W.** 2012b. 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.
- Chan, A.Y.W.** 2016. How Much do Cantonese ESL Learners Know About the English Article System? *System* 56: 66-77.
- Chan, A.Y.W. and Y. Loong.** 1999. Establishing Criteria for Evaluating a Learner's Dictionary. Berry R., B. Asker, K. Hyland and M. Lam (Eds.). 1999. *Language Analysis, Description and Pedagogy*: 298-307. Hong Kong: Hong Kong University of Science and Technology.
- Covitt, R.** 1976. *Some Problematical Grammar Areas for ESL Teachers*. M.A. in TESL Thesis. Los Angeles: University of California.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- Davis, L.M.** 1982. American Social Dialectology: A Statistical Appraisal. *American Speech* 57(2): 83-94.
- Dziemianko, A.** 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Lexicographica Series Maior 130. Tübingen: Niemeyer.

- Dziemianko, A.** 2012. *Noun and Verb Codes in English Monolingual Dictionaries for Foreign Learners: A Study of Usefulness in the Polish Context*. Poznań: Wydawnictwo Naukowe UAM.
- Gauton, R.** 2008. Bilingual Dictionaries, the Lexicographer and the Translator. *Lexikos* 18: 106-118.
- Greenbaum, S. and G. Nelson.** 2009. *An Introduction to English Grammar*. Third edition. Harlow: Pearson Longman.
- Hall, C.J., D. Schmidtke and J. Vickers.** 2013. Countability in World Englishes. *World Englishes* 32(1): 1-22.
- Hartmann, R.R.K.** 1994. Bilingualised Versions of Learner's Dictionaries. *Fremdsprachen Lehren und Lernen* 23: 206-220.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hua, D. and T.H.-T. Lee.** 2005. Chinese ESL Learners' Understanding of the English Count–Mass Distinction. Dekydtspotter, L., R.A. Sprouse and A. Liljestrang (Eds.). 2005. *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference, April 15–18, 2004, Indiana University Bloomington*: 138-149. Somerville, MA: Cascadilla Proceedings Project.
- Huebner, T.** 1983. *A Longitudinal Analysis of the Acquisition of English*. Ann Arbor: Karoma.
- Ionin, T., M.L. Zubizarreta and S.B. Maldonado.** 2008. Sources of Linguistic Knowledge in the Second Language Acquisition of English Articles. *Lingua* 118: 554-576.
- James, G.** 1994. Towards a Typology of Bilingualized Dictionaries. James, G. (Ed.). 1994. *Meeting Points in Language Studies: A Festschrift for Ma Tailai. Working Papers*: 184-196. Hong Kong: Hong Kong University of Science and Technology Language Center.
- Lightbown, P.M. and N. Spada.** 2013. *How Languages are Learned*. Fourth edition. Oxford: Oxford University Press.
- Lock, G.** 1996. *Functional English Grammar: An Introduction for Second Language Teachers*. Cambridge: Cambridge University Press.
- Marello, C.** 1998. Hornby's Bilingualized Dictionaries. *International Journal of Lexicography* 11(4): 292-314.
- Master, P.** 1987. *A Cross-linguistic Interlanguage Analysis of the Acquisition of the English Article System*. Unpublished Ph.D. Thesis. Los Angeles: University of California.
- Miller, J.** 2005. Most of ESL Students Have Trouble with the Articles. *International Education Journal*. ERC2004 Special Issue 5(5): 80-88.
- Miller J.** 2006. An Investigation into the Effect of English Learner's Dictionaries on International Students' Acquisition of the English Article System. *International Education Journal* 7(4): 435-445.
- Murphy, S.** 1997. *Knowledge and Production of English Articles by Advanced Second Language Learners*. Ph.D. Thesis. Austin: University of Texas.
- Ogawa, M.** 2008. The Acquisition of English Articles by Advanced EFL Japanese Learners: Analysis Based on Noun Types. *Journal of Language and Culture* 3: 133-151.
- Parrish, B.** 1987. A New Look at Methodologies in the Study of Article Acquisition for Learners of ESL. *Language Learning* 37: 361-383.
- Pica, T.** 1985. The Selective Impact of Classroom Instruction on Second-Language Acquisition. *Applied Linguistics* 6(3): 214-222.
- Robertson, D.** 2000. Variability in the Use of the English Article System by Chinese Learners of English. *Second Language Research* 16(2): 135-172.
- Schneider, E.W.** 2011. *English Around the World: An Introduction*. Cambridge: Cambridge University Press.

- Snape, N.** 2008. Resetting the Nominal Mapping Parameter in L2 English: Definite Article Use and the Count–Mass Distinction. *Bilingualism: Language and Cognition* 11(1): 63-79.
- Tang, E.Y.** 2006. Linguistic Environment as a Determinant of English Countability. *Proceedings of the 11th Conference of Pan-Pacific Association of Applied Linguistics*: 266-280. Retrieved from <http://www.paaljapan.org/resources/proceedings/PAAL11/pdfs/22.pdf>.
- Thomas, M.** 1989. The Acquisition of English Articles by First- and Second-Language Learners. *Applied Psycholinguistics* 10: 335-355.
- Wisniewski, E.J., C.A. Lamb and E.L. Middleton.** 2003. On the Conceptual Basis for the Count and Mass Noun Distinction. *Language and Cognitive Processes* 18(5/6): 583-624.
- Xue, M.** 2010. Countable or Uncountable? That is the Question — Lexicographic Solutions to Nominal Countability in Learner's Dictionaries for Production Purposes. *Lexikos* 20: 540-558.
- Yoon, K.K.** 1993. Challenging Prototype Descriptions: Perception of Noun Countability and Indefinite vs. Zero Article Use. *IRAL* 31(4): 269-289.
- Zobl, H.** 1980. The Formal and Developmental Selectivity of L1 Influence on L2 Acquisition. *Language Learning* 30(1): 43-57.
http://www.hkeaa.edu.hk/DocLibrary/HKALE/Subject_and_Syllabuses/2013/2013as-e-ue.pdf. Retrieved on 17/10/2016.

Appendix I

Noun Countability Task

For each of the following nouns, you need to

- i. decide how the noun should be used in each of the given three sentences by choosing the most appropriate answer from the given options and writing the answer (e.g. a, or b, or c) in the blank;
 - ii. decide whether the noun is used as a countable noun (C) or an uncountable noun (U) in each of the three sentences by circling the appropriate answer (i.e. C or U) after each.
1. awareness
 - A. The body is an organism with _____ of itself (Oxford Online).⁹ C/U
 - a. intense awareness
 - b. intense awarenesses
 - c. **an intense awareness**
 - B. Politicians now have _____ of these problems (Deuter et al. 2002: 49). C/U
 - a. **much greater awareness**
 - b. much greater awarenesses
 - c. a much greater awareness
 - C. She always seems to have _____ of her sidekick status (Oxford Online). C/U
 - a. affecting awareness
 - b. affecting awarenesses
 - c. **an affecting awareness**
 2. behavior
 - A. Teachers can't always respond effectively to _____ (Deuter et al. 2002: 62). C/U
 - a. **problem behavior**
 - b. problem behaviors
 - c. a problem behavior
 - B. Children should be rewarded for _____ (Manser 2009: 34). C/U
 - a. **good behavior**
 - b. good behaviors
 - c. a good behavior
 - C. These eating patterns are _____ (Sinclair 2003: 117). C/U
 - a. learned behavior
 - b. learned behaviors
 - c. **a learned behavior**
 3. feeling
 - A. It all feels so lacking in _____, genuine emotion, genuine interest (Oxford Online). C/U
 - a. **genuine feeling**
 - b. genuine feelings
 - c. a genuine feeling

- B. There _____ of dissatisfaction with the government (Cambridge Online). C/U
- a. is feeling
 - b. are feelings
 - c. **is a feeling**
- C. It's incredible that Peter can behave with such stupid lack of _____ (Sinclair 2003: 526). C/U
- a. **feeling**
 - b. feelings
 - c. a feeling
4. knowledge
- A. All of these plans require _____ in order to carry out the operation in a timely and accurate manner (Oxford Online). C/U
- a. **insider knowledge**
 - b. insider knowledges
 - c. an insider knowledge
- B. He has _____ of French (Cambridge Online). C/U
- a. limited knowledge
 - b. limited knowledges
 - c. **a limited knowledge**
- C. She has _____ of the Asian market (Deuter et al. 2002: 446). C/U
- a. intimate knowledge
 - b. intimate knowledges
 - c. **an intimate knowledge**
5. reason
- A. They had _____ to believe that there could be trouble (Sinclair 2003: 1192). C/U
- a. **reason**
 - b. reasons
 - c. a reason
- B. The police have _____ to believe that he is guilty (Cambridge Online). C/U
- a. **reason**
 - b. reasons
 - c. a reason
- C. Their goal is simply to cause terror without _____ (Oxford Online). C/U
- a. justified reason
 - b. justified reasons
 - c. **a justified reason**
6. thought
- A. She doesn't give any _____ to her appearance (Cambridge Online). C/U
- a. **thought**
 - b. thoughts
 - c. a thought
- B. Alice had been so deep in _____ that she had walked past her car without even seeing it (Sinclair 2003: 1508). C/U
- a. **thought**
 - b. thoughts
 - c. a thought

- C. If he wasn't there physically, he was always in _____ (Sinclair 2003: 1508). C/U
- a. her thought
 - b. **her thoughts**
 - c. a thought
7. understanding
- A. They have to have _____ of computers in order to use the advanced technology (Sinclair 2003: 1579). C/U
- a. basic understanding
 - b. basic understandings
 - c. **a basic understanding**
- B. We had not set a date for marriage but there _____ between us (Sinclair 2003: 1579). C/U
- a. was understanding
 - b. were understandings
 - c. **was an understanding**
- C. There _____ between Wilson and myself (Sinclair 2003: 1579). C/U
- a. **was complete understanding**
 - b. were complete understandings
 - c. was a complete understanding

Appendix II

Introspective Questionnaire (for Part II of the Noun Countability Task)

This questionnaire is meant to be completed in the course of your dictionary consultation. Answer the following questions after you have finished each sentence.

Noun (_____)

1. On ending the search, how do you feel?
 - A. Sure that my decision was correct (go to Questions 2 to 5)
 - B. Not sure whether my decision was correct (go to Questions 6 to 8)
 - C. I don't think I got the relevant info from the dictionary (go to Questions 9 to 11)
2. From which part of the dictionary entry did you find the information you wanted to look for? You can choose more than one.
 - A. English Definitions
 - B. English Examples
 - C. Chinese translations of definitions
 - D. Chinese translations of examples
 - E. Codes/Abbreviations
 - F. Special features (e.g. bold, italics, etc.)
 - G. Others (please specify: _____)
3. Write out the **definitions, examples, codes/abbreviations, etc.** which led you to your decision. If you used different definitions, examples, etc. for the three different sentences (i.e. A, B or C), please specify clearly which definitions, etc. are for which sentence.

4. How did your answer to Question 3 show you that your decision should be correct? Please specify the sentences (i.e. A, B, or C) you are discussing if necessary.
5. Is there any other information from the entry (other than your answer to Q.4) which showed you that the other options should be incorrect? If so, what?

(Please continue by answering Questions 12–17)

6. Which of the following helped you make your decision? You can choose more than one.

- A. English Definitions
- B. English Examples
- C. Chinese translations of definitions
- D. Chinese translations of examples
- E. Codes/Abbreviations
- F. Special features (e.g. bold, italics, etc.)
- G. Others (please specify: _____)

7. Write out the **definitions, examples, codes/abbreviations, etc.** which led you to your decision. If you used different definitions, examples, etc. for the three different sentences (i.e. A, B or C), please specify clearly which definitions, etc. are for which sentence.

8. Why were you doubtful about your decision? Please specify the sentences (i.e. A, B, or C) you are discussing if necessary.

(Please continue by answering Questions 12–17)

9. Did you consult the following information during the search?

- | | |
|--|--------|
| A. English Definitions | Yes/No |
| B. English Examples | Yes/No |
| C. Chinese translations of definitions | Yes/No |
| D. Chinese translations of examples | Yes/No |
| E. Codes/Abbreviations | Yes/No |
| F. Special features (e.g. bold, italics, etc.) | Yes/No |
| G. Others (please specify: _____) | |

10. What difficulties did you encounter when doing the search? Please specify the sentences (i.e. A, B, or C) you are discussing if necessary.

11. How did you make your final decision? Please specify the sentences (i.e. A, B, or C) you are discussing if necessary.

(Please continue by answering Questions 12–17)

12. How would you rank the overall usefulness of the English definitions in doing the search? (Please circle the answer.)

Very useful	Useful	Of little use	Of no use
4	3	2	1

13. How would you rank the overall usefulness of the English examples in doing the search? (Please circle the answer.)

Very useful	Useful	Of little use	Of no use
4	3	2	1

14. How would you rank the overall usefulness of the Chinese translations of the definitions in doing the search? (Please circle the answer.)

Very useful	Useful	Of little use	Of no use
4	3	2	1

15. How would you rank the overall usefulness of the Chinese translations of the examples in doing the search? (Please circle the answer.)

Very useful	Useful	Of little use	Of no use
4	3	2	1

16. How would you rank the overall usefulness of the codes and abbreviations (e.g. U, sing) in doing the search? (Please circle the answer.)

Very useful	Useful	Of little use	Of no use
4	3	2	1

17. How would you rank the overall usefulness of the special features (e.g. bold) in doing the search? (Please circle the answer.)

Very useful	Useful	Of little use	Of no use
4	3	2	1

— End of Current Questionnaire —
— Thank you —

Deutsches Wortgut in rumänischen Pflanzennamen. Ein Bericht aus der rumänischen Wörterbuchpraxis

Radu Drăgulescu, *Department für romanische Studien,
Fakultät für Philologie und Theaterwissenschaften,
„Lucian Blaga“ Universität, Sibiu, Rumänien
(radu.dragulescu@ulbsibiu.ro)*

Zusammenfassung: Mit dem Vorführen vielfältiger Spracheinflüsse in volkstümlichen Pflanzennamen nimmt folgender Beitrag eine wichtige, jedoch ungenügend erforschte Erscheinung — Sprachschöpfung als Ergebnis der Pflanzenbezeichnungen und Ausdruck volkstümlicher Prägung — in den Blick. Ein Phänomen, das, wie von E. Coșeriu betrachtet wird, nicht genug hervorgehoben wird (da der einzelne Sprecher Schöpfer der Sprache/Poesie wurde, wann immer er eine Blume nannte). Unserer Meinung nach ist ein botanisches Lexikon einer bestimmten Region, einmal bekannt, nicht nur ein Thesaurus, aber auch ein Dokument der ansässigen Bevölkerung und Orte der Vergangenheit. Die botanische Volksterminologie hat in erster Linie einen praktischen Wert, der die Elemente des Pflanzenreiches innerhalb der gegebenen natürlichen Herrschaft kennzeichnet, unterscheidet und kategorisiert. Sie hat aber auch eine hohe theoretische Bedeutung für Linguisten, sowohl betreffend die Etymonen, zu denen sie zurückschicken, als auch die metaphorischen Bedeutungen, die die Phytonymen meist haben. Der Beitrag fokussiert auf einen ausgewählten Bereich der rumänischen Speziallexikografie und bietet einen Einblick in die jüngste lexikografische Erfassung rumänischer *Pflanzennamen*. Da jedoch das deutsche Wortgut in den herkömmlichen usuellen Pflanzenbezeichnungen von der rumänischen Fachliteratur eher marginal berücksichtigt und deren Besonderheiten nur vereinzelt erfasst und beschrieben worden sind, greifen die Ausführungen ausschließlich entlehnte und/oder dem Rumänischen angepasste Pflanzenbezeichnungen deutscher Herkunft auf. Andere außersprachliche Entsprechungen, Verwechslungen, irrtümliche Zuordnungen, sonstige Pflanzenarten oder Bildungen mit den Begriffen *german/nemțesc* („deutsch“) oder *săsesc* („sächsisch“) u.a. wurden hier nicht berücksichtigt. In unsere Arbeit soll dieses reiche Inventar an Pflanzennamen am Beispiel des hier vorkommenden deutschen Wortguts vorgestellt und erläutert werden.

Stichwörter: LEXIKOGRAFISCHE PRAXIS, BEDEUTUNGSERKLÄRUNG, SPEZIALWÖRTERBUCH, PFLANZENBEZEICHNUNGEN, ENTLEHNUNGSPROZESS, LEHNPRÄGUNG, LINGUISTISCHER EINFLUSS, DEUTSCH, RUMÄNISCH, SIEBENBÜRGERSÄCHSISCH

Abstract: **German Vocabulary in Romanian Plant Names. A Report of Romanian Lexicographic Practice.** This paper examines an important but insufficiently investigated phenomenon — lexical innovation as a result of plant naming and the expression of

national customs and traditions — by showing various linguistic influences in common plant names. E. Coşeriu considers this phenomenon to be given insufficient importance by linguists, given that fact that individual speakers became creators of language/poetry whenever they named a plant. I am of the opinion that a botanical lexicon of a particular region, once known, is not only a thesaurus, but can also be regarded as a document of a resident population or place in the past. First and foremost, botanical folk terminology has a practical value in that it identifies, distinguishes and categorizes members of the plant kingdom within a given natural realm. It, however, also has great theoretical significance for linguists, both with regard to the etymons from which they are derived, and to the metaphorical meanings that the phytonyms usually have. The article focuses on a selected area of Romanian specialized lexicography and offers an insight into the most recent lexicographic recording of Romanian plant names. Since, however, the German vocabulary in the commonly used plant names has been rather marginalized in the Romanian specialized literature, and since its characteristics have only been recorded and described loosely, the explanations are based solely on borrowed and/or adapted Romanian plant names of German origin. Other extraneous correspondences, mix-ups, erroneous assignments, other plant species or formations with the terms German/*nemţesc* ('deutsch') or Transylvanian-Saxon/*săsesc* ('sächsisch') etc., were not considered here. In my paper this rich inventory of plant names will be presented and exemplified using the German vocabulary listed here.

Keywords: LEXICOGRAPHIC PRACTICE, MEANING EXPLANATION, SPECIALIZED DICTIONARY, PHYTONYMS, LOAN PROCESS, CALQUE, LINGUISTIC INFLUENCE, GERMAN, ROMANIAN, TRANSYLVANIAN SAXON IDIOM

1. Vorbemerkungen

Folgende Ausführungen gehen von der Erkenntnis aus, dass mit der Erfassung des volkstümlichen Pflanzeninventars verschiedener Gebiete Rumäniens Sprachschätze und wertvolle Sprachzeugnisse vergangener Sprachgemeinschaften bewahrt werden. Das empirisch erhobene Sprachmaterial, eine Fundgrube insbesondere für Linguisten¹, verdeutlicht Eigenheiten, die der wissenschaftlichen Fachwelt und auch dem interessierten Nichtfachpublikum zugänglich gemacht werden müssen. Auch gibt es hier einen großen Bedarf an noch zu erbringenden Forschungen: z.B. Erforschung der Etyma und des metaphorischen Sprachgebrauchs, Erfassung der geografischen Verteilung usueller Pflanzennamen. Bei der Prägung der botanischen Terminologie waren Besonderheiten der Pflanzenwelt in der betreffenden Landschaft für die Sprecher-gemeinschaft ausschlaggebend.

Das hier vorgestellte Material geht auf meine Zusammenarbeit mit Constantin Drăgulescu anlässlich der Ausarbeitung eines erklärenden Wörterbuchs rumänischer Pflanzennamen *Dicţionarul explicativ al fitonimelor româneşti* (2010) zurück. Dieses Wörterbuch ist eine wertvolle Ergänzung des 1968 von Borza veröffentlichten Wörterbuchs der Pflanzennamen (*Dicţionar etnobotanic*). Es erfasst einige tausend Pflanzennamen ungarischer, sächsischer, deutscher, französischer, englischer, russischer, ukrainischer, serbischer, bulgarischer und

türkischer Herkunft und auch 10906 rumänische Bezeichnungen für insgesamt 2095 Pflanzenarten. Mit der Veröffentlichung dieses Wörterbuchs und eines weiteren, 2014 erschienenen Wörterbuchs (*Dicționar de fitonime românești*) dokumentiert Drăgulescu 21839 rumänische Pflanzennamen für 3227 heimische und exotische, auch gezüchtete Pflanzenarten und legt die bisher umfangreichste lexikografisch erfasste Belegsammlung vor. Darüber hinaus wurden weitere 3070 Pflanzengattungen und 612 Bezeichnungen für Pflanzenteile berücksichtigt. Somit liegt ein wertvolles Korpus rumänischer (autochthoner) Pflanzennamen vor, das über 26000 Stichwörter umfasst.

Fast die Hälfte der bisher aufgenommenen Pflanzenbezeichnungen ist auch im Wortschatz der rumänischen Gegenwartssprache auszumachen. Ihr Geltungsbereich lässt sich wie folgt umreißen: Etwa 7000 werden im alltäglichen Sprachgebrauch verwendet, ungefähr 1000 nur von Fachleuten (Botanisten, Agronomen, Fachleute für Wald- und Forstwirtschaft, Gartenbaufachleute, Apotheker oder Ärzte) und ca. 2200 sind landesweit üblich. Die überregionale Verbreitung dieses Inventars ist vorwiegend den Hirten und Kaufleuten zu verdanken. Eine wichtige Rolle spielten auch die Lehrbücher und populärwissenschaftlichen Abhandlungen zu Heilpflanzen oder zum Anbau von genießbaren Pflanzen oder Zierpflanzen.

2. Geschichtlicher² Abriss der Überlieferung rumänischer Phytonyme

Die erste Erwähnung eines Pflanzennamens geht auf das 14. Jahrhundert zurück und ist im Toponym *Cornățel* (1306) belegt. Weitere Belege sind das Anthroponym *Brad* (1348), später auch das Toponym *Gura Brădățelului* (1456) und das Phytonym *brad* (1490). Ebenfalls aus dem 14. Jahrhundert belegt sind die Toponyme *Răchita* (1360), *Jaleșul* (1385), *Sălcioara* (1391–1392), *Aniniș* (1392), *Ceret* oder *Nucet* (1400). Aus dem 15. Jahrhundert sind über 30 Siedlungsnamen (Oikonyme), Berg- und Gebirgsnamen (Oronyme) und Gewässernamen (Hydronyme) überliefert, die aus Pflanzennamen hervorgegangen sind (Drăgulescu 2010). Hinzu kommen weitere Anthroponyme wie *Laur* (1443), *Cimbru* (1462), *Săcară* (1479) bzw. *Secară* (1482, 1487), *Dzârnă* (1488), *Ceapă* (1492), *Brândușă* (1494) oder *Ghinda* (1497) (Drăgulescu 2010).

Aus dem 16. Jahrhundert stammen u.a. die Toponyme *Frăsinet de Dumbravă* (1502, 1504), *Curpinel* (1512), *Fântâna Teiul* (1517), *Mănăstirea Brusturi* (1518), *Măceșul* und *Pelinul* (1520), *Scoruș* (1519), *Călinet* (1548), *Jugastru* (1555), *Zmeuret* (1567), *Părul* (1571) und das Anthroponym *Mălai* (1507, 1517), die generische Pilzbezeichnung *bureți* (Pl. 1509), die Pflanzenbezeichnungen *chimen* („Kümmel“), *grâu* („Weizen“), *mărar* („Dill“), *măslin* („Olivenbaum“), *mujdar* („Knoblauch“), *pătrânjen* („Petersilie“), *smochin* („Feigenbaum“), *urzică* („Brennnessel“) (1551), *călin* („Schneeballstrauch“; 1557), *cer* („Zerreiche“; 1577), *arciariu* (*arțar*) („Ahorn“; 1579), *pepene* („Melone“; 1598).³ In dem aus dem 16. Jahrhundert stammenden Sprachdenkmal *Palia de la Orăștie* (1582), die erste Übersetzung ins Rumänische der ersten zwei Bücher des Alten Testaments, unter dem Einfluss der Reforma-

tion entstanden, sind die Pflanzenbezeichnungen *ceapă* („Zwiebel“), *coliandru* (sic!) („Koriander“), *dafin* („Lorbeer“), *mesteacăn* („Birke“), *stejar* („Eiche“) belegt.

In den bedeutendsten Sammlungen und Enzyklopädien des 17. Jahrhunderts sind fast 700 rumänische Pflanzenbezeichnungen überliefert worden.⁴ Hier muss auf das reichhaltige Werk *Dictionarium Valachico-Latinum* des Anonymus Caransebesiensis (ca. 1693-1700) verwiesen werden, das 240 Pflanzennamen vermerkt. Auch Corbeas *Wörterbuch* (ca. 1691–1697) umfasst ca. 500 rumänische Bezeichnungen und gehört damit zu den umfangreichsten Sammlungen.⁵ Vereinzelt erscheinen rumänische Phytonyme auch in den Schriften von Varlaam (z.B. *agud* „Maulbeere“; *horciță* „Senf“) und Dosoftei (z.B. *alac* „Dinkel“; *marulă* „Lattich“; *migdal* „Mandelbaum“) belegt.

Aus dem 18. Jahrhundert sind viele rumänische Bezeichnungen in Kochbüchern überliefert.⁶ Im *Calendariu pe 112 ani scos din multe feluri de cărți* („Der Kalender für 112 Jahre aus vielen Büchern hervorgegangen“), der 1785 in Iași erschien, werden für 120 Heilpflanzen rumänische Bezeichnungen verzeichnet, viele davon sind unter dem Einfluss russischer Pflanzenbezeichnungen entstanden. Das wichtigste Werk „lateinischer, ungarischer und rumänischer“ Phytonyme hat J. Benkő 1783 veröffentlicht. Hier werden 612 rumänische Bezeichnungen für 429 Pflanzengattungen erfasst, wobei vorwiegend Dorfbewohner aus Siebenbürgen und der Walachei als Gewährspersonen herangezogen wurden. Dieses Werk diente Sigerus (1791), Neustädter (1795) oder Veszelki (1798) als Vorbild, sodass Ende des 18. Jahrhunderts 1500 rumänische Pflanzenbezeichnungen aus fast allen Gebieten Rumäniens (Siebenbürgen, Banat, Walachei und Oltenien) für ca. 1000 Pflanzengattungen belegt sind.⁷

Ab dem 19. Jahrhundert kann eine eingehendere Beschäftigung und Sammeltätigkeit registriert werden. So sind über 3600 Bezeichnungen für 1520 Gattungen dokumentiert. Gh. Șincai (ca. 1810) übernimmt 188 rumänische Phytonyme für 167 Gattungen, die Sigerus verzeichnete und J.C.G. Baumgarten (1816) veröffentlicht in seinem Werk zur Siebenbürgischen Flora 296 Bezeichnungen für 257 Pflanzenarten. Die Mehrzahl der Bezeichnungen, die nicht von Benkő übernommen worden sind, sind Lehnbildungen, die auf fachliche oder lateinische Pflanzenbegriffe zurückgehen. J. Leonhard (1818) übernimmt ebenfalls einige rumänische Phytonyme von Sigerus und erfasst aber auch einige neue. Bobb (1822–1823) vermerkt in seinem Werk 635 Phytonyme, die entweder in einigen Vorgängerwerken bereits angeführt worden sind oder Lehnbildungen darstellen.

Das Lexikon *Lexiconul românesc, latinesc, unguresc, nemțesc*, 1825 in Budapest erschienen, führt außer den Phytonymen aus dem Werk von Corbea zahlreiche Lehnbildungen nach lateinischen, ungarischen oder deutschen Bezeichnungen an, davon auch einige, die als autochthon gelten und die im *Dicționarul Borza* bereits vermerkt wurden.⁸

In seinem Manuskript aus dem Jahr 1841 nimmt Szabó 180 Pflanzennamen aus der Moldau auf. Einige Jahre später ergänzt Fuss (1847) die von Sigerus und Baumgarten erfassten rumänischen Bezeichnungen. Diese Samm-

lung umfasst insgesamt 462 Phytonyme für 399 Pflanzenarten. Cipariu (1847) nimmt in seinem Glossar Bezeichnungen auf, die bereits im *Dicționar Corbea* oder im *Lexiconul de la Buda* und in den Arbeiten von Sigerus und Fuss belegt sind. Czihac und Suțu (1849) erfassen über 600 rumänische Phytonyme, viele davon einheimische, aus der Moldau stammende Bezeichnungen und solche, die auf Lehnbildung zurückgehen. Bariț (1858-1859) hat ausgehend von Vorgängerwerken ebenfalls eine Liste mit 976 Bezeichnungen für 733 Pflanzenarten zusammengestellt und Sava Popovici-Barcianu (1868) ergänzt dessen Liste mit selbst gesammelten Bezeichnungen, sodass die Anzahl bekannter rumänischer Pflanzennamen 1065 beträgt. Das Verzeichnis von Fătu (1870) umfasst wenige einheimische Bezeichnungen, während Baronzi (1872) rumänische und französische Pflanzenarten, darunter Bezeichnungen für Obstbäume und Weinreben, anführt. Einen wichtigen Beitrag zur Erfassung einheimischer rumänischer Pflanzennamen aus der Region Caraș-Severin leistet Manguica (1874). Er kritisiert dabei die Autoren, die Pflanzenverzeichnisse veröffentlicht haben und nicht selbst als Sammler tätig waren. Diese haben sich vorwiegend auf Übersetzungen aus dem Griechischen, Lateinischen oder Deutschen beschränkt. Brândză (1879–1883, 1882) hat die volkstümlichen Bezeichnungen mit mehreren Hunderten Lehnbezeichnungen ergänzt und rumänische Bezeichnungen durch Lehnübersetzung griechischer oder lateinischer Fachbegriffe geprägt. Auch Marian (1881–1908) und Hlibowski (1890) verzeichnen Hunderte von einheimischen Bezeichnungen aus der Bukowina. Porcius (1880–1882) ergänzt die Liste der Pflanzenbegriffe aus dem *Lexiconul de la Buda* mit denen von Bariț și Baumgarten. Er verzeichnet jedoch auch viele aus Năsăud gesammelte Bezeichnungen oder Eigenbildungen. Gleichfalls hat Crăiniceanu (1889, 1892–1893) autochthone Bezeichnungen und Lehnbegriffe aufgenommen, aber auch einige Namen, die auf Verwechslungen beruhen.⁹

Ende des vorigen Jahrhunderts sind bereits über 15000 rumänische Pflanzennamen für 2800 Arten belegt.¹⁰ Der Botanist Z.C. Panțu hat die Informationen aus dem *Lexiconul de la Buda* und die von Baumgarten, Fuss, Czihac/Suțu, Barițiu, Brândză, Porcius u.a. zusammengetragenen und mit selbst gesammelten Bezeichnungen aus dem dörflichen und urbanen Umfeld ergänzt. Daher galt sein 1929 veröffentlichtes Nachschlagewerk, das 4650 rumänische Pflanzenbezeichnungen für fast 2000 Pflanzenarten erfasst, bis zum Erscheinen des Wörterbuchs von Al. Borza (1968; *Dicționarul etnobotanic*) als das umfangreichste.

3. *Wandernde Namen. Autochthones und fremdes Wortgut in rumänischen Pflanzennamen*

In der rumänischen Ethnobotanik sind die auf Fachbegriffe oder auf fremde volkstümliche Bezeichnungen beruhenden (Lehn-)Bildungen unterrepräsentiert. Annähernd 80% rumänischer Pflanzennamen sind einheimische Bezeichnungen, die Mehrzahl davon sind Eigenbildungen der Dorfbewohner. Weitere

20% sind bildungssprachlich von Fachleuten (Botanisten, Apotheker, Ärzte, Philologen, Volkskundler, Kaufleute u.a.) geprägt worden. Diese sind z.B. im Deutschen, Französischen oder Englischen weitaus zahlreicher. Originelle rumänische Pflanzennamen belegen das kreative Potenzial und verdeutlichen Eigenheiten der benannten Pflanzen. Über 300 Bezeichnungen haben ihren ursprünglichen Geltungsbereich nicht verlassen, viele sind als Bildungen von Gelehrten oder Lehnbildungen sowie Okkasionalismen auszuweisen. Weitere 1500 Bezeichnungen haben Al. Borza und C. Drăgulescu als künstliche Bildungen, Übersetzungen oder Anpassungen von Fachbezeichnungen gewertet und in ihren Werken ausgeschlossen.¹¹ Das Wörterbuch von Drăgulescu ignoriert auch 301 rumänische Phytonyme, die irrtümlicherweise bestimmten Pflanzenarten zugeordnet wurden sowie weitere ca. 300 Bezeichnungen, die in den Vorgängerwerken beim Abschreiben schriftlich falsch erfasst worden sind. Die häufigsten falsch transkribierten Phytonyme verzeichnet die Arbeit von Pârvu, *Universul plantelor* (1991) (Drăgulescu 2010).

Mit der Einstellung der Vermarktung einiger Heilpflanzen und Gewürze oder der Pflanzen für die Herstellung von Tinkturen sind über hundert Phytonyme aus dem mündlichen Sprachgebrauch verschwunden. Allerdings haben über 500 Bezeichnungen — insbesondere für Zierpflanzen oder zum Verzehr — vorwiegend im 20. Jahrhundert Eingang ins Rumänische gefunden und gelten als weit verbreitet.

Die Wandlungen in der Namensgebung sind durch vielfältige Faktoren beeinflusst, da naturgegebene, geografische, klimatische oder pedologische Veränderungen auch Transformationen in der Zusammensetzung der Flora bedingen. Auch in der Ernährungsweise der Bevölkerung sind tiefgreifende Veränderungen auszumachen. Gleichfalls vermögen wirtschaftliche, soziale, historische, kulturelle und nicht zuletzt sprachliche Entwicklungen Veränderungen in der Prägung von Pflanzennamen herbeizuführen. Die über 100 Gemüse- und Obstarten, die nicht in Rumänien angebaut oder noch vor zwei Jahrhunderten nicht importiert wurden oder die neuen Pflanzen, die Gärten und Parks schmücken, erweckten die Phantasie vieler Bewunderer, die ihnen treffende Namen gaben.¹² Andererseits wurden bestimmte traditionelle Pflanzen nicht mehr angebaut oder verwertet wie z.B. Hirse, Dinkel, Buchweizen, bestimmte Apfel-, Birnen-, Pflaumen-, Weizen- oder Rebarten u.a.¹³

Außer den einheimischen Bezeichnungen verzeichnet das rumänische Inventar an Phytonymen auch Übernahmen aus den Sprachen der Nachbarländer oder der in Rumänien ansässigen Völkergruppen, vorwiegend Ungarn, Sachsen, Serben, Bulgaren, Ukrainer, Russen, Türken. Bildungssprachliche Bezeichnungen gehen vorwiegend auf die Gelehrten zurück.

Die fremden Wörter und Pflanzenbezeichnungen — auch innerhalb ein und derselben Sprache — wurden bedauerlicherweise unterschiedlich schriftlich erfasst. Vorwiegend ungarische und deutsche Autoren verzeichneten rumänische Pflanzenbezeichnungen in einer ihrer Sprache üblichen Form, manchmal genau — z.B. *znyeure, gyiotsěj, jitze de jie, maku csari, frunsze de tsints*

desitye, juáre, heráne für *znierură, ghiocci, iifă de iie, macu ciorii, frunză de cincii deştie, iuoară, hărană* oder *Brendusch, Kretzischor, Tschireschelle* für *brânduşi, creţisor, cireşele* — oder falsch (z.B. *szklesti mits* anstatt *scălci mici, ferukutze* für *fereguţă/feriguţă*). Mitunter können die verzeichneten Pflanzennamen verschiedenartig gelesen werden. So z.B. *flóre szóroluj* für *floarea soarelui* und auch *floarea sorului, stedse* für *şteje/ştegie*; *szásztyiu* erscheint als *sastiu* und *sachiu, rujen* wurde als *ruien* und *rujen* transkribiert, *selnike* wurde als *şálnică-şálnică* gedeutet. *Sálnică, szimszene, Szinsiánie* erscheinen in rumänischer Graphie mit *i*, anstatt *sâmzene, sânzienie*. Es sind auch Fälle auszumachen, wo die rumänische Schreibvariante nicht rekonstruiert werden konnte: z.B. *jijetz mere* für *Pulsatilla, virczid'sa* für *Polypodium vulgare*. Oft begegnen Inkongruenzen oder Auslassungen des bestimmten Artikels (des Genitiv-i). Z.B. *lilie vunet* (anstatt *lilie vânăta*), *jarbe kodruluj* (*iarba codrului*), *lapte kinyeluj* (*laptele cânelui*), *puine babe* (*pâinea babii*).

Manche rumänische Pflanzenbezeichnungen lassen sich indirekt auf einen indoeuropäischen Ursprung zurückführen. Viele entstammen dem Sprachgebrauch der in Rumänien lebenden Minderheiten (Ungarn, Sachsen, Roma) oder lassen sich als Ergebnis von Rückentlehnungsprozessen ausweisen. Etliche rumänische Bezeichnungen unterschiedlicher Pflanzen sind untergegangen und leben *nur noch* im Sprachgebrauch der Ungarn, Sachsen oder Roma fort. Manche Übernahmen aus dem Wortschatz der Minderheiten auf dem Gebiet Rumäniens zeigen Veränderungen in der Bedeutung, Lautung und Schreibweise.

Bei der Herleitung der Herkunft von Pflanzennamen haben die Lexikografen vergleichbare Bezeichnungen aus dem Sprachgebrauch der Minderheiten herangezogen und als direkte Übernahmen ausgewiesen, wobei oft unberücksichtigt blieb, dass der betreffende Name seinerseits von diesen Sprachen aus dem Rumänischen entlehnt wurde. Auch Pflanzenbezeichnungen aus anderen Sprachen sind berücksichtigt worden.

Die im Wörterbuch kodifizierten Herkunftsangaben verweisen vielfach auch auf außersprachliche Faktoren. Sie bieten Hinweise zu Formen und Farben einiger Pflanzenteile, zu ihrem Umfeld, zu Vorgängen und Lebenszyklen, zu ihrem Gebrauch und der mit ihnen verbundenen Bräuche, Legenden, rituellen Handlungen, mehr noch, die Pflanzenbenennungen sind von Fachbezeichnungen begleitet, da diese oft eine wichtige Rolle spielen und bei der Unterscheidung der Pflanzenarten mit gleichen Phytonymen hilfreich sind.

Die in den rumänischen Pflanzennamen belegten Präpositionen, Konjunktionen, Adverbien und Pronomina sind (vor)lateinischen Ursprungs: Präp. *de* (,von') < lat. *de*; *la* (,nach'; ,von') vgl. lat. *illac* oder aus der Präp. *a* „la“ (,zu') aus der lat. *ad*, i.-e. Wurzel **ad*; *cu* (,mit') < lat. *cum*, i.-e. Wurzel **ku*; *sub* (,unter') < lat. *sub*, *subtus*; *în* (,in') < lat. *in*; *din* aus der Zusammensetzung *de+in*; *pe* (,auf') < lat. *per*, i.-e. Wurzel **per*; Konj. vgl. lat. *sic* oder vorlateinisch; Adv. *ca* (,wie') < lat. *quam*, altind. *ca*; Pron. *care* (,welche, welcher, welches') vgl. lat. *qualis*; *se* (,sich') < lat. *se*, i.-e. Wurzel **se*).

Im Folgenden soll dieses reiche Inventar an Pflanzennamen exemplarisch am Beispiel des hier vorkommenden deutschen Wortguts vorgestellt und erläutert werden. Es handelt sich hierbei um entlehnte und/oder dem Rumänischen angepasste Pflanzenbezeichnungen deutscher Herkunft. Andere außersprachliche Entsprechungen, Verwechslungen, irrtümliche Zuordnungen, sonstige Pflanzenarten oder Bildungen mit den Begriffen *german/nemțesc* („deutsch“) oder *săsesc* („sächsisch“) u.a. wurden hier nicht berücksichtigt.

Das reiche Lehngut lässt sich wie folgt gliedern:

- (1) **Direkte phonetisch unveränderte Übernahmen:** *acație* (*Robinia pseud-acacia*) < dt. Akazie; *bambus* (*Phyllostachys viridiglauca*, *Sasa palmata*) < dt. Bambus (franz. bambou < malayez. bambu); *carfiol* (*Brassica oleracea* var. *botrys*) < dt. Karfiol; *chifăr*, *chifer* (*Pinus silvestris*) < dt. Kiefer; *Diptam* (*Dic-tamnus albus*) < dt. Diptam; *dost* (*Origanum vulgare*) < sächs./dt. Dost; *edăl-vais* (*Leontopodium alpinum*) < dt. Edelweiss; *gerbera* (*Gerbera hybrida*) < dt. Gerbera (Fachbegriff/zur Ehrung des deutschen Botanikers Traugott Gerber); *inghe* (*Callistephus chinensis*) < dt./sächs. Inge (Antoponym); *lilie* (*Iris florentina*, *Iris germanica*, *Lilium bulbiferum*, *Lilium candidum*, *Lilium martagon*) < dt. Lilie; *mangold* (*Beta vulgaris* var. *cicla*) < dt. Mangold; *pizang* (*Musa* spp.) < dt. Pisang „Bananenpflanze“; *ramie* (*Boehmeria nivea*) < dt. Ramie; *vasărlilie* (*Hosta plantaginea*) < dt. reg. Wasserlilie; *vermut* (*Artemisia absinthium*) Wermut; *zegras* (*Zostera marina*) < dt. Seegras.
- (2) **Phonetisch veränderte Übernahmen** aus der deutschen Sprache: *acacie* (*Robinia pseudacacia*) < dt. Akazie; *bacolți* (Jungbirke der *Betula pendula*) < dt. Bachholz; *bonă* (*Vicia faba*) < dt. Bohne; *borda miresii* (*Campanula fenestrella*) < bartă, bortă, beartă, bertă („Hauptband, Hauptschmuck, Blumenkranz“) < dt. Barten, Berthe; *Brinolă* (*Mentha pulegium*) < dt. Brunelle; *bufă* (*Quercus* spp.); *bufoi*, *bufon(iu)* (*Fagus silvatica*) < dt. Buche (*Fagus*); Bedeutungserweiterung; gegenwärtig für ein beliebiges Gebüsch oder als Bezeichnung für einen jungen Wald (im Banat); *buhaci*, *buhaș* („Jungexemplare“ von *Abies alba*, *Fagus silvatica*, *Picea abies*), *buhășer* (*Picea abies*; „Jungexemplar“) < dt. Buche (*Fagus*); *cartfiol*, *cartofiol*, *cartofiom* (*Brassica oleracea* var. *botrys*) < dt. Karfiol; *cartoafe* (*Dahlia cultorum*, *Solanum tuberosum*) < dt. Kartoffel; *castan* (*Aesculus hippocastanum*, *Castanea sativa*) < dt. Kastanienbaum, wissenschaftl. Benennung *castanea*; *cherbel*, *cherbăl* (*Chaerophyllum aureum*); *chervăl* (*Anthriscus cerefolium*) < dt. Kälber(kropf), Kerbel bzw. Wald-/Wiesen-Kerbel < *cerefolium* < griech. *chairophyllon* „angenehmes Blatt“; *cocarde* (*Gaillardia picta*) < dt. Kokardenblume; *crețăruș(i)* (*Bellis perennis*) < cre(i)țar, crăițar < dt. Kreuzer („Münze“), ung. *krajcár*; *crumciri*, *crumpe*, *crumpei*, *crumpei de lună* (*Solanum tuberosum*), *crumpene* (*Helianthus tuberosus*, *Solanum tuberosum*), *crumpeni*, *crumperi*, *crumpi*, *crumpiri*, *crumple*, *crumpli*, *crupre* (*Solanum tuberosum*) < dt. Grundbirne, dt. dial. *krompire*, *krumbiir*, *krumpa*; *crumpei* (*Elaphomyces* spp.) ein knolliger Pilz, der so benannt wurde, weil er wie eine Grundbirne aussieht; *cule* (*Solanum*

tuberosum) Pflanzenname aus Kreis Hunedoara, eine Nachbildung nach dt. Knolle, Knulle, Kulen; *cuțcrișoară, cuțcrișor* (*Pulmonaria officinalis*) < dt. Kuhkreuz; *ențură, enzură* (*Gentiana* spp.) < *ințură, ghințură* unter dem Einfluss des dt. Enzian; *eră* (*Triticum aestivum*) Name des Weizen, mit dem dt. Ähre und griech. er, ear „Frühling“ zu vergleichen; *erdăpane* (*Solanum tuberosum*) < dt. Erdapfel; *feărminț* (*Mentha piperita*) < dt. Pfefferminze; *fenicel* (*Foeniculum vulgare*) wahrscheinlich unter dem Einfluss des sächs. *fienchəl*, dt. Fenchel; *fetică* (*Ficaria verna* heute *Ranunculus ficaria, Valerianella locusta*)¹⁴; *fiache* (*Viola odorata*) vgl. dt. Veilchen, russ., slow. *fialka*; *fidireie* (*Solanum tuberosum*) zu vergleichen mit dem dt. Feldeier, Erdeier; *finc(h)en, finchăn, fenhiel, finhien, finchil, finc(h)in, fincăl* (*Foeniculum vulgare*) < sächs. *fin(t)chəl, fienchəl*, dt. Fenchel; *flos* (*Linum usitatissimum*) < dt. Flachs; *garva, garvă* (*Achillea millefolium*) < dt. Schafgarbe, Garbe, sächs. *guar, guərblâm*, altdt. *garwe, garwa*; *ghințuire* (*Gentiana punctata*), *ghințură* (*Centaurium erythraea, Euphrasia stricta, Gentiana* spp.), *ghințurea* (*Centaurium erythraea, Gentiana cruciata*), *ghințurele* (*Gentiana clusii*), *ghințurică* (*Gentiana frigida, Gentiana lutea, Gentiana nivalis*) wahrscheinlich aus *jințură/ințură/ențură* (vgl. dt. *Gentian* und *Enzian* und die wissenschaftliche Bezeichnung *Gentiana* nach dem Namen des illyrischen Königs *Gentius*); *gonele* (*Solanum tuberosum*) aus dem dt. *Knolle*, oder **gogonele* („unreife Tomaten“); *grap* (*Rubia tinctorum*) < dt. Krapp; *hăidănuș(ă)* (*Viburnum opulus*); besonders die Früchte < dt. Heidenkirsch(en); *hulcă* (*Triticum aestivum* var. *muticum*) möglich vom dt. *Hülle, Hülse*; *ințură* (*Gentiana* spp.), *ințurea* (*Gentiana* spp.), *ințurică* (*Gentiana lutea*) < rum. *ghințură*; *levcoaie* (*Cheiranthus cheiri*) < dt. *Levkoje*; *libistoc* (*Levisticum officinale*) < dt. *Liebstock*; *liză* (*Pelargonium zonale*) < dt. *Liese* (Anthroponym); *maieran, maioran, maiorean, măierean(ă), maiorană, măgheran* (*Majorana hortensis, Origanum vulgare*), *măgherean* (*Majorana hortensis*), *măghieran, măghiran, măghieran, măgieran, măgiran* (*Majorana hortensis*), *măioreană* (*Majorana hortensis*): < dt./sächs. *Majoran, Meieran*; *minuțe* (*Bellis perennis*) < dt. *Münze* (Gänseblümchen); *mutmel, mutruel, muturel* (*Triticum dicoccum*): vgl. dt. dial. *Mudelweissa* (*Triticum aestivum* var. *muticum*) (Zweikorn) bzw. rum. *mudmel, mutmel* („feines Weizenmehl“) < dt. *Mudmehl*; *nudli* (*Portulaca grandiflora*) < dt. *Nudel*¹⁵; *ortanță, ortensă, ortensie, ortenzie* (*Hydrangea opuloides*) < wissenschaftl. *Hortensia opuloides*, dt. *Hortensie*, fr. *hortensia*; *petersil* (*Petroselinum hortense*) < dt. *Petersil(ie)*; *pilzene* (*Boletus edulis*) < dt. *Pilz, Pilze*; *pomaranță, pomeranță* (*Citrus aurantiaca*) < dt. *Pomeranze* (> russ. *pomeranetj*); *pospănc, pospang, puspan, pușpan, sospa(i)n* (*Buxus sempervirens*) < dt. *Buchsbaum* > ung. *puspáng, puszpáng*; *rapăn* (*Capsella bursa-pastoris*) < dt. *Rappe, Rappen* („Münze“), da die Früchte so aussehen; *răpac* (*Brassica napus, Brassica rapa*), *răpă, rāpe* (*Beta vulgaris, Solanum tuberosum*), *rips* (*Brassica napus*) < dt./sächs. *Rübe, R(i)ep* (*Beta vulgaris*) vgl. dt. *Raps, Reps(kohl)*, ukr. *rāpa* (*Brassica rapa*), slaw. *rēpa* „Speiserübe“; *rebizlă, ribiză, ribizil, ribizle, ribizli, ribizlă, ribuțe, rābiză, ribizle* (*Ribes rubrum*) < dt. *Ribisel*, landl. *Ribiseln* (< *Ribes* < arab. *ribās*), davon

auch die ung. *ribizli*, *ribizke*, slovak. *ribezľa*; *rosmailin* (*Rosmarinus officinalis*), *rosmalin* (*Artemisia abrotanum*, *Rosmarinus officinalis*), *rosmarin* (*Artemisia abrotanum*, *Artemisia annua*, *Rosmarinus officinalis*), *rosmarint*, *rosmălin*, *rosmărin*, *rosmolin* (*Rosmarinus officinalis*), *rozmalin* (*Artemisia abrotanum*, *Rosmarinus officinalis*), *rozmarin* (*Artemisia abrotanum*, *Asparagus officinalis*, *Asparagus sprengeri*, *Daphne cneorum*, *Rosmarinus officinalis*), *rozmolin* (*Rosmarinus officinalis*) aus dt. Rosmarin, it. *rosmarino*, davon auch ung. *rosmarin(t)*, russ. *rozmarin*, möglicherweise auch aus der wissenschaftlichen Benennung *Rosmarinus* < griech. *rhops* ‚kleiner Strauch‘; griech. *myrinos* ‚riechbar, geruchsvoll, balsamisch‘ oder lat. *ros* ‚Tau‘ und lat. *marinus* ‚des Meeres‘; *ruben* (*Beta vulgaris*) < dt. Rübe, Pl. Rüben; *silarei* (*Apium graveolens*) < dt. Sellerie; *simsă* (*Juncus effusus*, *Juncus inflexus*) < dt. Simse; *soloagăr* (*Helianthus annuus*) wahrscheinlich aus dt. Sonn(e)augen, auf die Blüten hinweisend; *spichinat*, *spichinel* (*Lavandula angustifolia* și *spp.*) dt. Spickenard, Spicke; *Lavandula spica* < lat. *spicum* ‚Ähre‘, auf die Blüten hinweisend; *spinac* (*Spinacia oleracea*) < Kontamination zwischen rum. *spanac* und dt. *Spinat*; *spinat* (*Spinacia oleracea*, *Atriplex hortensis*) < dt. Spinat, sächs. *špinât* (*Spinacia oleracea*); *strelia* (*Strelitzia sp.*) < dt. Strelitzie bzw. wissenschaftliche Benennung¹⁶; *șalotă* (*Allium ascalonicum*) < dt. Schalotte(n), fr. *echalote*, engl. *Shallot*; *șlaierblume* (*Gypsophila paniculata*) < dt. Schleierblume > skr. *ślajer* (trava); *șpic* (*Lavandula angustifolia*) < dt. Spicke < *Lavandula spica*; *ștainglezăr* (*Melilotus albus*, *Melilotus officinalis*) vgl. dt. Steinklee; *titenchi*, *titenghi* (*Coprinus atramentarius*) das rum. Phytonym könnte eine Verzerrung der deutschen Bezeichnung sein oder vom i.-e. Wurzel **teng-* ‚(weich) machen, feuchtmachen‘ abstammen; *toibă* (*Bromus secalinus*) < dt. Taubhafer; *triftărele* (*Clitocybe infundibuliformis*, *Craterellus cornucopioides*, *Petunia hybrida*): die ersten zwei Arten bezeichnen trichterförmige Pilze (< dt. Trichter), die letzte ist eine Zierpflanze mit einer trichterförmigen Blumenkrone; *trimpăn*, *trinpăn* (*Sedum acre*) scheint eine Kontamination zwischen der deutschen Benennung *Tripmadam* und die rum. *trânjen*, *trânjin* zu sein; *țâncraus* (*Equisetum arvense*) < dt. Zinnkraut; *țâtroană* (*Citrus limon*, *Philadelphus coronarius*), *țâtron* (*Hedera helix*, *Philadelphus coronarius*, *Ruscus hypoglossum*, *Viscum album*), *țitroană*, *țitrom* (*Citrus limon*), *țitron* (*Citrus limon*, *Philadelphus coronarius*) < dt. Zitrone, Zitron und ung. *citrom*, *citron* (*Citrus limon*) < dt. Zitrone, Zitron (*Citrus limon*), im Süden Siebenbürgens bezeichnet *zitron* eine immergrüne Pflanze, die Blumen der *Philadelphus coronarius* haben einen zitronenähnlichen Geruch; *țicorie* (*Cichorium intybus*) < dt. Zichorie > ung. *cikoria*; *țimat*, *țimărt*, *țimăt*, *țimet*, *țămărt* (*Cinnamomum zeylanicum*) < dt. Zimmet; *țucroș*, *țucuran*, *țucurechi* (*Beta vulgaris* var. *altissima*) < dt. Zuckerrübe, sächs. *zackerriep*, ung. *cukorrépa*, *cukrosrépa*.

- (3) **Übernahmen aus dem siebenbürgisch-sächsischen Dialekt:** *apricops*, *apricos*, *apricoz*, *apricoză*, *abricoji*, *abricoz* (*Armeniaca vulgaris*; heute: *Prunus armeniaca*) sächs. *Aprikôs* < dt. Aprikose; *banat*, *bănat* vgl. sächs. *banat*,

bənat, bonât, banötek, banədick, bənâti(k), dt. Bonat, Banat, Benat, Bonot < dt. Benedikt oder i.-e. *man- (> *ban-) („gut“), die Pflanze riecht sehr angenehm; *berten*, *bertin* (*Artemisia dracunculus*) < sächs. biertrem, dt. Bertram; *bonă* (*Vicia faba*) aus rum. *boană*, *bob*, *boabă* < sächs. bon, bun, dt. Bohne; *boreț* (*Borago officinalis*) < sächs. boretsch, serb. porec, poreč, skr. borač, dt. Boretsch, Burets, ung. burecs, aus dem lat. *boletis, -em (anstatt *boletus*, nach dem griech. βολιτης); *calarabă*, *calarambă*, *călărabe*, *caralabă*, *cararabe*, *cărălăbă* (*Brassica oleracea* var. *gongyloides*), *chelărabe* (*Brassica oleracea* var. *acephala*), *chelarabă*, *chelărabe* *vinete*, *chelărabi*, *chelărabie* (*Brassica oleracea* var. *gongyloides*), *chelărade* (*Brassica oleracea* var. *acephala*, *Brassica oleracea* var. *gongyloides*), *chelerab* (*Brassica oleracea* var. *gongyloides*) < sächs. kalərabən, kalərăbi, dt. Kohlrabi (> ung. kalaráb, karalábé, skr. keleraba, poln. kalarepa u.a.); *cârneauă*, *hernave* (*Hypericum perforatum*): vgl. sächs. kirnjān, harnāo, harnā, harnau und dt. dial. Karnoul, Hartenau, Hertenau; *chel*, *chil*, *chiel*, *chil* (*Brassica oleracea* var. *sabauda*) sächs. ki(e)l, kīl, dt. Köhl-(kraut) > skr. kelj, ung. kelkáposzta; *cherbel*, *cherbăl* (*Chaerophyllum aureum*), *chervăl* (*Anthriscus cerefolium*) < dt. Kälber(kropf), dt. Kerbel, sächs. kiərwəl, kerwəl, Wald-/Wiesen-Kerbel, < *cerfolium* < griech. chairophyllon („angenehmes Blatt“), davon auch fr. cerfeuill; *clocăboambă* (*Aquilegia vulgaris*) < sächs. kliokəbläom, klokəblommən, dt. Glocke(n)blume; *cricin*, *crihim*, *crihin*, *crichin* (*Prunus insititia*) < sächs. kriehe(n), krächən; *crumciri*, *crumpe*, *crumpei*, *crumpene* (*Helianthus tuberosus*, *Solanum tuberosum*), *crumpeni*, *crumperi*, *crumpi*, *crumpiri*, *crumple*, *crumpli*, *crupre* (*Solanum tuberosum*) vgl. skr. krumpir, ung. krumpli < sächs. krumpirn, dt. Grundbirne, dt. reg. *krompire*, *krumbiir*, *krumpa* beeinflusst auch von ung. *kolompér*, *kolompéra*, *krompé*, *krumpi*, *krumpli* < dt. Grundbirne, an manchen Orten bedeuten diese Namen „krumme Knollen“ (in Anlehnung an dt. krumm; Krummbirne); *dost* (*Origanum vulgare*) vgl. sächs. Dost; *fenicel* (*Foeniculum vulgare*) aus der wissenschaftlichen Benennung, lat. foeniculum, wahrscheinlich vom sächs. fienchəl, dt. Fenchel beeinflusst; *finc(h)en*, *finchăn*, *fenhiel*, *finhien*, *finchil*, *finc(h)in*, *fincăl* (*Foeniculum vulgare*) < sächs. fin(t)chəl, fienchəl; *foalcer* (*Mathiola incana*), *foalchine*, *foalâne*, *foaline* (*Mathiola incana*), *foaltăr*, *foalter* (*Mathiola incana*), *foaltine* (*Cheiranthus cheiri*, *Mathiola incana*), *foahine* (*Tropaeolum majus*), *foanchine* (*Tagetes erecta*) < sächs. foaltchər, vältchər, foalen < dt. Veilchen (*Mathiola incana*), sächs. giel foalen, dt. gelbes Veilchen (*Cheiranthus cheiri*), sächs. voälchen, dt. Veilchen (*Viola* spp.); *garva*, *garvă* (*Achillea millefolium*) < sächs. guar, guərblām < dt. Schafgarbe, Garbe, altdt. garwe, garwa („Garbe, Bündel“); *grubă*, *grube*, *grumbă(i)*, *grumbănă*, *grumbe*, *grumcizi*, *grumpă*, *grumpe*, *grumpene*, *grumpeni*, *grumpi*, *grumpin(i)* (*Solanum tuberosum*) < sächs. grumpir(τə)n, grumpir, grompír, dt. Grundbirne, Grunbir, Gromper; *heber* (*Lagenaria siceraria*) < sächs. Heber/hebər; *heciunpeci*, *heciunpeci* (die Früchte und die Konfitüre von *Rosa canina*) < sächs. hätschenpätsch, hetschumpetsch; *holăr* (*Sambucus nigra*) < sächs./dt. Holler; *inghe*, *inguțe* (*Callistephus chinensis*) < sächs./dt. Inge (Anthroponym); *liurbă(u)*, *l(i)urben*, *l(i)urber*, *liorbe*, (*Laurus nobilis*) <

sächs. lührber(buhm), lu(i)rbən, lu(i)rbər, liurbər, dt. Lorbeer; *liurben*, *lurbău* (*Syringa vulgaris*) < sächs. lorbeer, lorbeerbaum (*Syringa vulgaris*); *liză* (*Pelargonium zonale*) < germ/sächs. Liese, *liurbă(u)*, *l(i)urben*, *l(i)urber*, *liorbe* (*Laurus nobilis*) < sächs. lührber(buhm), lu(i)rbən, lu(i)rbər, liurbər, dt. Lorbeer; *liurben*, *lurbău* (*Syringa vulgaris*) < sächs. lorbeer, lorbeerbaum (*Syringa vulgaris*); *mure* (*Daucus carota*) < sächs. murr, murrən, dt. Möhre, slaw. murky; *părădaice*, *părădaisăne*, *părădaise*, *părădăi*, *părădaiși*, *părădazine*, *părădăici*, *parad(a)ise*, *paradaice*, *paradaise*, *paradice*, *paradici*, *bărădăi*, *porobici*, *porodăi*, *porodăici*, *porodeici*, *porodică*, *porodici*, *porodiciuri*, *porodiță* (*Lycopersicon esculentum*) < sächs. Parədeis, dt. Paradeis(apfel), Paradies(apfel), ung. paradicsom < dt. Paradeisen, unter den rumänischen Pflanzenbenennungen gibt es Mischbildungen zwischen sächs. parədeis und ung. paradicsom, manchmal auch serb. paradajz, mit der Grundform vom griech. parádeisos (,umschlossenes Gebiet reserviert für den König Persiens') abstammend¹⁷; *rabarbară* (*Rheum officinale*), *rabarbură* (*Rheum officinale*, *Rheum palmatum*), *rabarbăr* (*Rheum rhabarbarum*, *Rheum rhaponticum*), *rebarbar(ă)* (*Rheum officinale*, *Rheum palmatum*, *Rheum rhaponticum*), *rebarboră*, *robarbă*, *robarbăr* (*Rheum palmatum*) < sächs. rabarbər, dt. R(h)abarber, it. rabarbaro < rhabarbarum, radix barbarum „Barbarenwurzel“; *spargă* (*Asparagus officinalis*) < sächs. spargəl, dt. Spargel > ung. spárğa; *șarlai* (*Salvia sclarea*), *șerlai* (*Salvia aethiopis*, *Salvia sclarea*), *șerlau* (*Salvia aethiopis*) < sächs. schârlôch, dt. Scharlachkraut (*Salvia sclarea*); *șnirling* (*Allium schoenoprasum*) < sächs. schnirling, schnittling, dt. Schnittlauch; *Taghete* (*Tagetes spp.*) < sächs./dt. Tagetes bzw. aus der wissenschaftlichen Benennung dieser Zierpflanze nach dem Namen des Halbgottes Tages; *Țelăr*, *țeler*, *țeleră* (*Apium graveolens*) < sächs. zellər, dt. Zeller; *țucroș*, *țucuran*, *țucurechi* (*Beta vulgaris* var. *altissima*), Zuckerrübe, sächs. zackerriep, ung. cukorrépa, cukrosrépa.

- (4) **Andere Wortformen fremden Ursprungs:** *abricoji*, *abricoz* (,Aprikose'; die Formen *abricoji*, *abricoz*, durch die Stimmhaftigkeit der Konsonant p erhalten, sind mit fr. *abricot*, *abricot-pêche* und *abricotier* „Aprikosenbaum“, span. *albaricoque*, port. *albricoque* zu vergleichen, alle vom arab. *al barquq* (al *birquq*) ,frühreif', vgl. lat. *praecoquum malum* ,frühreifer Apfelbaum' und lat. *apricus* ,sonnenbescheint; reif'; *carfiol*, *cartifiol*, *cartofiol*, *cartofiom* (*Brassica oleracea* var. *botrys*) < dt. Karfiol < ital. dial. *cavol fior*, sächs. kardiviol(ə)n, kardiwiol > ung. kárfiol, kartifiol, skr. karfijol, ukr. karafiol; *cartoafe* (*Dahlia cultorum*, *Solanum tuberosum*) die erste Pflanze hat in der Erde knollige Wurzeln, die zweite Knollen < dt. Kartoffel < it. *tartufoli*, *tartiffolo* „Trüffeln“, weil die Knollen diesen unterirdischen Pilzen ähnlich aussehen; *castan* (*Aesculus hyppocastanum*, *Catanea sativa*) < dt. Kastanienbaum, paleoslav. *kastanъ*, bulg. *kástan*, *kesten*, wissenschaftliche Benennung *castanea* > ngriech. *kástanon* ,Kastanie' in Verbindung mit altind. *kashta-* ,Wald'; *cocarde*, *coarde* (*Gaillardia picta*) < dt. Kokardenblume < fr. *cocarde*; *crețăruiș(i)* (*Bellis perennis*) < cre(i)țar, *crăițar* ,Kreuzer'

(Münze) < dt. Kreuzer, ung. krajcár; *fenicel* (*Foeniculum vulgare*); *fiache* (*Viola odorata*); *lilie* (*Iris florentina*, *Iris germanica*, *Lilium bulbiferum*, *Lilium candidum*, *Lilium martagon*), *lilioane* (*Lilium candidum*), *liliom* (*Hemerocallis fulva*, *Lilium candidum*), *lilion* (*Hemerocallis fulva*, *Iris germanica*, *Lilium candidum*), *liliu* (*Lilium candidum*), *liliuă* (*Iris germanica*), *liliuță* (*Anthericum ramosum*), *lilom* (*Iris spp.*) < dt. Lilie, sächs. Leljen < lat. *lilium* ‚Lilie‘, verbreitet auch unter dem Beinamen *liliom*, *lilion*, unter ungarischem Einfluss der Benennung *liliom*, alle Pflanzen sind Lilienarten oder damit verwandt; *Maieran*, *maioran*, *maiolean*, *măierean(ă)*, *maiorană*, *măgheran* (*Majorana hortensis*, *Origanum vulgare*), *măgherean* (*Majorana hortensis*), *măghieran*, *măghiran*, *măghieran*, *măgieran*, *măgiran* (*Majorana hortensis*), *măioleană* (*Majorana hortensis*) < dt./sächs. Majoran, Meieran, poln. und skr. majeran, russ. maioran, skr. majoran, ung. Majorán, alle < wissenschaftliche Benennung *majorana*, arabischer Herkunft, oder aus lat. *majorana*, welches auf lat. *major* ‚größer, kräftiger‘ zurückgeht oder von Dioscorides erwähnten altlat. *meyurana* bzw. weil es sich um eine bevorzugte aromatische Pflanze handelt, mit dem mgriech. *mágeiros*, ngriech. *mageirion* ‚Küche‘ zu vergleichen; *Pizang* (*Musa spp.*) < dt. Pisang < *pisang* ist der Name der Banane bei den Malaysiern und Indonesiern; *porobici*, *porodăi*, *porodăici*, *porodeici*, *porodică*, *porodici*, *porodiciuri*, *porodiță* (*Lycopersicon esculentum*) < ung. *paradicsom* < dt. Paradeisen; *rabarbară* (*Rheum officinale*), *rabarbură* (*Rheum officinale*, *Rheum palmatum*), *rabarbăr* (*Rheum rhabarbarum*, *Rheum rhaponticum*), *rebarbar(ă)* (*Rheum officinale*, *Rheum palmatum*, *Rheum rhaponticum*), *rebarboră*, *robarbă*, *robarbăr* (*Rheum palmatum*) < sächs. *rabarbær*, dt. R(h)abarber, it. *rabarbaro* < *rhabarbarum*, *radix barbarum* ‚Barbarenwurzel‘; *ramie* (*Boehmeria nivea*) < dt. Ramie, fr., engl. *ramie* < malaysisch *rameh*; *strălucită* (*Gypsophila paniculata*) aus dem rum. Verb *a străluci* (Präfix *stră-* und „a luci“ < lat. *lucere*) ‚(aus)strahlen‘, oder Kombination zwischen i.-e. *stra-*, vgl. dt. Strahl und das rum. Verb *a luci*, die Pflanze hat viele, reiche, weiße oder rötliche Blüten.

- (5) **Lehnprägungen** sind Benennungen, die unter Verwendung von sprachlichen Mitteln der Empfängersprache zustandekommen sind. Hierbei handelt es sich um eine Nachbildung eines fremdsprachlichen Inhalts mit den Mitteln der Empfängersprache (hier: des Rumänischen). Dabei kann die Adaption der Übernahmen auf verschiedene Weise erfolgen:

(a) **Lehnbedeutung**: Die Anpassung geschieht durch Wandel bzw. Erweiterung der Bedeutung heimischer Wörter, d.h. es geht um eine Bedeutung, die ein Wort unter fremdsprachlichem Einfluss annimmt, wodurch eine Umdeutung der ursprünglichen Bedeutung bzw. eine Bedeutungserweiterung stattfindet. Bei einer Lehnbedeutung wird nur die Bedeutung und nicht das Wort übernommen und auf ein einheimisches Wort übertragen. Lehnbedeutungen entstehen oft durch Interferenzen bei Sprachträgern, durch den hohen Verwandtschaftsgrad zwischen den Sprachen. Auffal-

lend ist hierbei, dass Bedeutungserweiterungen von Lexemen einer Sprache nach dem Vorbild einer anderen Sprache nicht nur auf der Klangähnlichkeit beruhen, sondern auch die inhaltsbezogene Teilidentität zwischen Lexemen der beiden Sprachen berücksichtigen: *coif* (*Aconitum* spp.) < dt. Eisenhut, rum. *coif* ‚Helm‘ < lat. *cofea*; weist auf die Blütenform hin; *focul drăguței* (*Lychnis calcedonica*) < dt. brennende Liebe; rum. *foc* ‚Feuer‘ < lat. *focus*, rum. *drăguței* ist die Genitivform von rum. *drăguță* ‚Geliebte‘ < drag < slaw. dragŭ; die Pflanze hat rote Blüten.

(b) **Lehnübersetzung:** Bezeichnet das Ergebnis einer genauen Glied-für-Glied wiedergebenden Übersetzung eines fremdsprachlichen Vorbildes. Dabei wird auch den identischen/ähnlichen und semantischen äquivalenten Wortbildungsmustern Rechnung getragen: *arbore/copac de ambră* (*Liquidambar orientalis*) < dt. Amberbaum; *arborele de bumbac* (*Bombax ceiba*) < dt. Baumwollbaum; *arbore de chinină* (*Cinchona* spp.) Chininbaum; *arbore de gumă* (*Ficus elastica*) < dt. Gummibaum; *arbore de mastic* (*Pistacia lentiscus*) < dt. Mastixstrauch; *arborele de pâine al maimuțelor* (*Adansonia digitata*) < dt. Affenbrotbaum; *arborele mamut* (*Sequoia gigantea*; heute: *Sequoiadendron giganteum*) < dt. Mammutbaum; *aripa îngerului* (*Caladium bicolor*) < dt. Engelsflügel; *castan galben* (*Aesculus octandra*, heute *Aesculus flava*) < dt. gelber Kastanienbaum, *castan indian* (*Aesculus indica*) < dt. indischer Kastanienbaum oder von der wissenschaftlichen Benennung, *castan japonez* (*Aesculus turbinata*) < dt. japanischer Kastanienbaum, *castan roșu* (*Aesculus carnea*) < dt. roter Kastanienbaum; *căcatul dracului* (*Ferula asa-foetida*) < dt. Teufelsdreck, die Pflanze hat einen unangenehmen Geruch (vgl. auch die wissenschaftliche Benennung *foetida* ‚stinkend‘), das rum. Hauptwort *căcat* (‚fecale‘) < lat. *cacatus* (könnte auch vorlateinisch sein); *cânepă de Manila* (*Musa textilis*) < dt. Manilahanf; *coada peștelui*, *coadă de pește* (*Caryota maxima*) < dt. Fischschwanzpalme, weist auf die Form der Blätter hin; *copacul furnicilor* (*Cecropia palmata*) < dt. Ameisenbaum; *copacul lanternă* (*Crinodendron hookerianum*) < dt. Laternenbaum, die Blumen sehen wie Laternen aus; *copacul leopard* (*Caesalpinia ferrea*) < dt. Leopardenbaum, die Rinde ähnelt dem Leopardfell; *copacul popii* (*Evonymus europaeus*) < möglicherweise dt. Pfaffenholz, die Früchte werden von den Rumänen *Pfarrer-kappe* genannt, weil sie der Kopfbedeckung der Priester ähnlich sind; auch die Sachsen nennen sie *Pfaffenkappchen*; *coroana împăratului* (*Fritillaria imperialis*, *Lilium bulbiferum*), *coroană împărătească* (*Fritillaria imperialis*) < dt. Kaiserkrone, rum. *coroană* ‚Krone, Kranz‘ < lat. *corona*; die Blüten sehen so aus; *crinul focului* (*Hemerocallis fulva*) < dt. Feuerlilie; rum. *crin* soll vom slaw. *krinŭ* abstammen; vgl. auch die wissenschaftliche Benennung *crinum* < griech. κρινον, *krinon* ‚Lilie‘, rum. *foc* < lat. *focus*; *două foi* (*Listera ovata*) < dt. Zweiblatt; die Pflanze hat zwei Blätter; rum. *foi* < lat. *folia*; *fasole calabar* (*Physostigma venenosum*) < dt. Kalabarbohne nach der gleichnamigen Küste in Westafrika, wovon diese Art abstammt; *fasole-soia* (*Glycine hispida*) < dt. Sojabohne aus ngriech. *fasóli* und lat. *phaseolus*, beide aus griech. *pháse-*

los ‚Boot‘, auf die Form der Früchte hinweisend; *fenicul de mare* (*Chrithmum maritimum*) < dt. Meerfenchel, der erste Teil der wissenschaftlichen Benennung lat. foeniculum, möglich auch unter dem Einfluss des sächs. fienchøl, dt. Fenchel; *feriga-sabie* (*Nephrolepis* sp.) < dt. Schwertfarn; *ferigea vărgată* (*Asplenium trichomanes*) Streifenfarn, das zweite Wort aus dem lat. variegata; *ferigă solzoasă* (*Notholaena maranthae* heute *Cheilanthes maranthae*) < dt. Schuppenfarn, solz vgl. lat. solidus ‚Goldmünze‘; *floarea broștei*, *floarea broșteii*, *floarea broștelor* (*Taraxacum officinale*) < dt. Krötenblume, die Pflanze wächst an feuchten Orten; rum. *broască* von lat. *broasca oder autochthon, da alb. breskë; *floarea Domnului* (*Dodecatheon pulchellum*) < dt. Gottesblume (*dodecatheon* bedeutet ‚zwölf Götter‘, auf die wunderschönen vielen Blüten hinweisend); *floarea flamingo* (*Anthurium* spp.) < dt. Flamingoblume, die Pflanze hat Blüten, die dem Flamingovogel ähnlich sind (*Phoenicopterus ruber*); *floarea gramofonului* (*Pharbitis purpurea*) < dt. Grammophonblume, die Blumen haben die Form des Grammophontrichters; *floarea lui Io(a)n* (*Hypericum perforatum*), *floarea lui Sf. Ion*, *floarea lui Sântion* (*Galium verum*) Johannisblume, die Blüten der Pflanzen erreichen ihre volle Pracht am 24. Juni; in Rumänien wird Johannes der Täufer an diesem Tag gefeiert; *floarea pasărea paradisului* (*Strelitzia* spp.) Paradiesvogelblume, wegen ihrer Schönheit so benannt; *fructul pasiunii* (*Passiflora edulis*) Passionsfrucht; *floarea Paștelui* (*Anemone nemorosa*) < dt. Ost(er)blume, sächs. Uistərblom; *frunza sângerie* (*Iresine herbstii*) < dt. Blutblatt, frunză ‚Blatt‘ vgl. lat. *frondia < lat. frons, -dis, rum. sângerie ‚blutartig‘ < sânge ‚Blut‘; vgl. lat. *sanguem < sanguis; *ghimber nemțesc* (*Arum maculatum*) < dt. deutscher Ingwer, da die Knolle der Wurzel dem Ingwer (*Zingiber officinale*) ähnlich ist; *grâul vacii* (*Melampyrum pratense*, *M. sylvaticum*) < dt. Kuhweizen, die Kühe, rum. vaca vgl. lat. vacca, altind. vaçā, fressen die jungen Pflanzen, die Butter bekommt eine gelbe Farbe und ist geschmackvoller; *hribul cizmarului* (*Boletus luridus*) < dt. Schusterpilz (*Boletus erythropus*), rum. hrib „Pilz“ < ukr. hryb, tschech., slow. hřib (*Boletus*), russ. grib(î) ‚Pilz‘, das Hauptwort cizmar ‚Schuster‘ < cizmă < ung. csizma, tc. çizme; *iarba buricului* (*Umbilicus* sp.) < dt. Nabelkraut; rum. buric < lat. *umbulucus < lat. umbilicus; rum. *iarbă* ‚Gras, Kraut‘ < lat. herba; *iarba ficatului* (*Hepatica nobilis*) < dt. Leberkraut, aus der wissenschaftlichen Benennung *hepatica* abgeleitet; *iarba inimii* (*Leonurus cardiaca*) < dt. Herzkraut, Herzgespann, von der wissenschaftlichen Benennung *cardiaca* beeinflusst; *iarba păretelui* (*Parietaria officinalis*) < dt. Mauerraute, lat. parietaria; *iarba piperiului* (*Satureja hortensis*) < dt. Pfefferkraut; *iarbă arzătoare* (*Clematis recta*) < dt. Brennkraut, a arde ‚brennen‘ lat. ardere; *iarbă de cositor* (*Equisetum* spp.) < dt. Zinnkraut (die Stiele enthalten Kieselsäure und werden zum Polieren von Zinngegenständen benutzt); *iarbă de molii* (*Verbascum blattaria*) < dt. Mottenkraut, Schabenkraut, lat. blatta ‚Motte‘, die Staubblätter sehen aus wie Mottenantennen; *iasomia nopții* (*Cestrum nocturnum*) < dt. Nachtjasmin, von der wissenschaftlichen Benennung *Jasminum*, ngriech. ghiasemi, iasemi < pers.

yasmine abstammend, die Pflanze sieht dem Jasmin (*Jasminum*) ähnlich und der Geruch ist in der Nacht stärker, rum. *noapte*; vgl. lat. *nox, noctis*; *isop de apă* (*Bacopa myryophylloides*) < dt. Wasserysop, sieht dem Ysop (*Hyssopus officinalis*) ähnlich und wächst im Wasser; rum. *apă* vgl. lat. *aqua* oder i.-e. Wurzel *ap- „apă“; *lacrima miresii* (*Briza media*) < dt. Brautzähre ‚Brautträne‘; die Pflanze hat Ähren, die wie Tropfen oder Tränen aussehen; *lapte de vrăjitoare* (*Chelidonium majus*) < dt. Hexenmilch, wegen dem Latex der Pflanze so benannt; *laur de munte* (*Umbellularia californica*) < dt. Berglorbeer, rum. *laur* < lat. *laurus* (*Laurus nobilis*), ist dem Lorbeer ähnlich aber nicht verwandt; *leuștean negru* (*Smyrniium olusatrum*) < dt. schwarzer Liebstöckel, das deutsche Phytonym selbst ist eine Lehnübersetzung < lat. *levisticum* < griech. *libyistikon* ‚(Pflanze) aus Ligurien‘, das rum. Adj. *negru* ‚schwarz‘ < lat. *nigrum*; *Lizuca cea harnică* (*Impatiens sultanii*) < dt. Fleißiges Lieschen, eine Zierpflanze mit reichen Blüten; *lucernă de Banat* (*Medicago sativa*) < dt. Banater Luzerne (im Banat bevorzugt); *mațele găinilor* (*Stellaria media*) < dt. Hühnerdarm, rum. *maț* ‚Darm‘ < lat. *matia*, *găină* ‚Huhn, Henne‘ < lat. *gallina*; die Pflanzenbenennung weist auf die Form des Stiels hin; *mazăre țucără* (*Pisum sativum*) < dt. Zuckerbohne; Russu (1981) legt das rum. Phytonym *mazăre* ‚Erbse, Bohne‘ als autochthon aus, rum. *țucără* < dt. Zucker, vgl. arab. *sukkar* ‚Zucker‘; *măcriș de mare* (*Nasturtium officinale*) < dt. Meerampfer, rum. *mare* ‚Meer‘ < lat. *mare*; die Pflanze wächst eigentlich nicht am Meer, sondern neben Bächen und Quellen; *paradis* (*Malus pumila* var. *paradisiaca*) < dt. Paradies-Apfel; *măr de balsam* (*Clusia major*) < dt. Balsamapfel, engl. *balsam apple*; *pară de pământ* (*Lathyrus tuberosus*) < dt. Erdbirnenchen, hat birnförmige knollige Wurzeln; *pintenul cocoșului* (*Crataegus crus-galli*, *Plectranthus fruticosus*) < dt. Hahnen-sporn < wissenschaftliche Benennung *plectranthus* < griech. *plektron* ‚Sporn‘ und *ánthos* ‚Blume‘; *piperul călugărilor* (*Vitex agnus-castus*) < dt. Mönchspfeffer, die Früchte haben einen brennenden, scharfen Geschmack und wurden besonders von Mönchen anstatt Pfeffer verwendet, rum. *piper* ‚Pfeffer‘ < (n)griech. *pipéri*, slaw. *piperŭ*, *piperi* < pers. *pippari* < altind. *pippali*; *plăcintăblam* (*Callistephus chinensis*) < sächs. *plətschintəblām*, dt. Palatschinkenblume; Verbindung zwischen rum. *plăcintă* ‚Palatschinken, Strudel‘ und dt. Blume; rum. *plăcintă* < lat. *placenta*, weist auf die Blütenform hin; *ploaie de aur* (*Forsythia suspensa*, *Laburnum anagyroides*) < dt. Goldregen, die Pflanze hat goldgelbe Blumen wie große Regentropfen; rum. *ploaie* ‚Regen‘ < lat. **plovía* < lat. *pluvia*; *pom de mandulă* (*Amygdalus communis* actualmente *Prunus dulcis*) < dt. Mandelbaum, rum. *pom* ‚Baum‘ < lat. *pomus*; *punguliță de stânci* (*Aethionema saxatilis*) < dt. Steintäschel, die Pflanze hat taschenförmige Früchte; rum. *pungă* ‚Tasche‘, rum. *punguță* ‚Täschchen‘; vgl. ngriech. *ponga/pongi*, lat. med. *punga*, slaw. *pangva*, gotisch. *Pugg*, rum. *stâncă* ‚Felsen‘ ist autochthon; *putelizi*, *pute Lizi* (*Pelargonium zonale*) < dt./sächs. Stinkende Liesel/Lieschen, Liese; die Pflanze hat einen unangenehmen Geruch; *rădăcina ciumei* (*Petasites albus*, *Petasites*

hybridus) < dt. Pestwurz; *rădăcina sângelui* (*Potentilla erecta*) < dt. Blutwurz; die Wurzeln sind rötlich und werden zur Herstellung einer roten Farbe benutzt, die Pflanze wurde auch zur Blutstillung verwendet; rum. *sânge* ‚Blut‘; vgl. lat. *sanguem < sanguis; *rădăcina șerpilor* (*Polygonum bistorta*, *Polygonum viviparum*) < dt. Schlangenwurz (*Dracunculus vulgaris*), Natterwurz, vgl. lat. herba serpentina; die Wurzelstöcke sehen den Schlangen ähnlich aus; rum. *șarpe* ‚Schlange‘ < lat. serpes < lat. serpens, kann aber auch vorlateinisch sein; die Staubblätter sehen wie Schlangenzungen aus; *rădăcina ursului* (*Ligusticum mutellioides*) < dt. Bärwurz; *rădăcina de sânge* (*Sanguinaria sp.*) < dt. Blutwurz; rum. *sânge* ‚Blut‘ vgl. lat. *sanguem < sanguis; *regele moliilor* (*Plectranthus fruticosus*) < dt. Mottenkönig, giftig für Insekten; *rută de grădină* (*Ruta graveolens*) < dt. Gartenraute/Rute < griech. rhutos ‚behütet‘; *salată de apă* (*Pistia stratiotes*) < dt. Wassersalat, rum. *apă* < lat. aqua, i.-e. Wurzel *ap- ‚Wasser‘; *scoarță de friguri* (*Cinchona officinalis*) < dt. Fiebrerrindenbaum; die Rinde wird als Heilmittel für Malaria verwendet; *smalț de sticlă* (*Salicornia europaea*) < dt. Glasschmelztz; die Pflanze sieht aus, als wäre sie aus Glas; rum. *sticlă* ‚Glas‘ < slaw. stiklo; *spanac englezesc* (*Rumex patientia*) < dt. Englischer Spinat < ngriech. spanáki < pers. aspanakh; *spinul crucii* (*Rhamnus cathartica*) < dt. Kreuzdorn; entweder weil die Dornen mit dem Ast ein Kreuz bilden, oder weil Jesus Krone vor der Kreuzigung aus dieser Pflanze gewesen sein soll; rum. *spin* ‚Dorn‘ < lat. spinus; *struța apei* (*Veronica anagallis-aquatica*), *struța mirelui*, *struțul mirelui* (*Erigeron canadensis*), *struțișori* (*Selaginella kraussiana*, *Selaginella selaginoides*), *struțușor* (*Lycopodium selago* heute *Huperzia selago*), *struțușori* (*Selaginella helvetica*, *Selaginella selaginoides*) < dt. Strauß; auch im Rumänischen sowohl Blumenstrauß als auch der afrikanische Vogel; rum. *struțișor* ‚Sträußlein‘ bedeutet sowohl ‚Blumensträußlein‘ als auch ‚Trauben‘; *trestie de Spania* (*Calamus caesius*) < dt. Spanisches Rohr; die Pflanze stammt aber aus Malayesien und Kalimantan; *trifoi persan* (*Trifolium resupinatum*) < dt. Persischer Klee; *tufă de foc* (*Hamelia petens*) < dt. Feuerbusch, engl. firebush < rum. *tufă* ‚Busch, Gebüsch‘ < lat. tufa; ein Strauch mit feuerroten Blüten; rum. *foc* ‚Feuer‘ < lat. focus; *umbra nopții* (*Solanum dulcamara*, *Solanum nigrum*) < dt. Nachtschatten (*Solanum nigrum*), lat. umbra noctis; die Pflanze ist dunkelgrün und hat schwarze Früchte; *urzică surdă* (*Lamium album*) < dt. Taubnessel; rum. *surdă* ‚taub‘ < lat. surda, rum. *urzică* < lat. urtica; *verdeata iernii*, *verdiața iernii* (*Chimaphila umbellata*, *Pyrola media*) < dt. Wintergrün.

(c) **Lehnübertragung:** Im Unterschied zur Glied-für-Glied-Wiedergabe der Lehnübersetzung basiert die Lehnübertragung auf einem freieren Umgang mit dem fremdsprachlichen Ausgangswort, das durch eine angenäherte Übersetzung oder aber genauere Ausdeutung wiedergegeben wird. D.h. es geht hier um Teilübersetzungen, freiere Übertragungen von Wörtern einer fremden Sprache: *arbore de friguri* (*Cinchona pubescens*) < dt. Fiebrerrindenbaum; *cănișoară* (*Nepenthes spp.*), *cănișoară purpurie* (*Sarracenia*

purpurea): Karnivorpflanze mit kannenförmigen Blüten als Falle für Insekten; rum. Hauptwort *cană* ‚Tasse, Kanne‘, scheint vorlateinisch zu sein; die rumänischen Pflanzenbenennungen sind Lehnübertragungen des dt. Kannenpflanze; *căpșună grasă* (*Fragaria moschata*) < dt. Riesenerdbeere, rum. *căpșună* ‚Erdbeere‘ < *căpșon < rom. căpșor < cap ‚Kopf, Haupt‘, rum. *grasă* ‚dick, fett‘ < lat. *grassus < lat. crassus, hier in der Bedeutung ‚groß‘; *gulerăș* (*Carpesium cernuum*) < dt. Kragenblume, rum. *gulerăș* ist die Diminutivform von rum. *guler* ‚Kragen‘ < ung. gallér, weist auf die Blüten hin; *floarea lui Cristos* (*Helleborus niger*) < dt. Christrose, Weihnachtsrose; *o mie de florini* (*Centaureum erythraea*) < dt. Tausendgüldenkraut; weist darauf hin, wie wertvoll diese Heilpflanze war; *finhienu bălții* (*Oenanthe aquatica*) < dt. Wasserfenchel; die Pflanze sieht *Foeniculum vulgare* ähnlich, wächst aber in Teichen; *floare de o zi* (*Commelina communis*) < dt. Tagblume, rum. *zi* ‚Tag‘ aus dem lat. die; *floare popească* (*Narcissus poeticus*) < dt. pfarrersche Narzisse, sächs. fuaresch narziss, rum. *popească* < *popă* < slaw. *popŭ*, lat. *popa*; *iarba miresii* (*Briza media*) < dt. Brautzähre ‚Brautträne‘, rum. *iarbă* < lat. *herba*; *iarbă de doctorit ficatul* (*Ageratum haustonianum*) < dt. Leberbalsam aus lat. *med. doctor* und lat. *ficatum*; *iarbă englezească* (*Lolium perenne*) < dt. Englisches Raygras, Englisches Weidelgras; *lacrima doamnei* (*Ophrys fuciflora*) < dt. (Hummel) Frauentränen, rum. *lacrimă* < lat. *lacrima*; *lacrimile miresei* (*Convallaria majalis*, *Dicentra spectabilis*) < dt. Brautzähre, sächs. brautzäir; rum. *lacrimile* ist die artikulierte Pluralform des rum. *lacrima* < lat. *lacrima*; *lemnul Sanctului Ioan* (*Ribes rubrum*) < dt. Johannisbeerstrauch; *păr de stâncă* (*Amelanchier ovalis*) < dt. Felsenbirne, rum. *păr* < lat. *pirus/pyrus* ‚Birnbäum‘; *pomul Domnului* (*Ailanthus altissima*) < dt. Götterbaum; eine Lehnübertragung von *moluc ailanto* („Himmelsbaum“).

(d) **Semantische Veränderungen. Volksetymologie:** *barba Saftei* (*Rheum officinale*, *Rheum palmatum*) < dt. Rhabarberschaft, das rum. Phytonym lautet ‚Saftas Bart‘; Safta ist ein weibliches Anthroponym; *clei(u)* (*Medicago sativa*, *Trifolium spp.*) < dt. Klee (*Trifolium*), Schneckenklee (*Medicago*); *Clocăboambă* (*Aquilegia vulgaris*) < sächs. *kliokəbläom*, *klokəblommən*, dt. Glocke(n)blume, rum. *clocă* ‚Glucke, Bruthenne‘ und rum. *boambă* ‚Weinbeere, Beere, Samen‘; *crinul fânului* < dt. Feuerlilie, das rum. Phytonym *crin* soll von slaw. *krinŭ* abstammen, aber es gibt auch die wissenschaftliche Benennung *crinum* < griech. *κρινον*, *krinon* ‚Lilie‘; *drăgan* (*Astragalus gummifer*) < dt. (Gummi)Tragant, Dragant; *floarea oștilor* (*Anemone nemorosa*) < dt. Ost(er)blume, sächs. *Uistərblom*; *franț* (*Galinsoga parviflora*) < dt. Franzosenkraut; *frățiori* (*Viola tricolor*). Dies sei nach Drăgulescu (2010) eine Eindeutung nach dt. Fratze; *măzărîche* (*Briza media*, *Nigella arvensis*) < sächs./dt. Maschehirsch < dt. Mönschenhirse (*Briza media*) bzw. sächs. *muasertcher* (*Nigella damascena*); *măzărăș* (*Aposeris foetida*) < sächs. *Huasngras* < dt. Hasengras oder Muasergas; *mierluță* (*Minuartia verna*) hat nichts mit dem rum. *mierlă* ‚Amsel‘ zu tun; es handelt sich hier um eine Volksetymologie von **mieruță* nach dem dt./sächs. *Miere*; *moș gras* (*Phalaris*

arundinacea var. *picta*) < sächs. moschongras, dt. Maschengras; *salată* (*Allium ascalonicum*) < dt. Schallote, Schalloten; *săgețel* (*Carex arenaria* heute *Carex colchida*, *Gladiolus gandavensis*) < dt. Segge, Segese ‚Sense‘; die Pflanzen haben schneidende Blätter < rum. *săgeată* ‚Pfeil‘ < lat. *sagitta*; *troacă* (*Cucurbita pepo*, *Lagenaria siceraria*), *trochițe* (*Lagenaria siceraria*) halbe Früchtenschalen wurden als kleine Troge benutzt, rum. Hauptwort *troacă* ‚Trog‘ < sächs. *troch*, dt. *Trog*.

(e) **Mischbildungen:** hier handelt es sich um direkte phonetisch unveränderte Übernahmen aus der deutschen Sprache oder aus dem siebenbürgisch-sächsischen Dialekt, eine angenäherte Übersetzung oder Ausdeutung und ein rumänisches (sogar autochthones) Wort: *acacie boierească* (*Sophora japonica*) < dt. Akazie, rum. *boierească* < Adjektivform des rum. *boier* < slaw. *bolijarinu* ‚Herr‘; *bostan de lompău* (*Lagenaria siceraria*): Kürbis, die Frucht wird dazu verwendet, sogenannte „lompău/lompă“ fertigzustellen < dt. Lampe, rum. *bostan* < tc. *bostan kabaği* ‚Gartenkürbis‘; *bușteanu(l) Ielelor*: vgl. sächs. *bumstam*, dt. Baumstamm und rum. *iele* ‚Feen/Hexen‘; *cârligei de somot* (*Viola wittrockiana*) die Blüten sind samtig, rum. *somot*, *somet* ‚Samt‘ < sächs. *sumət*, dt. *Sammet*, *Samt*, rum. *cârlig* ‚Hacken‘; nach Russu (1981) auf das i.-e. Wurzel *(s)ker- ‚zurückkehren, beugen, biegen‘ zurückführbar; *chel varză creață* (*Brassica oleracea* var. *sabauda*) < sächs. *ki(e)l*, *kîl*, dt. Köhl(kraut); *cherbăl auriu* (*Chaerophyllum aureum*), *cherbăl sălbatic*, *chervăl sălbatic* (*Anthriscus sylvestris*) < dt. Kälber(kropf), dt. Kerbel, sächs. *kiərwəl*, *kerwəl*, *Wald/Wiessekerbel*, < *cerefolium* < griech. *chairphyllon* ‚angenehmes Blatt‘; *chifăr roșu* (*Pinus sylvestris*), *chifer negru* (*Pinus nigra*) < dt. Kiefer; *corfa jupânesii* (*Tropaeolum majus*) < rum. *corfă* ‚Korb‘ < sächs. *korf*, dt. *Korb* > ung. *karfa*, es gibt auch ein griech. *korban* und lat. *corbis* davon fr. *corbeille*, rum. *jupâneasă* ‚Bojarin; Frau aus der Stadt‘ < Femininform von rum. *jupân* < slaw. *županü*; so genannt wegen der korbförmigen Blumen; *crumpei de lună* (*Solanum tuberosum*), *crumpene porcești*, *crumpi porcești*, *crumpene sălbatic* (*Helianthus tuberosus*) < sächs. *krumpirn*, dt. Grundbirne, unter dem Einfluss des ung. *kolompér*, *kolompéra*, *krompé*, *krumpi*, *krumpli* < dt. Grundbirne; *feldera pământului*, *ferdela pământului* (*Cyathus* spp.) rum. *felderă*, *ferdelă* ‚Maßeinheit für Getreide‘ < sächs. *fyrdeł*, dt. Viertel¹⁸; *floare din șanț* (?*Sieglingia decumbens*) wächst auf den Schanzen, *ghințură crăcoasă* (*Gentianella austriaca*), *ghințură de primăvară* (*Gentiana verna*), *ghințură galbenă* (*Gentiana lutea*), *ghințură gușată* (*Gentiana utriculosa*), *ghințură pătată* (*Gentiana punctata*), *ghințură tomnatică* (*Gentiana pneumonanthe*), *ghințură vânăță* (*Gentiana cruciata*) < dt. Enzian; *foalchin de iarnă* (*Cheiranthus cheiri*), *foaltine bogate* (*Mathiola incana* ‚mit reichen Blüten‘), *foaltine sărace* (*Mathiola incana* ‚mit einfachen Blüten‘) < sächs. *foaltchər*, *vältchər*, *foalen* < dt. Veilchen (*Mathiola incana*), sächs. *giel foalen*, dt. gelbes Veilchen (*Cheiranthus cheiri*), sächs. *voälchen*, dt. Veilchen (*Viola* spp.); *grâu spelt* (*Triticum spelta*) < dt. Spelt, Spelz, vgl. i.-e. Wurzel *pel-, pel-t- ‚Staub, Mehl‘; *iarbă de*

tripăl (*Digitalis grandiflora*) zur Behandlung von Tripper verwendet; *ințură galbenă* (*Gentiana lutea*, *Gentiana punctata*) < dt. Enzian; *levcoai* *galbenă* (*Cheiranthus cheiri*), *levcoai* *roșie* (*Matthiola incana*) < dt. Levkoje < altgriech. *leucoion* (*Matthiola incana*), aus *leucós* und *ion* gebildet, mit der Bedeutung ‚weiße Veilchen‘; es gibt aber auch Arten mit gelben (rum. *galben*) und roten (rum. *roșu*) Blumen; *lilie albă* (*Lilium candidum*), *lilie galbenă* (*Iris pseudacorus*, *Lilium bulbiferum*), *lilie roșie* (*Lilium bulbiferum*), *lilie sălbatică* (*Iris germanica*), *lilie vânătă* (*Iris florentina*, *Iris germanica*), *lilium alb* (*Lilium candidum*), *lilion alb* (*Convallaria majalis*, *Lilium candidum*), *lilion roșu* (*Hemerocallis fulva*), *lilion albastru* (*Iris germanica*), *lilion bun* (*Convallaria majalis*), *lilion galben* (*Hemerocallis fulva*), *lilion pășăresc* (*Iris variegata*), *lilion sălbatic* (*Majanthemum bifolium*), *lilion vânăt* (*Iris germanica*), *liliu galben* (*Iris pseudacorus*), *liliu vânăt* (*Iris germanica*) < dt. Lilie, sächs. Leljen < lat. *lilium* ‚Lilie‘; *lipscănoaice*, *lipscănoaie* (*Coreopsis tinctoria*) < rum. *Lipsca* < Leipzig unter dem Einfluss des slaw. Lipskū; *lurbău de toamnă* (*Phlox paniculata*) < sächs. lührber(buhm), lu(i)rbæn, lu(i)rbær, liurbær, dt. Lorbeer, hat dem Flieder (*Syringa vulgaris*) ähnliche Blumen bis spät im Herbst; rum. *toamna* < lat. *autumnus*; *mierluță pitică* (*Minuartia recurva*, *Minuartia sedoides*), *mierluță țepoasă* (*Minuartia setacea*); *măghiranul pământului* (*Lysimachia nummularia*) ist *Majorana hortensis* nicht ähnlich (einziges gemeinsames Merkmal sind die runden Blätter); der Name *măghiranul pământului* ‚Erdmajoran‘ weist darauf hin, dass die Art wild ist; *mihlele* (*Lycopersicon esculentum*): nach der Ortschaft Mihla in Deutschland, woher diese Tomaten nach Rumänien eingeführt worden sind, vgl. das Anthroponym Mihle; *ortul popii* (*Lysimachia nummularia*) < rum. *ort* ‚Münze‘ < dt. Ort, poln. *ort*, rum. *popă* < slaw. *popŭ*; die Blätter sehen wie kleine Münzen aus, vgl. die wissenschaftliche Benennung *nummularia* < lat. *nummulus* ‚wenig Geld, oder von geringem Wert‘; *părădăici roșii* < sächs. Parædeis, dt. Paradeis-(apfel), Paradies(apfel), ung. *paradicsom* < dt. Paradeisen; *rosmalin cu miros mândru și frumos* (*Philadelphus coronarius*) < dt. Rosmarin; *rabarbăr sălbatic* (*Polygonum cuspidatum*) ist *Rheum rhabarbarum* ähnlich; *rozmarin de baltă* (*Ledum palustre*), *rozmarin de munte* (*Gnaphalium silvaticum*), *rozmarinul câmpului* (*Ajuga chamaepytis*) manche werden angebaut, andere wachsen wild auf Feldern, im Gebirge oder in den Mooren; *spargă aninată* (*Asparagus sprengeri*) < sächs. spargel, dt. Spargel > ung. *spárga*, rum. Verb *a anina* ‚(auf)hängen‘; *țâtron de friguri* (*Sempervivum tectorum*), *țâtron domnesc* (*Veratrum nigrum*), *țâtron sălbatic* (*Hedera helix*, *Ruscus hypoglossum*), *țitroană*, *țitrom* (*Citrus limon*), *țitron* (*Citrus limon*, *Philadelphus coronarius*) < dt. Zitrone, Zitron, im Süden Siebenbürgens ist mit *zitron* eine immergrüne Pflanze gemeint; *țelăr sălbatic* (*Peucedanum sp.*, wahrscheinlich *Peucedanum oreoselinum*) < sächs. zellær, dt. Zeller, rum. *sălbatic* ‚wild‘ < lat. *silvaticus* ‚aus dem Wald, waldig‘ < lat. **salvaticus*; *urzică de șură* (*Urtica urens*) < sächs. schyren, dt. Scheuer, Schur, vgl. ukr. *šura*, rum. *urzică* ‚Nessel‘ < lat. *urtica*.

Die Transferenzen können nach dem Ausmaß ihrer Integration in das rumänische Sprachsystem (Assimilation, Eingliederung, „Rumänisierung“) gegliedert werden in: (a) phonemische Integration: *acacie, acație; bambus; bacolți; berten, bertin; bonă; bordă; boreț; bufă, bufoi, bufon(iu); buhaci, buhaș, buhășer; calarabă, calarambă, călărabe, caralabă, cararabe, cărălabă, chelărabe, chelarabă, chelărabi, chelărabie; carfiol; cărneauă, hernaive; cartifiol, cartofiol, cartofiom; cartoafe; castan; chel, chil, chiel, chil; chelărade, chelerab; cherbel, cherbăl, chervoăl; chifăr, chifer; clocăboambă; cocarde; crețaruș(i); cricin, crihim, crihin, crichin; crumciri, crumpe, crumpei, crumpene, crumpeni, crumperi, crumpi, crumpiri, crumple, crumpli, crupre; crumpei; cule; cuțcrișoară, cuțcrișor; diptam; dost; edălvais; ențură, enzură; eră; erdăpane; fefărminț; fenicel; fenicel; fetică; fiache; fidireie; finc(h)en, finchăn, fenhiel, finhien, finchil, finc(h)in, fincăl; flos; foalcer, foalchine, foalâne, foaline, foaltăr, foalter, foaltine, foahine, foanchine; garoa, garoă; gerbera; ghințuire, ghințură, ghințurea, ghințurele, ghințurică; gonele; grap; grubă, grube, grumbă(i), grumbănă, grumbe, grumcizi, grumpă, grumpe, grumpene, grumpeni, grumpi, grumpin(i); hăidănuș(ă); heber; heber; heciumpeci, heciunpeci; holăr; holăr; hulcă; inghe, inguțe; ințură, ințurea, ințurică; levcoaie; libistoc; lilie; liurbă(u), l(i)urben, l(i)urber, liorbe; liurben, lurbău; liză; mace; maieran, maioran, maiorean, măierean(ă), maiorană, măgheran, măgherean, măghieran, măghiran, măghieran, măgieran, măgiran, măioreană; mandulă; mangold; minuțe; mure; mutmel, mutruel, muturel; nudli; ortanță, ortensă, ortensie, ortenzie; părădaice, părădaisăne, părădaise, părădăi, părădăiși, părădazine, părădăici, parad(a)ise, paradaice, paradeise, paradice, paradici, bărădăi, porobici, porodăi, porodăici, porodeici, porodică, porodici, porodiciuri, porodiță; petersil; pilzene; pizang; pomaranță, pomeranță; pospănc, pospang, puspan, pușpan, sospa(i)n; rabarbară, rabarbură, rabarbăr, rebarbar(ă), rebarboră, robarbă, robarbăr; ramie; râpac, râpă, râpe, rips; rapăn; rebizlă, ribiză, ribizil, ribizle, ribizli, ribizlă, ribuțe, răbiză, ribizle; rosmailin, rosmalin, rosmarin, rosmarint, rosmălin, rosmărin, rosmolin, rozmalin, rozmarin, rozmolin; ruben; șalotă; șarlai, șerlai, șerlau; silairei; simsă; șlaierblume; șnirling; soloagăr; spargă; spic; spichinat, spichinel; spinac; spinat; ștainglezăr; streliția; taghete; țâncraus; țătroană, țătron, țătroană, țitrom, țitron; țelăr, țeler, țeleră; țicorie; țimat, țimărt, țimăt, țimet, țămărt; titenchi, titenghi; toibă; trifțărele; trimpăn, trinpăn; țucroș, țucuran, țucurechi; vasărlilie; vermut; zegras u.a.; (b) graphemische Integration: die rumänische Rechtschreibung gründet auf dem phonologischen Prinzip, sodass die oben aufgezählten Pflanzennamen genauso geschrieben wie sie ausgesprochen werden; (c) semantische Integration: Übernahmen mit einer von der Herkunftssprache teilweise abweichenden Bedeutung sind oft bemerkbar (vgl. die Beispiele oben); (d) sprachsoziologische Integration: Übernahmen mit einem geringen Integrationsgrad aufgrund des Vorkommens in der fachgebundenen Kommunikation und in exklusiven Kreisen (Eliten) kommen auch vor. Die meisten Formen werden von den Dorfbewohnern benutzt und sind, mit wenigen Ausnahmen, nur in Siebenbürgen belegt.*

4. Fazit

Das Inventar an rumänischen Pflanzennamen ist auch durch die Übernahmen

aus dem Deutschen und seiner Standardvarietät Österreichisches Deutsch sowie den deutschsprachigen Dialekten Siebenbürgens (das Siebenbürgisch-Sächsische und das Landlerische) bereichert worden. Bei der Herleitung der im Wörterbuch aufgenommenen Bezeichnungen wurde ihr Herkunftsgebiet bzw. die Ortschaft berücksichtigt, um eventuelle Wortvarianten eruieren zu können, die keine Pflanzenbezeichnungen darstellen und folglich nicht zum Forschungsgegenstand gehören. Dabei sind als Vergleichsgrundlage oft Pflanzenbezeichnungen der Siebenbürger Sachsen herangezogen worden.

In meiner bisherigen wissenschaftlichen Tätigkeit bin ich auf die enge Beziehung zwischen Mensch und Pflanzenwelt eingegangen, wobei ich auch Quellen der (sekundären) Nominaton, Denotation und Konnotation erfasst habe. Mit dem hier lexikografisch umrissenen Bereich, der Ethnobotanik, tut sich ein weites Forschungsgebiet auf, das für anthropologische, psychologische oder psycholinguistische Untersuchungen verwendbar ist, da die Pflanzennamen ein linguistisch wertvoller Schatz darstellen, auf vielfältige Repräsentationen hinweisen und den affektiven Unterbewußtsein einer Gemeinschaft enthüllen.

Endnoten

1. Mehr über das Verhältnis zwischen Linguistik und Lexikographie siehe Hartmann (2015: 73-74).
2. Eine hervorragende Forschung hat Reichmann (2012) unternommen. Siehe dazu auch Smit (2016).
3. Mehr dazu siehe Drăgulescu (2010: 18).
4. Vgl. hierzu *Lexiconul lui Mardarie Cozianu* (1649) 42 Bezeichnungen, *Codex Sturdzanus* (ca. 1660–1670), *Lexiconul Marsilian* (ca. 1700) oder das Lexikon aus dem *Fragmentul de la Bistrița* (ca. 1700) 75 Bezeichnungen, siehe Drăgulescu (2010: 18).
5. Vgl. Borza (1958) und (1961).
6. Vgl. das Kochbuch von Mihaiu (1749), *Istoria poamelor* (1773) oder die Arbeiten von Grisellini (1780) und Raicevich (1788), mehr darüber in Drăgulescu (2010: 18).
7. Drăgulescu (2010: 18).
8. Zum *Dicționarul Borza* vgl. Chivu (2008) und (2010 I: 333-340).
9. Drăgulescu (2010: 18).
10. Dies ist der Sammeltätigkeit zahlreicher Fachleute, Biologen, Ärzte, Volkskundler, Philologen, zu verdanken: hierzu ausführlicher Drăgulescu (2010: 18).
11. Es handelt sich hierbei um Bezeichnungen, die D. Brândză, J. Chr. Baumgarten, Fl. Porcius, G. Barițiu, A. Fătu, N. Șuțu, I. Czihak, G. Crăiniceanu, I. Bob u.a. anführen.
12. So kam ein Inventar von über 1000 rumänischen Pflanzenbezeichnungen zustande.
13. Vgl. Drăgulescu (2010).
14. C. Drăgulescu (2010) meint, dass die Benennung nicht mit dem rum. Hauptwort „fată“ (Mädchen) in Verbindung gebracht werden kann. Er verweist auf dt. *fettig*, auf die „fetten Blätter“, die als Salat verzehrt wurden; möglich auch dt. *Feldsalat*.
15. Möglicherweise eine schwäbische Variante, da die Benennung im Kreis Caras-Severin verbreitet ist. Die Blätter sehen wie Nudeln aus.
16. 1773 wurde sie zu Ehren der britischen Königin Sophie Charlotte, einer geborenen Prinzessin von Mecklenburg-Strelitz und Gemahlin Königs Georg III., benannt.

17. Vgl. awest. *pairi-daeza* („mit Mauern umschlossener Ort“, später ‚Jagdpark, Jagddomäne‘) und auch die Bildungen mit *para* und *diza* („umschlossener Ort, Burg“) in zahlreichen trakischen Toponymen.
18. Diese Pilze werden auch von den Deutschen *Erdviertel* genannt.

Literatur

- Borza, Al.** 1958. Numiri românești de plante în vocabulare și dicționare din sec. XVII-XVIII. *Cercetări lingvistice*. Cluj, III.
- Borza, Al.** 1961. Aportul etnobotanic al lexicografiei vechi slavo-române și al altor lucrări din secolul al XVIII-lea. *Studii și comunicări*. București, III und *Apulum*. Alba Iulia, IV.
- Borza, Al.** 1968. *Dicționar etnobotanic*. București: Academia R.S.R.
- Chivu, Gh.** 2008. *Dictionarium Valachico-Latinum. Primul dicționar al limbii române*. Studiu introductiv, ediție, indici și glosar de Gh. Chivu. București: Academia Română.
- Chivu, Gh.** 2010. Nume de plante în *Dictionarium valachico-latinum. Limba română. Controverse, delimitări, noi ipoteze*. București: Universității din București.
- Drăgulescu, C.** 2010. *Dicționarul explicativ al fitonimelor românești*. Sibiu: Editura Universității „Lucian Blaga“.
- Drăgulescu, C.** 2014. *Dicționar de fitonime românești*. Sibiu: Editura Universității „Lucian Blaga“.
- Hartmann, R.R.K.** 2015. Lexicography and Its Interdisciplinary Contacts, with Special Reference to Linguistics and Onomasiology. *Lexikos* 15: 70-89.
- Reichmann, O.** 2012. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen*. *Studia Linguistica Germanica* 111. Berlin/Boston: Walter de Gruyter.
- Russu, I.I.** 1981. *Etnogeneza românilor*. București: Științifică și Enciclopedică.
- Smit, M.** 2016. 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. *Lexikos* 26: 446-455.

Visualisation of Collocational Preferences for Near-Synonym Discrimination

Mariusz Piotr Kamiński, *Basic Science Centre,
University of Applied Sciences, Nysa, Poland*
(mariusz.kaminski@pwsz.nysa.pl)

Abstract: This paper aims to explore the potential usefulness of two techniques that visualise collocational preference for the purpose of synonym discrimination. Given the fact that collocation is one of the most important markers of meaning difference, it is used as the criterion for distinguishing between near-synonyms. Collocational preferences for a set of near-synonyms (*artificial, fake, false, and synthetic*) were visualised using two techniques: correspondence analysis plot and collocational network. The collocations were retrieved from BNC corpus by using a distributional method. An advantage of the graphs is that they allow lexicographers to spot similarities and differences in collocational preference of several words in a single diagram. Such visualisations may be used as an alternative way to a tabular form of data presentation to avoid information overload which arises when lexicographers prepare synonym essays for productively-oriented dictionaries. The visualisations can be used as a starting point for exploring semantic differences between semantically similar words.

Keywords: VISUALISATION, CORRESPONDENCE ANALYSIS, COLLOCATIONAL NETWORK, SYNONYMY DISCRIMINATION, COLLOCATION, AUTOMATIC RETRIEVAL OF SYNONYMS, NEAR-SYNONYM, COLLOCATION PREFERENCES

Opsomming: Die visualisering van kollokasionele voorkeure by die onderskeiding van ampersinonieme. Hierdie artikel beoog om die potensiële bruikbaarheid van twee tegnieke te ondersoek wat kollokasionele voorkeure visualiseer met die oog op sinoniem-onderskeiding. Aangesien kollokasies een van die belangrikste merkers van betekenisverskil is, word dit gebruik as maatstaf om tussen ampersinonieme te onderskei. Kollokasionele voorkeure vir 'n stel van ampersinonieme (*artificial, fake, false en synthetic*) is gevisualiseer deur twee tegnieke te gebruik: ooreenkomsanalise-stipping en kollokasionele netwerk. Die kollokasies is onttrek uit die BNC korpus deur gebruik te maak van 'n verdelingsmetode. 'n Voordeel van die grafieke is dat hulle leksikograwe in staat stel om ooreenkomste en verskille ten opsigte van kollokasionele voorkeur raak te sien vir verskeie woorde in 'n enkele diagram. Sulke visualiserings kan gebruik word as 'n alternatiewe manier vir 'n tabellariese manier van datavoorstelling wat inligtingsoorlading voorkom wat ontstaan wanneer leksikograwe sinoniemreekse voorberei vir woordeboeke wat op produksie fokus. Die visualiserings kan gebruik word as 'n vertrekpunt om semantiese verskille te ondersoek tussen woorde wat semanties soortgelyk aan mekaar is.

Sleutelwoorde: VISUALISERING, OOREENKOMSANALISE, KOLLOKASIONELE NETWERK, SINONIEMONDESKIEDING, KOLLOKASIE, OUTOMATIESE ONTTREKKING VAN SINONIEME, AMPERSINONIEM, KOLLOKASIEVOORKEURE

1. Introduction

Synonym discrimination may be a challenge for lexicographers constructing synonym essays for productively-oriented dictionaries. In such essays, words of similar meaning are brought together and their meaning discriminated by specifying the particular points where they overlap or differ.¹ According to Landau, the inclusion of synonym discrimination in general dictionaries "makes a great deal of sense" and they are "a welcome superfluity" (2001: 141).

For many years, synonym discrimination, as a lexicographer's task, was laborious and time-consuming, as it was carried out manually and relied largely on the lexicographer's introspection and intuition.² However, recent developments in computer technology offer support to this process by automating data excerption, visualisation, and analysis. Given the fact that collocation is one of the most important markers of meaning difference (see section 2) and easily accessible from corpora, contemporary lexicographic software uses collocation behaviour as the criterion for distinguishing between near-synonyms (Kilgarriff 2013).

Sketch Engine, one of the widely used systems for writing dictionaries, offers a tool called Sketch Diff, which supports comparing and contrasting two words by producing word sketches (see Kilgarriff 2013).³ These are summaries of the collocational behaviour of each word, presented in a tabular form, and accompanied by a statistical measure of collocational strength for each co-occurring word. Using this software, lexicographers are provided with collocational information in a clear and accessible way, which is essential for further analysis of semantically similar words. However, the tabular form of contrasting words, which is used in Sketch Diff, helps lexicographers compare two words, rather than three or more at once, which means that lexicographers engaged in writing panels for several synonyms need to conduct several analyses on all possible word pairs in order to draw conclusions as to collocational overlap and differences. Thus, the question arises whether it would be possible to compare several synonyms in a single graph, by showing their collocational preference in such a way that the investigator can see at a glance which words are closer to each other by virtue of their grammatical-collocational behaviour.

This paper aims to explore the potential usefulness of two visualisation techniques: correspondence analysis plot and collocational network. The focus is to discriminate between a set of words (*artificial*, *fake*, *false*, and *synthetic*) by visualising their collocational preferences. Correspondence analysis (henceforth CA) is a frequency-based multivariate analysis used in statistics and quantitative studies. Its output has a form of a plot which reveals associations in corpus data on the basis of co-occurrence of different features (see Glynn

2014: 443, Greenacre 2007). In turn, collocational network is a diagram that displays links between lexical units, showing how a given unit relates syntagmatically to other units in the lexical structure. Both ways of presentation of data help the analyst to spot collocational overlap between near-synonyms, which is obscured in a traditional textual or tabular form. The focus on collocational overlap is motivated by the need to establish similarities in collocation preferences across synonyms.

Despite enormous developments in computer lexicography, the above two forms of visualisation remain underexplored in lexicography. In particular, correspondence analysis is novel to dictionary making, though it has recently been used in related studies, i.e. in dictionary comparison, to visualise similarities in definition style (Kamiński 2015), and in corpus linguistics, to display semantic content of a set of moderators (Desagulier 2014). As for collocational networks, they have been employed in a few dictionary projects, for example, in constructing dictionary entries in *E-Advanced Learner's Dictionary of Verbs in Science* (see Alonso et al. 2011), and in building lexical systems as graph models of lexicons (Polguère 2014). The potential of the two visualisation techniques is worth exploring now, as computer technology is opening up new opportunities for lexicographic projects.

2. Synonymy and collocation

Synonymy is a relationship between two or more lexical items which have the same meaning. Because complete synonymy presupposes a rather unlikely situation in language, namely the substitutability of lexical items in all contexts and identity of denotative and connotative elements (Bussmann 1996: 1164), it is more convenient to extend the definition of synonymy to cover items that are close enough in meaning to be substitutable in some contexts, without changing the meaning of the sentence (Crystal 2003: 450, Lewandowska-Tomaszczyk 1990). In other words, it is possible to identify candidates for synonyms by selecting words that have the same collocational pattern, and disregarding differences in connotation due to particular regional, socio-dialectal or stylistic variation (Bussmann 1996: 1165). Of all the features distinguishing between near-synonyms, such as language variety, register, emotive content and degree of specificity (Moon 2013: 261), collocation is one of the easiest to retrieve from a corpus (Sinclair 1991: 170). As Firth (1957: 11) neatly put it: "you shall know a word by the company it keeps." According to Firth (1968: 181), "collocations of a given word are statements of the habitual or customary places of that word". Firth was the first to use the term collocation, and his definition above arguably implies a quantitative nature of collocation study (cf. Krishnamurthy 2000: 32; Xiao and McEnery 2006: 105). Although collocation analysis has become a well-established area of corpus linguistics, it is only relatively recently that collocation has been systematically exploited in dictionary making as a criterion for synonym discrimination.⁴

3. Collocation retrieval

Collocations can be acquired automatically from a corpus in at least two ways: one that depends on proximity of two words (Kilgarriff and Kosem 2012), and the other on grammatical relationships. The former rests on the assumption that two words are collocates of one another if they co-occur with a frequency far greater than chance (Atkins and Rundell 2008: 369). In this approach, which is used on an untagged corpus, a candidate for a collocate of a node word is established within a given span around the node by calculating the association strength between the two words. A commonly used measure of association strength is Mutual Information (MI), which is based on the probability of occurrence of the combination of the node and the collocate compared to the probabilities of the two words separately (Church and Hanks 1989). Apart from MI, there are other statistics, including t-score, z-score, log-likelihood ratio, the Dice coefficient, and logDice (Rychlý 2008). However, this proximity-based approach, which is based solely on calculating collocation measure, has a disadvantage of yielding a certain amount of noise in the form of rare words (in the case of MI) or function words (log-likelihood) (Kilgarriff and Kosem 2012). Because collocates are gathered together irrespective of the functional relations to the node word, extra effort is needed to identify the collocates that stand in specific relations to the node word. Furthermore, this approach requires one to set up a span within which the collocates are sought, which may be problematic for lexicographers who have no idea of how far from the node word they should look for (Kilgarriff and Tugwell 2002: 127).

The other approach, which is based on grammatical relationships, does not have the shortcomings of the former. It relies on distribution of words, that is only the collocates that stand in a relation of grammatical dependency to the node word are retrieved. For that to be possible the corpus has to be syntactically parsed with grammatical structure of sentences identified. As a result, one can retrieve lists of subjects, objects and other functional categories of collocates that a given word relates to. This approach has been used by Kilgarriff et al. (2004) in Sketch Engine. In this software, collocations extracted from a parsed corpus are accompanied by the measure of collocation strength, which helps the lexicographer reduce the amount of information overload.

In this paper the distributional approach will be used in extracting collocations. Collocation types and their frequencies will be supplemented by a statistical measure of collocation strength (Rychlý's logDice⁵).

4. Method

4.1 Material and preparation of data

The data were retrieved from BNC XML Edition⁶ and processed according to the following procedure:

- (1) retrieving full sentences containing the adjectives from the corpus,
- (2) parsing the corpus sentences with the Stanford Parser (Chen and Manning 2014),
- (3) retrieving 10 most frequent collocates for each adjective, in either attributive or predicative positions,
- (4) creating a matrix of the adjectives and their top-frequency collocates, ordered according to the overall frequency,
- (5) converting the raw frequencies into a measure of collocational strength (for collocational network),
- (6) visualising the data

The sentences containing either of the adjectives under study were retrieved from the corpus and cleaned up of all BNC tags. They were then parsed with the Stanford Parser, a program that outputs grammatical dependencies within a sentence. Of all 50 typed dependencies produced by the parser, only two were of interest to us, namely the ones in which the adjective stands in either attributive or predicative position with respect to the noun. In the Stanford Parser, the above relationships are indicated as "amod" and "nsubj", respectively. Below is an example output for two sentences "*But that was something of a false dawn*" and "*The style is artificial*". In the first sentence, the relationship between *false* and *dawn* is indicated by "amod", and in the second one the relationship between *style* and *artificial* by "nsubj".

```
cc(something-4, But-1)
nsubj(something-4, that-2)
cop(something-4, was-3)
root(ROOT-0, something-4)
case(dawn-8, of-5)
det(dawn-8, a-6)
amod(dawn-8, false-7)
nmod:of(something-4, dawn-8)

det(style-2, The-1)
nsubj(artificial-4, style-2)
cop(artificial-4, is-3)
root(ROOT-0, artificial-4)
```

In order to reduce the parsing time, the maximum sentence length to parse was set to 100, which still allowed us to retrieve over 90% of the relevant sentences. From the parser output, a list of all collocates was extracted for each adjective. Function words such as "one", "it", "they", "that", etc., which stood in the subject position, were excluded from the list. A subset of the top 10 most frequent collocates for each adjective was ordered according to the overall frequency. As a result, a table of most frequent collocations was obtained (see Table 1). An advantage of constructing such a table is that one can see which adjectives

share the same collocates. However, such a table is more difficult to read, when more data are included. In order to enhance the interpretation of the data, they were visualised using CA plot and network.

4.2 Visualisation

Table 1 brings together raw frequencies of collocations. The table was used as input to correspondence analysis (see Fig. 1). CA converts the frequencies of feature co-occurrence to distances and displays them in a two-dimensional graph. The graph shows correlations between data points through distance and proximity. The same input data were used to generate the collocational network (see Fig. 2 and Fig. 3), with the difference being that the raw frequencies were converted into Rychlý's logDice scores, a measure of correlation strength (see Table 2). For better clarity, the collocational network in Fig. 3 was generated only for strong collocations by removing the collocations with logDice lower than 5.

	collocate\adjective	artificial	fake	false	synthetic
1	intelligence	97	0	0	0
2	pretences	0	0	82	0
3	teeth	4	0	68	0
4	impression	0	0	62	1
5	sense	2	0	57	0
6	starts	0	0	59	0
7	consciousness	0	0	57	0
8	start	0	0	50	0
9	alarms	0	0	48	0
10	statement	0	0	48	0
11	name	0	1	46	0
12	light	41	1	0	0
13	chemicals	3	0	0	30
14	insemination	33	0	0	0
15	fibres	7	0	0	22
16	materials	3	0	0	21
17	ventilation	24	0	0	0
18	pitch	12	0	0	10
19	flowers	15	3	1	1
20	selection	18	0	0	0
21	lighting	17	0	0	0
22	limbs	16	0	0	0
23	tan	0	13	2	0
24	fur	2	10	0	2
25	oligonucleotides	0	0	0	14

26	material	0	0	0	11
27	fabrics	1	0	0	9
28	pearls	3	4	3	0
29	polymers	1	0	0	9
30	drugs	0	0	0	9
31	accent	2	5	1	0
32	brushes	0	0	0	8
33	goods	0	6	1	0
34	notes	0	4	2	0
35	bombs	0	5	0	0
36	permits	0	4	0	0
37	bomb	0	3	0	0

Table 1: Input for correspondence analysis: contingency table with raw frequencies of the adjectives and their collocating nouns

	collocate\adjective	artificial	fake	false	synthetic
1	intelligence	9	0	0	0
2	pretences	0	0	10	0
3	teeth	4	0	8	0
4	impression	0	0	8	3
5	sense	2	0	6	0
6	starts	0	0	8	0
7	consciousness	0	0	8	0
8	start	0	0	6	0
9	alarms	0	0	9	0
10	statement	0	0	7	0
11	name	0	0	6	0
12	light	6	1	0	0
13	chemicals	5	0	0	9
14	insemination	9	0	0	0
15	fibres	6	0	0	9
16	materials	4	0	0	7
17	ventilation	8	0	0	0
18	pitch	6	0	0	7
19	flowers	6	4	2	3
20	selection	6	0	0	0
21	lighting	7	0	0	0
22	limbs	8	0	0	0
23	tan	0	9	4	0
24	fur	4	8	0	5
25	oligonucleotides	0	0	0	9
26	material	0	0	0	5

27	fabrics	4	0	0	8
28	pearls	5	7	5	0
29	polymers	4	0	0	8
30	drugs	0	0	0	6
31	accent	4	7	3	0
32	brushes	0	0	0	8
33	goods	0	4	1	0
34	notes	0	4	3	0
35	bombs	0	7	0	0
36	permits	0	7	0	0
37	bomb	0	5	0	0

Table 2: Input for collocational network: contingency table with (logDice) scores of collocation strength

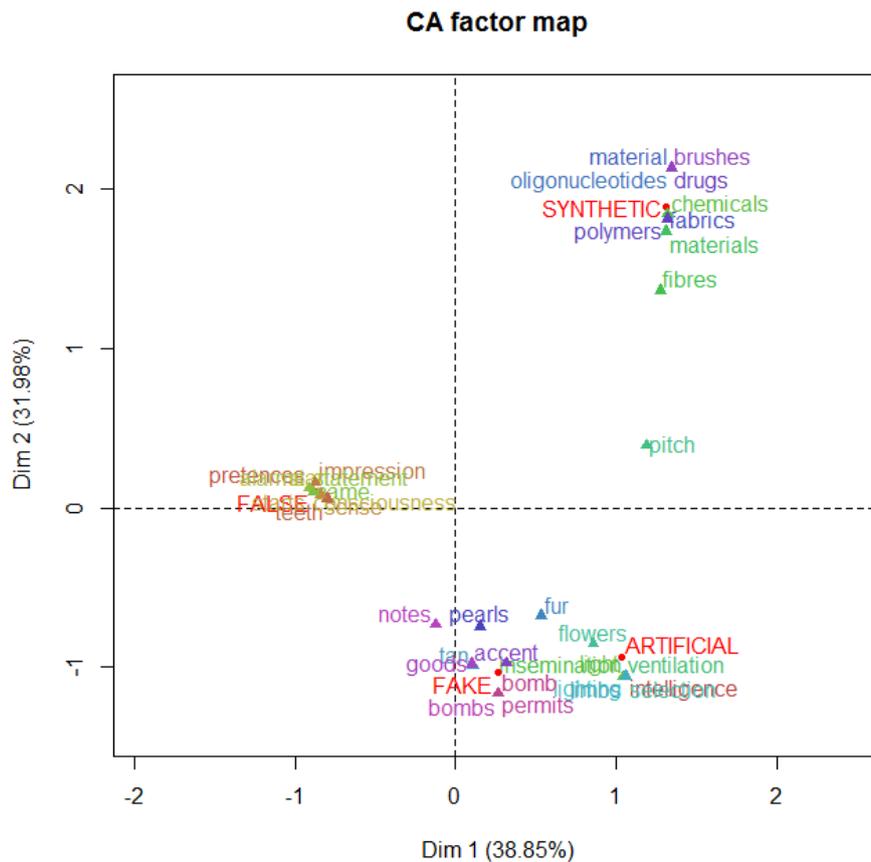


Figure 1: Correspondence analysis plot for the adjectives and their collocates⁷

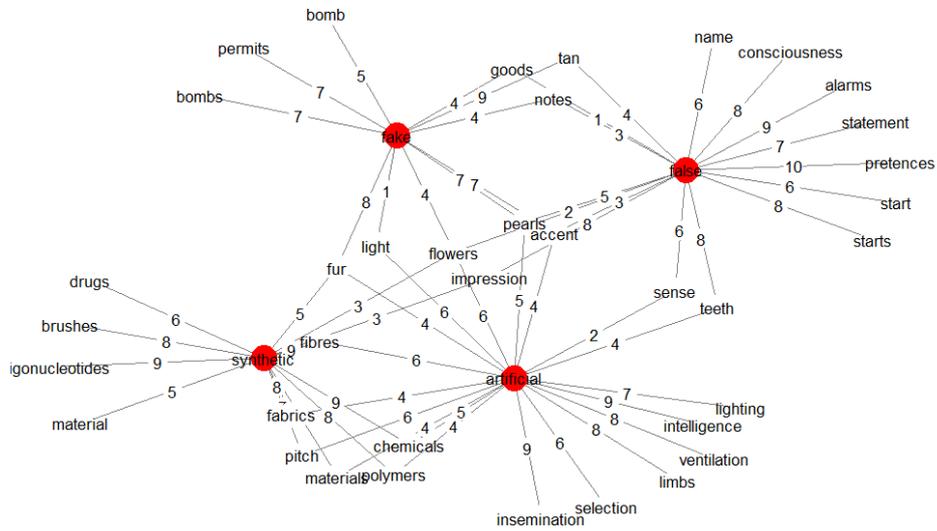


Figure 2: A collocational network with logDice scores

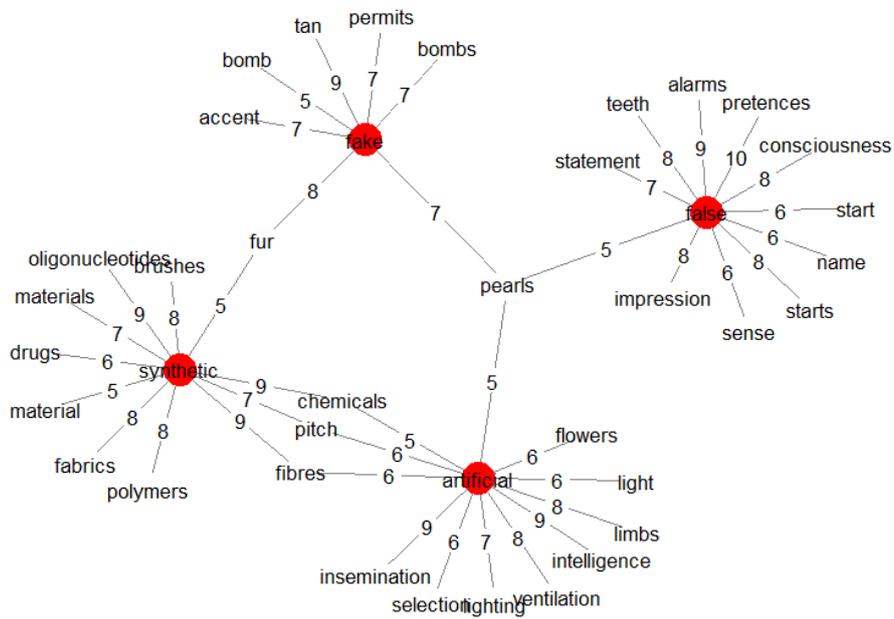


Figure 3: A collocational network with logDice higher than 4

5. Results and discussion

The interpretation of the diagrams in Fig. 1 and 2 is straightforward. In CA plot, the data points that appear in proximity to each other show strong association, while those that are far apart are unrelated. The association is stronger for points located away from than close to the centre. The plot shows that the collocates cluster around their respective adjectives with which they are distinctly associated. This is especially noticeably for *synthetic* and *false*. As for adjectives *artificial* and *fake*, they are relatively close to each other, meaning that they share certain collocates, in particular *accent*, *pearls*, *fur*, and *flowers*. Given this fact, the adjectives are more similar to each other than to other adjectives. Apart from the shared collocates, there are nouns distinctly associated with either *artificial* or *fake*, for example (*artificial*) *selection*, *intelligence*, and (*fake*) *goods*, *bomb*, *permits*, *tan*, though it is not made explicit in this plot.

The position of the data points relative to the horizontal and vertical dimensions is also meaningful. Because the adjectives *synthetic* and *artificial* are placed in the same half of the plot, quite away from the centre, they have certain collocates in common. One of them is *pitch*, which is located half way between *artificial* and *synthetic*, meaning that it enters into a strong collocation with both of the adjectives.

When it comes to the collocational network in Fig. 2 and Fig. 3, one can see that each adjective enters into distinct collocations. In particular, the collocates which are not shared by any other adjective may be indicative of semantic differences between the adjectives. For example, *fake* makes collocation with *bombs*, *permits*, and *accent*, which may suggest the idea of tricking as a semantic feature of *fake*.

Fig. 2 also shows that the adjectives are highly interwoven. Worth noticing is the collocation overlap between *synthetic* and *artificial*, which is partly visible in CA plot (note the position of *pitch* mentioned earlier). The overlapping nouns such as *pitch*, *materials*, *polymers*, *fibres*, *fabrics*, and *chemicals* strongly correlate with *synthetic* and somewhat less so with *artificial*. This finding suggests that *synthetic* and *artificial* are semantically close in the sense when they are used to describe materials and substances produced in an artificial rather than natural way.

A good strategy in interpreting collocational overlap in Fig. 2 is to compare the collocation scores for a noun that enters into collocation with more than one adjective. For example, *fur* collocates with *fake*, *synthetic* and *artificial*, but the strongest collocation is made with *fake* (8 vs. 5 and 4). Another example can be *light*, which collocates with *artificial* with a higher collocation strength than with *fake* (6 vs. 1). By comparing the scores, one can identify the strongest collocations and consider them for inclusion in a dictionary. To obtain a clearer picture, it is possible to remove the collocations with lower collocation scores, as in Fig. 3.

Based on collocation evidence in Figures 1-3, one can draw tentative con-

clusions about semantic preferences of each adjective. An adjective *artificial* is used in the context of man-made things that have the same features as their natural counterparts (e.g. *intelligence, light, respiration, ventilation*). In this regard, the use of *artificial* overlaps with that of *synthetic* (*synthetic fibres*), as well as of *false* (*false teeth*). In turn, a word *synthetic* collocates with the words denoting substances and materials specific to the field of chemistry and biology (e.g. *chemicals, oligonucleotides, polymers*), which are produced in an artificial or unnatural way. With regard to *false*, the word carries the core meaning of being "untrue" or "wrong" in the collocation *false statement*. However, the word has other meanings which are more or less related to the core one, such as *false pretences*, which is semantically more distant to the core meaning than *false impression*. The collocation with *teeth*, which is also common to *artificial*, suggests that *false* is a good candidate for synonymy with *artificial*. Given the above collocational preferences, *artificial* is semantically more similar to *synthetic* than to *false*, and these two adjectives are semantically simpler than polysemous *false*. Finally, *fake* is used to describe valuable objects (*fur*) or is used to deceive people (*permit, bomb*). The feature of deceiving is also present in some collocations of *false* (*notes, statement*).

On the whole, the results of correspondence analysis overlap with those of collocational network, but certain points of difference emerge. CA reduces the number of dimensions to just two, which allows one to spot major differences and parallels in the distribution of the data. As a result, it highlights strong collocations and obscures the weak ones. For example, while the noun *flowers* is shared by all the adjectives in the network, in the CA plot it appears close to just *artificial*, on the grounds that there is a much higher frequency of *artificial flowers* than *fake flowers, synthetic flowers, or false flowers* (see Table 1). This fact is left implicit in the network, in which *flowers* is linked with all the adjectives. Likewise, *teeth* is linked with both *false* and *artificial* in the network, but it correlates with just *false* in the CA plot. Thus, the network tends to obscure strong collocations in favour of weak ones, treating all of them as equal. In turn, an advantage of CA is that it highlights the collocations that contribute significantly to the variation of the data and disregards the less distinctive co-occurrences.

The analysis of visualisations of collocational preferences is a starting point for a detailed investigation of meaning differences through the analysis of a wider context of word use.

6. Conclusions

Correspondence analysis plot and collocational network have their strengths and weaknesses. They allow researchers to spot similarities and differences in collocational preferences of several words in a single diagram. They help lexicographers deal with information overload which arises while comparing more than two words. Using the diagrams, one can identify which words are con-

textually similar to or different from each other and see what collocational patterns underlie the semantics of the words. The visualisations may be used as an alternative way to the presentation of data in tabular form. They save a great deal of time, as they show what one can discover by perusing thousands of concordance lines or by analysing summaries of collocational behaviour generated by Sketch Engine for each pair of adjectives.

Because the visualisations are based exclusively on collocational behaviour, they do not provide lexicographers with ready-made solutions to such semantic details as sense division. The semantic boundaries, however, can be discovered through an in-depth analysis of the diagrams in combination with the frequency tables and concordance lines. Such an analysis would be enhanced by making the diagrams interactive, that is by associating hyperlinks with data points on the plot to enable instant access to corpus evidence. Finally, the visualisations are potentially useful as long as they are restricted to a subset of data. Visualising a large number of data points may result in the diagram being illegible.

Endnotes

1. In contemporary general dictionaries, synonym discriminations are provided in the middle matter. Sometimes brief discriminations supplement definitions, together with illustrative examples. For example, in an EFL *MacMillan Dictionary Online*, synonyms for *false* are provided in a panel titled "Synonyms: false", with a definition, collocations or an illustrative example for each synonym.
2. Synonym discrimination was a feature of early synonym dictionaries such as Trusler's *The Difference, Between Words, Esteemed Synonymous, in the English Language* (1766) and Piozzi's *British Synonymy: or An Attempt at Regulating the Choice of Words in Familiar Conversation* (1794) (Landau 2001: 135). These dictionaries attempted to discriminate between related words by discussing differences in meaning, and sometimes indicating collocations. Apparently for practical reasons, this practice was soon abandoned by Roget and his followers, who limited themselves to enumerating related words without specifying what distinguished them (Hartmann 2005: 83). Roget's *Thesaurus of English Words and Phrases* (1852) groups words together to the effect that the user understands the meaning of the headword through these words. However, like in many contemporary thesauri, the choice of synonyms is highly dependent on the compiler's subjective judgement (see McArthur 1986, Piotrowski 1994). This intuition-based approach leaves a great deal to be desired, as there is little agreement among lexicographers with regard to which synonyms should be recorded in an entry (Church et al. 1994: 156).
3. Distributional Thesaurus, which is another Sketch Engine tool, is used for automatic generation of a list of synonyms for a given word (Kilgarriff et al. 2004). The list is ranked according to a similarity score, and visualised in the form of a word cloud, with most similar words in the centre and others in the periphery. The similarity score is computed on the basis of a distributional criterion, by comparing quantitatively grammatical and collocational behaviour of the words.
4. According to Landau (2001: 137), the fact that dictionaries of cumulative synonymy are easier to prepare than dictionaries of synonym discrimination is a reason why publishers are more

willing to compile the former type of reference works. However, as corpus processing tools allow lexicographers for efficient exploitation of texts, contemporary dictionaries address the problem of synonym discrimination in a more principled way. For example, Urdang's *Oxford Thesaurus* (1991) divides synonyms into groups according to whether they are "more or less substitutable for the main entry in the illustrative sentence".

5. This is an association measure which was proposed by Rychlý (2008). It is based on the frequency of collocates, nodes, and the collocational pair. The theoretical maximum of the score is 14, but it is usually less than 10 (*ibid.*). In the network shown in Fig. 2, values equal to or lower than 0 are omitted, as these indicate little or no significance of collocation.
6. Data cited herein have been extracted from the British National Corpus, distributed by the University of Oxford on behalf of the BNC Consortium. All rights in the texts cited are reserved.
7. To obtain the CA plot, I used R package FactoMineR designed by Lê et al. (2008), available from <https://cran.r-project.org/web/packages/FactoMineR/index.html>.

References

- Alonso, A., C. Millon and G. Williams. 2011. Collocational Networks and their Application to an e-Advanced Learner's Dictionary of Verbs in Science (DicSci). Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011, Bled, 10–12 November 2011*: 12-22. Ljubljana: Trojina, Institute for Applied Slovene Studies.
- Atkins, B.T. Sue and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- The British National Corpus. Version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on Behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>.
- Bussmann, H. (Ed.). 1996. *Routledge Dictionary of Language and Linguistics*. (Translated and edited by Trauth, G. and K. Kazzazi). London/New York: Routledge.
- Chen, D. and C.D. Manning. 2014. A Fast and Accurate Dependency Parser Using Neural Networks. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*: 740–750. Doha, Qatar: Association for Computational Linguistics.
- Church, K.W., W. Gale, P. Hanks, D. Hindle and R. Moon. 1994. Lexical Substitutability. Atkins, B.T.S. and A. Zampolli (Eds.). 1994. *Computational Approaches to the Lexicon*: 153-177. Oxford: Clarendon Press.
- Church, K. and P. Hanks. 1989. Word Association Norms, Mutual Information and Lexicography. *ACL Proceedings, 27th Annual Meeting, Vancouver, Canada, June 26–29, 1989*: 76-83. Stroudsburg, PA: Association for Computational Linguistics.
- Crystal, D. 2003. *A Dictionary of Linguistics and Phonetics*. Fifth edition. Oxford: Blackwell Publishing.
- Desagulier, G. 2014. Visualizing Distances in a Set of Near-Synonyms: *Rather, Quite, Fairly, and Pretty*. Glynn, D. and J.A. Robinson (Eds.) 2014. *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*: 145-178. Amsterdam: John Benjamins.
- Firth, J.R. 1957. *Papers in Linguistics*. Oxford: Oxford University Press.
- Firth, J. 1968. A Synopsis of Linguistic Theory 1930–1955. Palmer, F.R. (Ed.). 1968. *Selected Papers of J.R. Firth 1952–59*: 1-32. Bloomington: Indiana University Press.

- Glynn, D.** 2014. Correspondence Analysis: Exploring Data and Identifying Patterns. Glynn, D. and J. Robinson (Eds.). 2014. *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*: 443-486. Amsterdam: John Benjamins.
- Greenacre, M.** 2007. *Correspondence Analysis in Practice*. Second edition. London: Chapman & Hall. DOI: 10.1201/9781420011234.
- Hartmann, R.R.K.** 2005. Lexicography and Its Interdisciplinary Contacts, with Special Reference to Linguistics and Onomasiology. *Lexikos* 15: 70-89.
- Kamiński, M.** 2015. In Search of Lexical Discriminators of Definition Style: Comparing Dictionaries through n-Grams. *International Journal of Lexicography* 29(4): 403-423.
- Kilgarriff, A.** 2013. Using Corpora as Data Sources for Dictionaries. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 77-96. London: Bloomsbury.
- Kilgarriff, A. and I. Kosem.** 2012. Corpus Tools for Lexicographers. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 31-55. Oxford: Oxford University Press.
- Kilgarriff, A., P. Rychlý, P. Smrz and D. Tugwell.** 2004. The Sketch Engine. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 105-116. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- Kilgarriff, A. and D. Tugwell.** 2002. Sketching Words. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins*: 125-137. Grenoble: EURALEX.
- Krishnamurthy, R.** 2000. Collocation: From Silly Ass to Lexical Sets. Heffer, C., H. Sauntson and G. Fox (Eds.). 2000. *Words in Context: A Tribute to John Sinclair on his Retirement*: 31-47. Birmingham: University of Birmingham.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York/Cambridge: Cambridge University Press.
- Lê, S., J. Josse and F. Husson.** 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* 25(1): 1-18. 10.18637/jss.v025.i01.
- Lewandowska-Tomaszczyk, B.** 1990. Meaning, Synonymy, and the Dictionary. Tomaszczyk, J. and B. Lewandowska-Tomaszczyk (Eds.). 1990. *Meaning and Lexicography*: 181-208. Amsterdam: John Benjamins.
- Macmillan Dictionary.** 2009–2015. Macmillan Publishers Limited. The Entry for *False*. June 2015. <http://www.macmillandictionary.com>.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer*. Cambridge: Cambridge University Press.
- Moon, R.** 2013. Braving Synonymy: From Data to Dictionary. *International Journal of Lexicography* 26(3): 260-278.
- Piotrowski, T.** 1994. *Z Zagadnień Leksykografii*. Warszawa: Wydawnictwo Naukowe PWN.
- Piozzi, H.L.** 1794. *British Synonymy: or An Attempt at Regulating the Choice of Words in Familiar Conversation*. London: Printed for G.G. and J. Robinson.
- Polguère, A.** 2014. From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography* 27(4): 396-418.
- Roget, P.M.** 1852. *Thesaurus of English Words and Phrases*. New York: Longmans, Green & Co./Dell Publishing.

- Rychlý, P.** 2008. A Lexicographer-friendly Association Score. Sojka, P. and A. Horák (Eds.). 2008. *Proceedings of Second Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2008*: 6-9. Brno: Masaryk University.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Trusler, J.** 1766. *The Difference, Between Words, Esteemed Synonymous, in the English Language*. London: J. Dodsley. January 2017. <https://archive.org/details/differencebetwe00giragoog>.
- Urdang, L.** 1991. *The Oxford Thesaurus. An A-Z Dictionary of Synonyms*. Oxford: Oxford University Press.
- Xiao, R. and T. McEnery.** 2006. Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics* 27(1): 103-129.

Intellectualization through Terminology Development*

Langa Khumalo, *Linguistics Program, School of Arts,
University of KwaZulu-Natal, South Africa*
(khumalol@ukzn.ac.za)

Abstract: The term intellectualization was famously used in the Prague School to describe a process that a language undergoes in its development and refinement. In our South African context intellectualization entails a carefully planned process of hastening the cultivation and growth of indigenous official African languages so that they effectively function in all higher domains as languages of teaching and learning, research, science and technology. This article critically examines the terminology development process that is being driven at the University of KwaZulu-Natal (henceforth UKZN) as one of the key agents of language intellectualization. The article critically evaluates the UKZN terminology development model that is used to harvest, consult and authenticate isiZulu terminology for Administration, Architecture, Anatomy, Computer Science, Environmental Science, Law, Physics, Psychology, and Nursing disciplines. Outflow platforms for the terminology in this development model are loosely listed as the 'database' and the 'development platform' but there is no clear end-user platform for students and lecturers, who seem to be the main end-user-targets of the whole terminology development initiative. The article will propose an improved model to cater for AnyTime Access, which is convenient for student needs between lectures, and improve the harvesting mechanism in the existing model.

Keywords: INTELLECTUALIZATION, TERMINOLOGY DEVELOPMENT, HARVESTING, CROWDSOURCING, CONSULTATION, VERIFICATION, AUTHENTICATION, ANYTIME ACCESS

Opsomming: Intellektualisering deur terminologieontwikkeling. Die term intellektualisering is oorspronklik deur die Praagse skool gebruik om 'n proses wat 'n taal tydens sy ontwikkeling en verfyning ondergaan te beskryf. In ons Suid-Afrikaanse konteks behels intellektualisering 'n sorgvuldig beplande proses om die ontwikkeling en groei van plaaslike amptelike Afrikatale te bespoedig sodat hulle effektief op alle hoëvlakterreine funksioneer as tale van onder- en leer, navorsing, wetenskap en tegnologie. Hierdie artikel doen 'n kritiese ondersoek na die terminologieontwikkelingsproses wat by die Universiteit van KwaZulu-Natal (voortaan UKZN) bestuur word as een van die sleutelfigure van taalintellektualisering. UKZN se terminologieontwikkelingsmodel word gebruik om isiZulu-terminologie binne die dissiplines Administrasie, Argitektuur, Anatomie, Rekenaarwetenskap, Omgewingswetenskap, die Regte, Fisika, Sielkunde en Verpleegkunde te oes, te raadpleeg en te staaf. Die artikel evalueer hierdie model krities. Afvoer-

* 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.

platforms vir die terminologie in hierdie ontwikkelingsmodel word losweg gelys as die 'databasis' en die 'ontwikkelingsplatform', maar daar is geen duidelike eindgebruikerplatform vir studente en dosente nie, wat waarskynlik die belangrikste eindgebruikerteikengroep van die hele terminologie-ontwikkelingsinisiatief vorm. Die artikel stel 'n verbeterde model voor wat deurlopende toegang sal bied, wat gerieflik vir studente se behoeftes tussen lesings sal wees, en wat die oesmeganisme in die bestaande model sal verbeter.

Sleutelwoorde: INTELLEKTUALISERING, TERMINOLOGIEONTWIKKELING, OES, SKARE-DEELNAME, KONSULTASIE, VERIFIKASIE, STAWING, DEURLOPENDE TOEGANG

1. Introduction

It can be argued that South African Higher Education (henceforth SAHE) is increasingly embracing the imperative of transmuting African languages to be the kernel of the academy. This has been recently articulated in the #FeesMustFall initiative (cf. Chetty and Knaus 2016). Hitherto the dominant languages in SAHE have been English and Afrikaans. It has been persuasively argued that the high attrition rate in Africa's education system is in part due to the fact that Africa is uniquely one of the few continents where children receive education in foreign tongues (cf. Finlayson and Madiba 2002). It is in this vein that the UKZN language policy and plan (2006 revised 2014) recognizes the prominent role that language plays in the teaching and learning. Through its language policy and plan, UKZN has taken an initiative to cultivate, modernize and elaborate isiZulu so that it becomes a vehicle in knowledge production and knowledge dissemination. One of the stated aims of the UKZN language policy is to "*achieve for isiZulu the institutional and academic status of English*" and to "*provide facilities to enable the use of isiZulu as a language of learning, instruction, research and administration.*" (Language Policy of the UKZN 2014: 2). It is in this regard that, while UKZN upholds the principle of multilingualism, English and isiZulu are the two official languages of the institution. It has thus been persuasively argued that in order for African languages to be used in education as languages of instruction, innovation, science, mathematics and logic, there has to be a clear, conscious and careful process of the intellectualization of these languages.

Through its language policy and plan, UKZN articulates a clear objective to remove language as a barrier of access and success. The notion of *access and success* is articulated in the 1994 UNESCO World Conference framework of action where the following was noted: that our education systems should

[...] accommodate all children regardless of their physical, intellectual, social, linguistic or other conditions. This should include disabled and gifted children, street and working children, children from remote or nomadic populations, children from linguistic, ethnic, or cultural minorities and children from other disadvantaged or marginalized area and groups. (UNESCO, 1994, Framework for Action on Special Needs Education: 6)

SAHE is thus encouraged to:

[...] recognize and respond to the diverse needs of their students, accommodating both different styles of learning and ensuring quality education to all through appropriate curricula, organizational arrangements, teaching strategies, resource use and partnerships with their communities. (UNESCO, 1994, Framework for Action on Special Needs Education: 11-12)

Hence UKZN's strategy to cultivate isiZulu so that it becomes a language of administration and all academic activity in order to eliminate problems of language being a barrier to access and success. In order to achieve this, UKZN has committed huge financial and human resources. This article thus critically examines the UKZN's language program. The article locates terminology development as one of the ways through which the University seeks to intellectualize isiZulu. The article further evaluates the limitations of the University's current terminology development model, and proposes an improved model to cater for the novel notion of AnyTime Access that we motivate here in order to satisfy end-user-needs.

2. Intellectualization

Intellectualization is a term originally used by Havranek (1932), a linguist from the Prague School, to characterize a process that a language undergoes in its advancement.

By the intellectualization of the standard language, which we could also call its *rationalization*, we understand its adaptation to the goal of making possible precise and rigorous, if necessary abstract, statements, capable of expressing the *continuity* and *complexity of thought*, that is, to reinforce the intellectual side of speech. This intellectualization culminates in *scientific* (theoretical) *speech*, determined by the attempt to be as *precise in expression* as possible, to make statements which reflect the rigor of *objective* (scientific) *thinking* in which the terms approximate concepts and the sentences approximate *logical judgements*. (My emphasis). (Havránek 1932: 32-84)

Intellectualization thus is a clear process of (functionally) cultivating, developing, elaborating and modernizing a language so that the terminology of the language can carry the full weight of scientific rigor and precision, and that its sentences can accurately express logical judgements resulting in a language that has the capacity to function in all domains. As the direct consequence of intellectualization the speakers of the language derive the pride, self-assurance and resourcefulness in the (new) ability to discuss the most complex of issues ranging from the mundane to academic and beyond.

Intellectualization has been famously associated with the development of Tagalog in the Philippines. The cultivation process involved Tagalog's lexical enrichment through terminology to enable its use in academia. Philippine linguistics and sociolinguists are recognized by Neville Alexander in Busch,

Busch and Press (2014) as the doyens in the scholarship of intellectualization. Sibayan (1999: 229) characterizes an intellectualized language as one:

[...] which can be used for educating a person in any field of knowledge from kindergarten to the university and beyond.

Thus, an intellectualized language has the capacity to discuss any issue regardless of its complexity. According to Finlayson and Madiba (2002: 53), in the South African context intellectualization is a meticulous procedure aimed at expediting the growth and development of hitherto underdeveloped African languages to augment their capacity to effectively interface with modern developments, theories and concepts. It is imperative to note that germane to this process is the development of discipline specific terminology. The paucity of such specialized terminology is often cited as the reason why African languages cannot be used as languages of teaching and learning, hence their discernment as shallow and inadequate (cf. Shizha 2012).

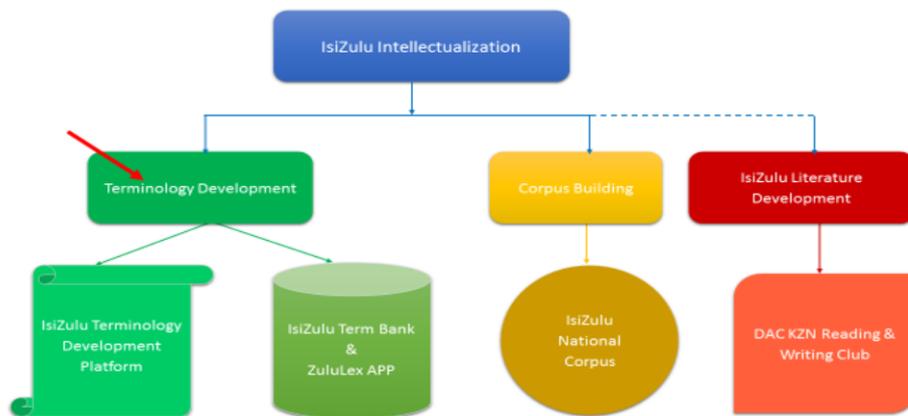
Intellectualization in our context thus means the radical transformation of the capacity and role of indigenous African languages in carrying and conveying all forms of knowledge in all spheres of life. While the government through the Constitution (1996, section 6) has expressed commitment "*to elevate the status and advance the use of*" these hitherto underdeveloped languages, very little has actually been done to improve their status and role in SAHE (cf. Olivier 2014). The argument on the status and role of these languages has been sharply brought back to the center of SAHE through the #FeesMustFall. Writing on the #FeesMustFall, Chetty and Knaus (2016) posit that at the center of the 2015 student protests was the disruption of the status quo with regards to "[...] institutions' language policies, high fees, structural inequalities and colonial symbols."

3. UKZN Language Program

UKZN has a University Language Board (henceforth ULB), which was set up through a committee charter (2006 amended 2010). The ULB meets quarterly and its mandate is to develop, implement, monitor and review the University Language Policy and Plan. The ULB is required to report to the University Senate once a year, and importantly its annual report to Senate must be in both English and isiZulu. It is chaired by a Deputy Vice-Chancellor (henceforth DVC), who is also a member of the University's Executive. As reiterated above, UKZN first adopted its Language Policy and Plan in 2006, which was subsequently revised in 2014 following the University's college reorganization and the introduction of the Use of Official Languages Act of 2012. In its strategy to institutionalize language planning and development, UKZN established the University Language Planning and Development Office (henceforth ULPDO), which is headed by a Director who is a member of the ULB, and reports to the DVC.

The program to intellectualize isiZulu so that it (ultimately) functions at par with English in all high function domains across the University is initiated by the ULPDO, and is reported quarterly at the ULB. The major thrust of the language program (see Fig. 1.) as approved by the ULB is the creation of discipline specific terminology in isiZulu, the building of an isiZulu National Corpus, and the development of a contemporary body of literature in isiZulu. Other language activities include the provision of training workshops, translation and (simultaneous) interpreting services, Sign Language advocacy, the Sesotho *Bua Le Nna* (Let's Talk) Program, language research, and the development of computational tools.

Figure 1: Language Program



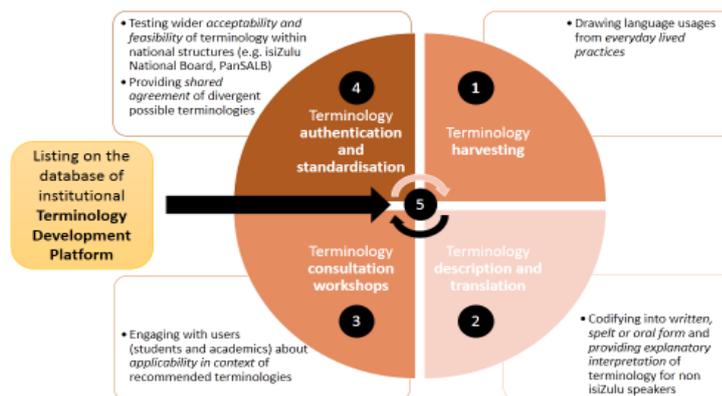
The article critically evaluates the UKZN terminology development model (see Fig. 2.) used to harvest, consult and authenticate isiZulu terminology for Administration, Architecture, Anatomy, Computer Science, Environmental Science, Law, Physics, Psychology, and Nursing to date. In this model, outflow platforms for the authenticated terminology are loosely indicated as the 'database' and 'development platform'. There is no clear end-user platform for students in the model, yet the students seem to be the main end-user-target of the whole terminology development initiative. Other important end-users are lecturers, teachers, lexicographers, educational interpreters and translators, isiZulu speakers, and authors. The article will propose an improved model to cater for AnyTime Access, which is convenient particularly for student needs between lectures, and propose an improved harvesting mechanism for the existing model. We critically examine the terminology development model in section 4, discuss in section 5 and conclude in section 6.

4. Terminology Development Model

UKZN has identified terminology development as one of the main cogs (together with corpus building and computer assisted language tools) to drive the goal of isiZulu intellectualization. One of the important principles in the development of specialized terminology at UKZN is to observe statutory mandatory processes. These processes are driven by the Pan South African Language Board (henceforth PanSALB), which is a Board constituted through an Act of Parliament. Through the provisions of the PanSALB Act 59, (1999), there exists an isiZulu National Language Body, UMZUKAZWE, which is short for *UMkhandlu WesiZulu Kuzwelonke*.

In light of these provisions, UKZN has designed a unique terminology development process with five important stages that incorporate the statutory processes facilitated by PanSALB through its KwaZulu-Natal Provincial office. There are five stages in the UKZN terminology development model and these include 1) *harvesting* of existing usage terms; 2) *description and translation* of terminology that has been harvested or created; 3) *consultation and verification* with end-users about terminology proposed, and 4) *authentication and standardization through official national structures*. The "finalization" of the process in the current model takes place in 5) through the listing of these terms on *the terminology databases* for wider institutional and national usage. The five stages are illustrated in Figure 2 below.

Figure 2: UKZN Terminology Development Model (cf. Khumalo 2016)



The current terminology development model as shown in Fig. 2 relies on the discipline lecturers, as experts in their fields, to initiate the harvesting of terms that are central to the module or course that they lecture in. While the UKZN

language policy exists as an instrument that enables the development of teaching materials in both English and isiZulu, the enforcement of the policy is tepid, cautious and not compulsory. Hence harvesting is thus done voluntarily by discipline lecturers who are committed to the principles of the UKZN language policy, and also realize the value in making their teaching materials available to students in both languages. The harvested terms are a wordlist of key terms created from a main course/module or a major reference work. Because of the expenses involved in convening a single terminology development workshop (each workshop ca. R65 000–R95 000), the ULPDO has imposed a minimum requirement of at least 500 terms to be submitted for them to be taken through the development processes. Each workshop is a minimum of three days and can go up to five days depending on the complexity of the subject or discipline.

Once the terms are harvested and have been submitted to the ULPDO, the latter convenes a consultative workshop with a panel of at least two discipline experts, lexicographers, linguists, terminologists, and students of up to 25 individuals to describe, create and codify the isiZulu terms. This approach can be rightly criticized as top-down, elitist and exclusionary because it seems to target experts to the exclusion of general isiZulu speakers and other stakeholders who are outside the academe. A counter-argument is that terminology development is inherently a specialized task for people with certain levels of expertise for use by people working or training in a particular specialized discipline. A balance of the top-down and bottom-up approach is argued for in section 5 below.

Once the terms have been described and codified the ULPDO convenes a second workshop for the PanSALB subcommittee of isiZulu national language body to consider the terms. This is called the verification workshop. After the verification committee has thoroughly engaged academics and students in these often grueling verification workshops, the agreed terms are then submitted to the full isiZulu national language body (UMZUKAZWE) for authentication and standardization. The UMZUKAZWE is convened by the PanSALB regional office through the request of the ULPDO to consider and approve the consulted and verified terminology. Once the full UMZUKAZWE has approved the terms in the presence of the discipline experts, ULPDO notifies the ULB, and afterwards the terms are ready to be used in the lecture room and in formal academic discourses. In the current terminology development model the authenticated terms are stored in a database. It is crucial to observe that these workshops (consultative, verification and authentication) are arduous to both academics and students, and the members of each committee involved in the process. It is through sheer individual commitment and dedication by academics in each disciplines, and the members of the isiZulu language committees, that UKZN has completed in a narrow sense (given that terminology development is a continuous process) the terminology development processes in the following disciplines: Administration, Architecture, Anatomy, Computer

Science, Environmental Science, Law, Nursing Science, Physics, Psychology, and Research.

The terminology development model is cyclical. Once a discipline has gone through all the processes, ULPDO moves on to the term list submitted by the next discipline experts and the process is repeated. It is clear that the process requires a huge investment both in terms of time and financial resources.

5. Discussion

In this section we critically evaluate the terminology development model. The current model relies on the discipline lecturers to initiate harvesting. This harvesting work is done over and above the discipline lecturer's other teaching, research, and administrative commitments. Those who have done so have demonstrated outstanding commitment. However the harvesting has become slow and erratic. This is attributable to the University's Teaching Workloads Framework (used to assess teaching) and Performance Measurers (used to assess research, among other things). According to the Teaching Workloads Framework the norm is that a lecturer must have at least 810 hours of teaching. The approved performance measures are 60 Productivity Units (henceforth PU) for lecturers, 90 PUs for senior lecturers, 120 PUs for associate professors and 150 PUs for professors per year. A single publication in a Department of Higher Education and Training (DHET) approved journal carries 60 PUs. This means a lecturer has to publish at least a single journal article per annum, while a professor has to publish two and a half articles in the same period. There are indeed some lecturers who have abandoned terminology development for their disciplines citing huge workloads, while expressing the obvious need for isiZulu terminology as a useful intervention to improve conceptual and epistemic access.

To mitigate erratic terminology harvesting, and the effects of a clearly top-down and exclusionist approach to terminology development, ULPDO is introducing (computational) technology in term harvesting. Crowdsourcing has been advanced as a useful strategy to harness discipline specific terminology. This argument was advanced when ULPDO was developing isiZulu terminology in computer science. The two discipline experts, Dr Maria Keet and Dr Graham Barbour created a novel method (cf. <http://www.commuterm.co.za/>) of harnessing terms in computer science using computational resources (cf. Keet and Barbour 2014). This is a useful strategy to improve the collection of terminology, which can also be further extended to include online commentary in order to augment the role of the consultative workshop, with a view of replacing it in the long term in order to reduce costs.

The current UKZN terminology development model can be criticized for being a top-down approach to language planning. Language planning is often viewed as a top-down process, which is characterized by Cooper (1989: 45), and Grin (1996: 31) as a considered, planned, systematic effort to influence and

affect the language behavior of a particular society through crafting and implementing regulations or policies, thus regulating a speech community's language behavior. According to Webb (2002: 42) for language planning to be effective, top-down processes need to be complemented by a bottom-up approach, which means that the interest, views, attitudes and linguistic competences of the targeted speech community must be accommodated.

In explicating the bottom-up approach, Kaplan and Baldauf (1997: 50), distinguish between activities of governments and agencies and activities of pressure groups and individuals in the language planning process (cf. Sithole 2017). According to these distinctions, a bottom-up approach to language planning refers to the language cultivation activities of individual and pressure group agencies. We argue in this article that a judicious mixture of the top-down and bottom-up approach may render a better outcome for the UKZN terminology development model.

Figure 3: Zulu Lexicon Screenshot



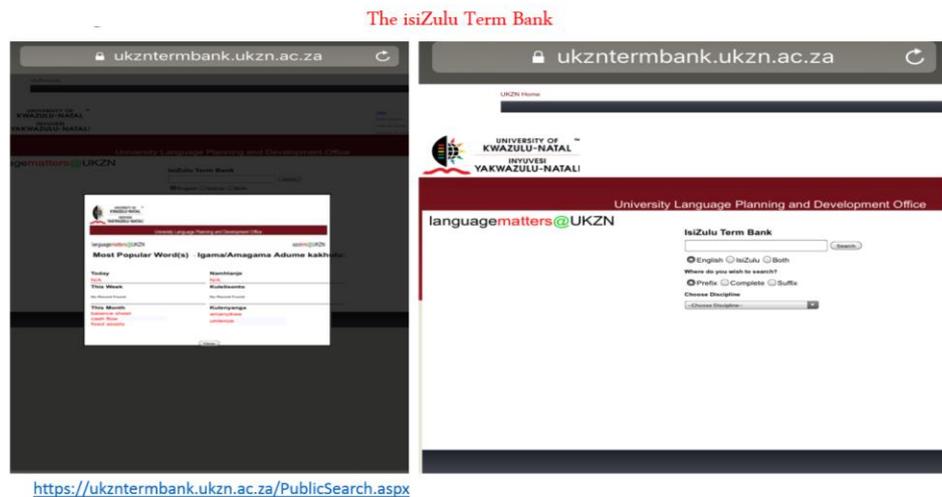
In order to also improve access to the authenticated and finalized terminology we propose a clearer end-user platform(s) for students, lecturers, teachers, lexicographers, educational interpreters and translators, and isiZulu speakers. Hitherto a lot of terminology has been developed by many universities in the country, the National Language Services in the Department of Arts and Culture (DAC), and the South Africa Norway Tertiary Education Development Program (SANTED) multilingual glossary project, among others. Terminology

developed is stored away in *some inaccessible databases* without any dissemination strategy. ULPDO has developed two open-source platforms to cater for a broad-based end-user uptake of its authenticated terminology. ULPDO has also introduced the notion of AnyTime access so that students and other users can access the terms anytime they need them. ULPDO has exploited technology to develop improved end-user access through the development of a mobile compatible application. The application is called the Zulu Lexicon, which is short for *isiZulu Lexicon*, and is a free download from iTunes and Google Play in Smartphone and Android. The Zulu Lexicon affords students, lecturers, and other end-users anytime access at their fingertips to all the authenticated (multidisciplinary) terminology, with search functions on the APP in both English and isiZulu as shown in Fig. 3.

Fig. 3 shows the landing page for the Zulu Lexicon App (on the left hand side) the word-of-the-day in the discipline of anatomy in the middle, and the a-z lexicon with random entries under the letter I of the Roman alphabet (on the far right). The APP has other filters to make searching very easy, and user-friendly for users to find the kind of information they want.

The second open source terminology resource is the isiZulu Term Bank. ULPDO developed the isiZulu Term Bank (ITB) as an open source database that is freely accessible on the following URL: <https://ukzntermbank.ukzn.ac.za>. It is a flexible system that can be incrementally populated with more authenticated terms for existing disciplines that are already in the database or with new disciplines whose terms have been developed and authenticated. Figure 4 below shows the ITB screenshot.

Figure 4: ITB Screenshot



The left screen shows the landing page for the isiZulu Term Bank. The landing page will flash the most popular word of the day, week and month on the screen. The screen window is closed upon clicking the close button at the bottom of flashing page. On the right side of Fig. 4 is the screen showing the home page with easy to use functionalities. The end-user can search for a term by language (English or isiZulu), word structure (prefix, suffix or complete word) or by discipline, by following a drop down menu. The Zulu Lexicon APP and the isiZulu Term Bank are two open source technology platforms that the ULPDO has developed in order to effectively disseminate the authenticated isiZulu terminology that has been developed to date in the disciplines of Administration, Architecture, Anatomy, Computer Science, Environmental Science, Law, Physics, Psychology, and Nursing Science. As open source resources, the Zulu Lexicon application and the isiZulu Term Bank are freely available to students, lecturers, teachers, lexicographers, educational interpreters and translators, isiZulu speakers, and authors.

6. Conclusion

The article has argued that intellectualization is a gradual process, which culminates in a scientific language capable of reflecting the rigor of objective thinking. While the Constitution (1996) recognizes the importance of developing the official indigenous African languages, the imperative to cultivate them for higher level functions in order to improve access and success at higher education institutions in order to arrest rampant attrition rates has been brought sharply to the fore in the recent #FEESMUSTFALL initiative. SAHE is thus embracing the importance of bringing these hitherto marginalized languages at the very center of teaching and learning. This is arguably a very length process.

The article has argued that terminology development is one of the vital cogs in the intellectualization process. The paucity of terminology is often cited as the reason why African languages cannot contribute meaningfully to the knowledge economy, and this is articulated in Bamgbose (2002: 2), when citing lack of political will by those in authority as an impediment: "[...] They are quick to point out that African languages are not yet well developed to be used in certain domains or that the standard of education is likely to fall [...]." We thus argue in this article that terminology development should not be exclusively a top-down and selective process involving workshops often with a few discipline experts, terminologists, lexicographers and linguists, which is very resource intensive, but must embrace a bottom-up approach through crowd-sourcing (cf. Keet and Barbour 2014) in cases of terminology harvesting, including usurping certain functions of the consultative workshops in order to improve efficiency and reach, enhance acceptance, and preclude exclusion while cutting back on expenditure. Thus, the UKZN terminology development process must embrace a mixed approach in which the top-down and bottom-up approaches complement each other.

It is clear that stages 1 and 2 in the current ULPDO terminology development model (term harvesting and consultation workshops) can be enhanced using computational resources. It must be argued that the discipline expert knowledge is required consistently throughout the terminology development process. In order for these experts to be able to also participate in the online platforms (and at a later stage in verification and authentication workshops), a process must be initiated by the ULPDO to submit to the ULB and ultimately to Senate an argument to formally recognize participation in the terminology development within the Teaching and Workloads Framework. This will ensure that expert standards are retained, and rebut the argument that terminology development for African languages is an exercise in dumbing down education.

The strategy to improve access to multidisciplinary authenticated terminology using a novel mobile application the Zulu Lexicon and the isiZulu Term Bank introduces to the UKZN terminology development model new end-user platforms. The multidisciplinary terms are now available anytime to students, lecturers, teachers, lexicographers, educational interpreters and translators, isiZulu speakers, and authors.

References

- Bamgbose, A.** 2002. Launch of the Activities of the African Academy of Languages: Mission and Vision of the ACALAN. *ACALAN Special Bulletin* 2002: 24-25.
- Chetty, R. and C.B. Knaus.** 2016. #FeesMustFall: SA's Universities are in the Grip of a Class Struggle. *Mail&Guardian*, 14 January 2016. Accessed on 31 July 2016.
- Cooper, R.L.** 1989. *Language Planning and Social Change*. Cambridge: Cambridge University Press.
- Finlayson, R. and M. Madiba.** 2002. The Intellectualization of the Indigenous Languages of South Africa: Challenges and Prospects. *Current Issues in Language Planning* 3(1): 40-61.
- Grin, F.** 1996. The Economics of Language: Survey, Assessment, and Prospects. Grin, F. (Ed.). 1996. *Economic Approaches to Language and Language Planning. International Journal of the Sociology of Language* 121: 17-44.
- Havranek, B.** 1932. The Functions of Literary Language and its Cultivation. Havranek, B and M. Weingart (Eds.). 1932. *A Prague School Reader on Esthetics, Literary Structure and Style*: 32-84. Prague: Melantrich.
- Kaplan, R.B. and R.B. Baldauf.** 1997. *Language Planning from Practice to Theory*. Clevedon: Multilingual Matters.
- Keet, C.M. and G. Barbour.** 2014. Limitations of Regular Terminology Development Practices: The Case of isiZulu Computing Terminology. *Alternation: Special Edition* 12: 13-48.
- Khumalo, L.** 2016. Disrupting Language Hegemony: Intellectualizing African Languages. Samuel, M., R. Dhunpath and N. Amin (Eds.). 2016. *Disrupting Higher Education Curriculum. Undoing Cognitive Damage*: 247-264. Rotterdam/Boston/Taipei: Sense Publishers.
- Olivier, J.** 2014. Compulsory African Languages in Tertiary Education: Prejudices from News Website Commentary. *Southern African Linguistics and Applied Language Studies* 32(4): 483-498.
- Shizha, E.** 2012. Reclaiming and Re-visioning Indigenous Voices: The Case of the Language of Instruction in Science Education in Zimbabwean Primary Schools. *Literacy Information and Computer Education Journal (LICEJ), Special Issue* 1(1): 785-793.

- Sithole, E.** 2017. *From Dialect to 'Official' Language: Towards the Intellectualization of Ndaue in Zimbabwe*. Unpublished D.Phil. Thesis. Grahamstown: Rhodes University.
- UNESCO.** 1994. *The Salamanca Statement and Framework for Action on Special Needs Education*. Paris: UNESCO.
- Webb, V.N.** 2002. *Language in South Africa: The Role of Language in National Transformation, Reconstruction and Development*. Amsterdam/Philadelphia, Penn: John Benjamins.

"Minimum Input, Maximum Output, Indeed!" Teaching Collocations Through Collocation Dictionary Skills Development

Sugene Kim, *Department of English Studies, Nagoya University of
Commerce & Business, Nisshin, Japan (sugene_kim@nucba.ac.jp)*

Abstract: This study examined the teachability of collocations through cultivating EFL learners' collocation dictionary skills. Fifty-nine EFL college students participated in the study, and they received two 75-minute instructions between pre- and post-tests: one on the definition of collocation and its importance, and the other on the skill of looking up collocational information in the *Naver Dictionary* — an English–Korean online dictionary. During the second instruction, the students were trained to consult the dictionary for collocation production in the order of node word selection, word sense distinction, collocate type location, and feasible collocate identification. A comparison of collocation production test scores through a paired-samples *t*-test indicated that teaching collocation dictionary skills substantially improves learners' ability to produce natural collocations regardless of proficiency differences. In addition, the survey data collected at the end of the semester suggested that the participants perceive the instruction as necessary and helpful in gaining collocational competence and that their dictionary consultation behaviors have changed after receiving the instruction. Equipping EFL learners with collocation dictionary skills was also found to help them raise a sense of learner autonomy.

Keywords: COLLOCATION, COLLOCATION DICTIONARY, DICTIONARY SKILLS, EFL WRITING, L2 WRITING, COLLOCATION ERRORS, L1 INTERFERENCE, COLLOCATIONAL COMPETENCE, LEARNER AUTONOMY

Opsomming: "Sowaar minimum inset, maksimum uitset!" Die onderrig van kollokasies deur woordeboekvaardigheidsontwikkeling ten opsigte van kollokasies. Hierdie studie het die moontlikheid ondersoek om kollokasies te onderrig deur EVT-leerders se woordeboekvaardighede te ontwikkel ten opsigte van kollokasies. Nege-en-vyftig EVT-kollegestudente het aan die studie deelgeneem. Hulle het twee opleidingsessies van 75 minute elk tussen die aanvanklike en uiteindelijke toetse ontvang: een oor die definisie van die kollokasie en sy belangrikheid, en die ander oor die vaardigheid om kollokasionele inligting in die *Naver Woordeboek* — 'n Engels–Koreaanse aanlyn woordeboek — op te soek. Tydens die tweede opleidingsessie is die studente opgelei om die woordeboek te raadpleeg vir die produksie van kollokasies in die volgorde van noduswoordseleksie, betekenisonderskeiding, die opsoek van kollokasiesoorte, en die identifisering van werkbare kollokasies. 'n Vergelyking van die resultate vir kollokasieproduk-

sie deur 'n *t*-toets vir gepaarde steekproewe het getoon dat die onderrig van woordeboekvaardig-hede ten opsigte van kollokasies die leerders se vermoë om natuurlike kollokasies te produseer aansienlik verbeter het ondanks vaardigheidsverskille. Daarbenewens het die ondersoekdata wat aan die einde van die semester ingesamel is, daarop gedui dat die deelnemers die opleiding as noodsaaklik en nuttig beskou het om kollokasionele bevoegdheid te verwerf en dat hulle manier om woordeboeke te raadpleeg ná die opleiding verander het. Daar is bevind dat die toerus van EVT-leerders met woordeboekvaardighede ten opsigte van kollokasies help om 'n gevoel van self-standigheid by leerders te wek.

Sleutelwoorde: KOLLOKASIE, KOLLOKASIEWOORDEBOEK, WOORDEBOEKVAARDIGHEDE, EVT-SKRYFWERK, L2-SKRYFWERK, KOLLOKASIEFOUTE, L1-INTERFERENSIE, KOLLOKASIONELE BEVOEGDHEID, SELFSTANDIGHEID BY LEERDERS

1. Introduction

Ever since studies in the 1980s demonstrated convincingly that collocations play an important role in characterizing overall language proficiency (Kennedy 1990, Sinclair 1991, Granger 1998), the importance of collocations has been at the center of attention among both English as a second language (ESL) and a foreign language (EFL) researchers (Wray 2002, Nesselhauf 2005). Amid burgeoning empirical evidence suggesting positive effects of teaching collocations, both ESL and EFL researchers agree that collocations need to be taught systematically (Nattinger and DeCaricco 1992, Martyńska 2004).

To date, studies have been mainly interested in finding answers to "which collocations to teach based on what technique at which learning stages" (Cowie 1998, Granger and Meunier 2008). For example, a web-based project commonly referred to as "collocator" (Wible et al. 2006) intended to help both learners and teachers know which collocations to attend to by highlighting the collocations from a page being viewed. In the study of Ackermann and Chen (2013), they came up with the academic collocation list with 2,468 entries compiled from the Pearson International Corpus of Academic English. Regarding how to teach collocations, the pedagogical experiments have suggested a few methods such as awareness-raising and attention-drawing. Both approaches consist of strategies aimed at bringing collocations to the learners' attention, with an expectation that learners who become aware of collocations will use them more beneficially. These approaches have been tried extensively in various contexts, confirming that learners with raised consciousness on collocations are indeed better at using them correctly (Jiang 2009, Ying and O'Neill 2009).

For learners studying English for academic purposes, it has been proved that teaching collocations from reading — specifically through consciousness raising — can raise their competence in academic interactions (Conzett 2000, Lewis 2000). However, teaching certain collocational expressions through raising consciousness alone is not sufficient enough when it comes to fostering learner autonomy. In reality, English learners encounter countless situations

where they are expected to produce English texts using a far larger number of collocations that can be covered in class. Furthermore, it is virtually impossible for one language teacher to meet each individual learner's needs when there are dozens of students with various levels of proficiency from different majors. As Durrant (2009) pointed out, imparting knowledge of certain "important" collocations without taking sufficient account of variation across disciplines cannot possibly enable learners to write successfully in a comparatively advanced language proficiency. To quote the well-known maxim, language teachers need to teach students how "to fish" than give them "a fish." In other words, rather than confining to what collocations to teach, language teachers need to provide an autonomous learning tool that students can use in their endeavors to produce an English text.

Despite recognized significance of collocations for both accuracy and fluency (Nation and Webb 2011), a review of the relevant literature indicates that there are a limited number of studies conducted in EFL contexts that suggest ways to increase learners' collocational competence. To address this research gap, the present study was undertaken to examine the teachability of collocations through fostering collocation dictionary skills. More specifically, this investigation attempted to answer the following research questions:

- (1) Does teaching collocation dictionary skills affect EFL learners' ability to produce collocations correctly?
- (2) How do EFL learners perceive the learning of collocation dictionary skills?

2. Literature review

2.1 Definition of collocation

The term collocation has its origin in the Latin verb *collocare*, which means "to set in order/to arrange" (Martyńska 2004), and is used to refer to binary lexical combinations that "occur physically together or have stronger chances of being mentioned together" (Sinclair 1991: 170). In this regard, Gitsaki (1999) reviewed the three main approaches to collocations — lexical, semantic, and structural.

The lexical approach was first proposed by Firth (1957) and is based on the idea of word meaning at the lexical level. Lexical collocations comprise two or more content words as in *noun + verb*, *verb + noun*, *adjective + noun*, *adverb + adjective*, *verb + adverb*, and *adverb + verb* combinations. Following Firth (1957), Halliday et al. (1964) also explained collocation as the tendency of a lexical item to co-occur with one or more words.

Unlike the lexical approach, the semantic approach perceives the meaning of a lexical item as its semantic properties, which determine possible collocates (Cruse 1986). However, this approach is criticized for being unable to explain the large number of idiosyncratic co-occurrences that are arbitrarily restricted (Gitsaki 1999).

While the lexical and semantic approaches focus only on lexical words, the structural approach takes both lexical and grammatical collocations into consideration (Benson et al. 2010). According to this approach, collocations are viewed as combinations of a content word and a preposition or a grammatical structure like an infinitive or a clause. The structural approach is considered more pedagogical than the lexical and semantic approaches.

According to Nesselhauf (2003), collocations are to be delimited from other types of word combinations — such as "free combinations" (Cowie 1994) and "idioms" — based on arbitrary restriction on substitutability criterion. In a *verb + noun* combination (e.g., *read a newspaper*), for instance, if the senses of the verb and the noun are both unrestricted so that they can be freely combined with a number of other nouns and verbs (e.g., *read a book/a letter/a sign/a manual/a leaflet*), it is regarded as a free combination. In the case of *perform a task*, on the other hand, while the sense of the noun is unrestricted, that of the verb is restricted and thus can be combined with certain nouns only from a semantic point of view (e.g., **perform a survey*). Therefore, the combination *perform a task* is a collocation. Unlike the aforementioned cases, if both the verb and the noun are used in a restricted sense so that substitution is either not at all possible or extremely limited, such a combination is classified as an idiom (e.g., *sugar / sweeten the pill*).

2.2 Second language (L2) learners' difficulty in collocation production

The importance of collocational knowledge is beyond dispute. It enables learners to produce L2 texts that are more understandable and native-like (Hunston and Francis 2000). However, empirical studies have repeatedly reported learners' problems with L2 collocational use across proficiency levels (Biskup 1992).

Understandably, one of the biggest hindrance to the learning of collocations lies in the idiosyncratic nature of collocations (Halliday 1966). As depicted in Wray (2002: 73), "in English you run a business, but in German you lead it ... In English you smoke a cigarette, but in Hindi you drink it ... In English you lie in the sun, but in Russian you lie on it." Since neither syntactic nor semantic rules of English have been able to account for collocations, from the learner's perspective they appear to be co-occurrence of "seemingly inexplicable lexical choices" underlying so-called collocations (Bartsch 2004: 19).

Furthermore, the lack of exposure to the target language poses an equally serious problem. While native speakers acquire the knowledge of lexical constraints leading to the co-selection of particular words as they grow up in their speech community, most L2 learners who learn English almost exclusively within the classroom environment are deprived of such an opportunity. Devoid of linguistic intuitions and the ability to produce appropriate collocations (Groom 2009), they cannot but turn to their native language (L1) (Koya 2003, Laufer and Waldman 2011). According to the results of the previous studies, L2 learners often falsely assume a one-to-one correspondence between

L1 and L2 collocational choices, unaware that different languages follow markedly different collocational rules. Consequently, unless the collocation in their L1 happens to match that in L2, negative transfer occurs, resulting in deviant L2 combinations (Bahns and Eldaw 1993, Ellis 2008, Gass and Selinker 2008).

3. The study

3.1 Participants

The participants were 59 EFL college students attending a large private university in Seoul, Republic of Korea. Before the start of the new school year, all incoming students took the school-organized English proficiency test and were assigned to take English courses according to their proficiency level. The participants were from two intact freshman English classes that the author taught: One ($n = 30$) was for basic to low-intermediate students (hereafter referred to as the "low-intermediate group"), and the other ($n = 29$) for intermediate to advanced students (hereafter referred to as the "advanced group"). The class met twice a week over the span of roughly 15 weeks, and each class lasted for 75 minutes.

The participants were 19–22 years old and had learned English both at elementary and secondary schools and private language institutions for an average of approximately 12 years when they enrolled in the course. English was a foreign language for all of them with no prior experience of living in an English-dominant country, except for two students who had lived in the US for four years and two and a half years, respectively. Regarding ethnicity, Koreans constituted 73%, Chinese 24%, and Japanese 3% of the participants. The students of Chinese and Japanese nationality were fluent enough in Korean to understand written Korean vocabulary used in the collocation production tests and the collocation instruction worksheets.

3.2 Instruments

3.2.1 Pre-semester survey

The participants took a pre-semester survey on the first day of the semester. The survey items were regarding personal data such as name, age, and experience living abroad; language learning history; and their familiarity with the term/concept "collocation."

3.2.2 Collocation production tests

The collocation production tests were administered before and after the experimental intervention at a 12-week interval (in Weeks 1 and 13). To see

whether or not teaching collocation dictionary skills attributes to the development of learners' collocation production ability, the test scores were compared.

For a pilot study conducted before the start of the semester, 36 items were written concerning three collocation types (*verb + noun*, *adjective + noun*, and *preposition + noun* combinations), with 12 items per each type. They were drawn from McCarthy and O'Dell (2005) and online resources with revisions. Nine students were recruited from the same study population that would be used for the subsequent study. They were asked to elicit without dictionary consultation a collocating verb for the first type, an adjective for the second, and a preposition for the third. When multiple answers were possible, they were required to give only one answer, and all the correct answers were awarded one point. For instance, for the blank in the item (과중한) *workloads are not uncommon in today's workplace*, any one of the adjectives *enormous*, *heavy*, and *huge* was graded as correct.

Based on the result of the pilot study, the author sifted out six sentences (two from each type) that were found to be relatively easy and entailed no dictionary consultation. Then, the remaining 30 items were divided into two test sets of comparable difficulty (Type A and Type B), each consisting of 15 questions (see Appendix 1). For convenience of reference, each item number was followed by its test type (e.g., 3A for the third question in the Type A test). In the main study, half of the participants, whose student ID ends with an odd number, took the Type A as a pre-test and the Type B as a post-test; and the other way around for the other half, whose student ID ends with an even number.

With an exception of test-taking situations, students engage in a writing task flexibly using resources according to their needs. Moreover, since the present study did not concern EFL learners' prior knowledge of the target collocations, the participants took the test in a computer lab, freely using the desktop computers or their smartphones as resource materials.

3.2.3 End-of-semester survey

The end-of-semester survey was administered on the last day of the semester. The participants responded to an open-ended survey question, which solicited comments on the instruction they received on collocation and collocation dictionary look-up skills. They were allowed to choose from either Korean or English when answering the question.

3.3 Design for collocation dictionary skills instruction

For two consecutive class sessions (Classes 1 and 2) in Week 5, the participants received instruction relevant to collocational competence development. During Class 1, they were taught "what collocation is" and "why it is important." For

Class 2, the class met in the computer lab and received the 4-step instruction devised by the author, drawing upon Chen's (2016) observations on the causes of unsatisfactory collocation dictionary consultation behavior of L2 learners. After learning "how to check for collocational information" using the *Naver Dictionary* — the online dictionary service offered by Naver, one of Korea's major web portals — the participants worked on the collocation production task comprising a set of exercise questions in the formats of fill-in elicitation and sentence correction. Additionally, since the first drafts of their first writing assignment were returned that day, the students were asked to correct the marked (highlighted with a proofreading abbreviation "C" written on top) collocation errors — along with other lexical, grammatical, and discoursal aspects of their written work — and submit a revision by the next class meeting. The subsections below outline the four steps that the students were introduced to in Class 2.

3.3.1 Node word selection

As a way to teach collocation dictionary skills, the instructor first explained the overall lexicographic presentation of the *Naver Dictionary*, which provides collocational information retrieved from the *Oxford Collocations Dictionary for Students of English* as a separate section entitled '함께 쓰이는 단어 (literally meaning "words used together")' at the bottom of the look-up result page complemented by examples for the respective collocation types. Then, she demonstrated how to locate a list of feasible collocates using a "node word" — the term used to refer to an entry word to look up in the collocation dictionary (Sinclair 1991).

In addition, the students were instructed on how to divide a sentence into meaningful segments to pick a correct node word and how to simplify it if necessary. For instance, to fill in the blank in the sentence *CEO of Apple Steve Jobs resigned (때문에) health reasons and Tim Cook officially took over the reigns on August 24, 2011*, the students were told to divide the sentence into four segments ([Segment 1] *CEO of Apple Steve Jobs resigned* [Segment 2] (때문에) *health reasons* [Segment 3] *and Tim Cook officially took over the reigns* [Segment 4] *on August 24, 2011*) and use *reason* — not its plural form, *reasons*, nor the immediately adjacent *resigned* or *health* — as a node word, which directed them to the possible preposition collocate — *for*. In the same manner, for the gapped sentence *While the names of some ingredients may sound exotic, many are (손쉽게) available (에서) any local supermarket*, the students were advised to use *available* as a node word for the first blank, and a simplified version of *supermarket* — *market* — for the second.

3.3.2 Word sense distinction

After learning how to choose a simplest possible node word for locating its collocate(s), the students were introduced to the layout of an entry for a polysemous word. Here, the instructor repeatedly urged the students to

examine all the word senses, not just the first one or two (Chen 2016), until they run into the correct meaning in the given context. For example, after typing in the node word *time* to fill in the first blank in *It was extremely difficult to* (정하다) *a time for a make-up class* (에) *weekdays*, the students were directed to the result page showing nine definitions of *time* when used as a noun, along with its meaning as a verb (see Figure 1). In this case, the correct sense of the word is the third one ("time when sth happens/should happen").

함께 사용되는 단어

[명사] 로 time이(가) 사용될 때

1. what is measured in minutes, hours, days, etc.의 의미인 경우

[time + 동사] elapse, go by, pass | fly | go | drag | heal sth

- As time went by we saw less and less of each other.

The changing seasons mark the passing of time.

- How time flies!

2. time shown on a clock의 의미인 경우

[형용사] good, perfect | local | daylight saving | British Summer, Eastern Standard, Greenwich Mean, etc.

- My watch keeps good time.

- The attacks were launched at 9 p.m. local time.

[동사 + time] tell(BrE)(AmE) | have(AmE) | make(BrE) | look at | check | keep

- Can he tell the time yet?

Can he tell time yet?

3. time when sth happens/should happen의 의미인 경우

[형용사] peak | prime | closing, opening | arrival, departure

- There are extra buses at peak times.

[동사 + time] fix(BrE), set | change

- We need to set a time for the next meeting.

[time + 동사] come

- You'll feel differently about it when the time comes.

[전치사] ahead of~ | behind~(BrE) | by the~ | in~ | on~ | ~for

- We had everything worked out ahead of time.

- The plane took off an hour behind time.

- By the time you get there the meeting will be over.

- We got home in time to see the end of the game.

- The trains are rarely on time.

- It's time for dinner.

Figure 1: Screenshot of the *Naver Dictionary* for the entry word *time*, noun

3.3.3 Collocate type location

After identifying the correct sense of the multiple-meaning node word, the participants were asked to locate the collate type they were looking for. Since

the Korean translation of the expected collocate was provided next to each blank, this step took almost no effort to complete and was finished immediately.

3.3.4 Feasible collocate identification

The last step for consulting collocational information was to identify the possible collocate(s) in the given context. For the gapped sentence *College students should (지다) responsibility for their own learning*, for example, the students first typed in *responsibility* as a node word. Then, the result page showed a list of feasible collocates divided by vertical bars, denoting that those enumerated within the vertical bars can be used interchangeably (see Figure 2). In this case, the correct collocates are the second set of verbs (*accept, acknowledge, assume, bear, shoulder, carry, recognize, shoulder, take, take on, take over*).

함께 사용되는 단어

[명사]로 responsibility이(가) 사용될 때

1. being responsible의 의미인 경우

[형용사] complete, full, total | awesome, big, enormous, grave, great, heavy, huge, important, serious, tremendous, weighty | direct | primary | overall, ultimate | special | diminished(law, esp. BrE) | exclusive, sole | collective, communal, joint, mutual, shared | individual, personal | corporate | criminal | security | constitutional, political | environmental, ethical, financial, fiscal, legal, moral, social

- It is a great responsibility caring for other people's children.
- Ultimate responsibility rests with the president.
- He was found not guilty of murder on the grounds of diminished responsibility.

[동사 + responsibility] have | accept, acknowledge, assume, bear, carry, recognize, shoulder, take, take on, take over | feel | share | lay, place | assign, delegate, devolve, give sb, hand over | retain | abdicate, escape, evade, shift | admit, claim | deny, disclaim, duck | allocate, assign (sb), give sb | attribute | burden sb with | absolve sb from, absolve sb of

- She has responsibility for public spending.
- The bank refuses to accept responsibility for the mistake.
Will you take responsibility for arranging the food?
- He feels a responsibility to his community.
- The government of the time placed responsibility for the poor on the Church.
- Responsibility is devolved down to the people who are affected.
- We retain all responsibility for any shortcomings.
- They wanted to shift responsibility for the failure onto their employees.
- No organization has yet claimed responsibility for the bomb attack.
- Ducking responsibility is fatal in a democracy.
- He attributed responsibility for the killing to the secret service.

[responsibility + 동사] fall on sb, fall to sb, lie with sb, rest with sb | come with sth

- With great power comes great responsibility.

[전치사] ~for | ~towards/toward

- Full responsibility for the fiasco lies with the PR department.
- He feels a strong sense of responsibility towards/toward his parents.

Figure 2: Screenshot of the Naver Dictionary for the entry word *responsibility*, noun

3.4 Data analysis

Four sets of data were collected and analyzed: the yes/no answers to the pre-semester survey question about familiarity with the term/concept collocation, the pre-test collocation production task scores, the post-test collocation production task scores, and the comments to the end-of-semester survey question. For the first data set, the number of "yes" responses was counted and its percentage was calculated. To examine "whether or not teaching collocation dictionary skills affects EFL learners' ability to produce collocations correctly" (Research Question 1), the pre- and post-tests were scored manually by the author. One point was awarded for each correct answer, up to the maximum total of 15. Then, the participants' pre-test scores were compared against their post-test scores by a paired-samples *t*-test, using an alpha level of .05. To probe into "how EFL learners perceive the learning of collocation dictionary skills" (Research Question 2), the end-of-semester survey data were analyzed. The comments obtained from the open-ended question were categorized; frequencies were counted and ranked.

4. Results

4.1 EFL college students' familiarity with collocation

The analysis of the pre-semester survey data indicated that collocation is a foreign concept to almost all participants. To the question "whether or not they have heard of the term 'collocation' before," 58 of 59 students responded "no." In Korean, *collocation* ("연어") and *salmon* ("연어") are homonyms. Although the survey item was written in both Korean and English, two respondents answered the question by asking in return, "Salmon as in salmon sushi?" Only one student answered that she had heard of it from her private English academy teacher. To the follow-up question "if and how she uses a collocation dictionary," however, the answer was negative. Confirming what has been reported unanimously in the literature, most EFL learners were found to "have no knowledge of collocation dictionaries or other potential resources" that they can use to address their collocation problems (Henriksen 2013: 42).

4.2 Comparison of pre- and post-test scores

To answer the first research question "whether teaching collocation dictionary skills affects EFL learners' ability to produce collocations correctly," a paired-samples *t*-test was performed on each group's pre- and post-test composite scores. The descriptive statistics and statistical analyses are presented in Table 1.

Table 1: Comparison of pre- and post-test composite scores

Group	N	Pre-test		Post-test		t	p
		M	SD	M	SD		
Low-intermediate	30	5.91	4.53	13.14	0.43	-16.830	.000
Advanced	29	7.71	3.10	13.37	0.38	-15.990	.000

The mean pre-test composite scores were 5.91 (out of 15) for the low-intermediate group and 7.71 for the advanced group. The mean post-test scores improved substantially to 13.14 and 13.37 for the low-intermediate and advanced groups, respectively. Not surprisingly, the *p*-values were reported far less than the significance level for both groups (*p* = .000).

Then, to see if the instruction positively affects learners' collocation production ability without reference to collocation types, the paired-samples *t*-test was performed one more time on the three pre- and post-test subscale scores for the *verb + noun*, *adjective + noun*, and *preposition + noun* combinations. The results are shown in Table 2.

Table 2: Comparison of pre- and post-test subscale scores

Group	Collocation type	Pre-test		Post-test		t	p
		M	SD	M	SD		
Low-intermediate	verb + noun	2.44	1.17	5.00	.00	-12.227	.000
	adjective + noun	1.96	.49	5.00	.00	-22.349	.000
	preposition + noun	1.51	1.79	3.14	.43	-5.662	.000
Advanced	verb + noun	3.00	.44	5.00	.00	-15.874	.000
	adjective + noun	2.71	1.17	5.00	.00	-11.159	.000
	preposition + noun	2.00	.59	3.37	.38	-7.258	.000

For the collocation type of the *verb + noun* combination, the mean subscale scores improved from 2.44 (out of 5) to 5.00 for the low-intermediate group and from 3.00 to 5.00 for the advanced group. For the collocation type of the *adjective + noun* combination as well, the participants achieved a considerable mean subscale score gain from 1.96 to 5.00 for the low-intermediate group and from 2.71 to 5.00 for the advanced group. As expected, the mean subscale score differences were statistically significant for both groups (*p* = .000). For the collocation type of the *preposition + noun* combination, the mean subscale scores increased from 1.51 to 3.14 for the low-intermediate group and from 2.00 to 3.37 for the advanced group. Although the participants' performance on this type

was not as impressive as on the first two collocation types, the mean subscale score gains were statistically meaningful for both groups ($p = .000$).

4.3 EFL learners' perception toward collocation dictionary skills instruction

To assess EFL learners' perception toward the learning of collocation dictionary skills, the answers to the open-ended survey question were analyzed. Since some participants offered multiple comments, the total number of responses is greater than the number of respondents. The comments written in Korean were translated into English by the author. Table 3 illustrates a summary.

Table 3: EFL learners' perception toward collocation dictionary skills instruction

Comment	Number of responses
I feel confident about English writing now that I can check the naturalness of a text by myself.	37
A collocation dictionary is a very useful tool for English composition.	32
The way I consult a dictionary to look up natural-sounding collocations changed completely after receiving the instruction.	16
Instruction on collocation dictionary skills should constitute an indispensable part of English curriculum.	11
There should be a bilingualised collocation dictionary to make it easier to locate a correct sense of a polysemous word.	6
A collocation dictionary should provide additional explanation, along with a list of collocates, for collocations involving prepositions.	4
Mobile collocation dictionary apps should be developed.	1

Overall, the respondents made favorable comments about the usefulness of the instruction targeting collocation dictionary skills development, and it was notable that none reported negatively about its helpfulness. A vast majority responded that the instruction helped them to gain or boost confidence with English writing, mainly thanks to increased collocational competence. Particularly, the participants valued the instruction, which — as one respondent put it — "freed" them from having to ask others to proofread their "humble" papers,

risking losing face. Such comments could be interpreted to mean that the instruction helps raise learner autonomy in collocation production. Over 23% of the participants ($n = 14$) responded that their dictionary consultation behavior changed fundamentally after receiving the instruction. Specifically, they commented that they no longer check pages of examples in an English–Korean dictionary in search of similar-meaning sentences.

In addition to expressing opinions about the instruction, some respondents made suggestive feedback. Six students suggested that a bilingualised collocation dictionary should be made available because it was overly demanding to use the monolingual English collocation dictionary, especially when distinguishing senses for polysemous words and choosing a collocating preposition. As illustrated in Figures 1 and 2, the *Naver Dictionary* offers collocation information in English only, with the exception of the Korean phrases 이(가) 사용될 때 ("when used as sth") and 의 의미인 경우 ("in the case where it means sth") used repeatedly in the section, and the parts of speech in Korean — e.g., 명사 ("noun"), 형용사 ("adjective"), 동사 ("verb"), and 전치사 ("preposition"). Also, one respondent addressed the need for developing a collocation dictionary app for the convenience of mobile-savvy college students.

5. Discussion and conclusion

The present study was an attempt to validate the effectiveness of the teaching of collocation dictionary skills in fostering EFL learners' collocational competence. In accordance with Laufer (2011) and Chen (2016), it was found that EFL learners — regardless of their proficiency level — significantly improved their ability to produce natural collocations with the aid of an online collocation dictionary. Also, the survey data suggested that the instruction targeting collocation dictionary skills development was positively perceived by the learners.

For the pre-test collocation production task, the overall correct answer rate was approximately 39% for the low-intermediate group and 51% for the advanced group. What was a little surprising was the unsatisfactory performance of the students in the advanced group, who were supposed to be relatively proficient in English. Besides, given that they could use a computer or a smartphone to consult online resources including dictionaries, the result that they managed to produce collocations only at a little over a 50% success rate was quite disappointing. In this regard, the result of the present study seems to lend further support to Howarth's (1996, 1998) and Nesselhauf's (2003) claim that collocations are challenging even to advanced-level learners, not to mention less proficient ones. Therefore, since a substantial number of word combinations that can be readily understood cause problems when producing them, it is deemed necessary for English instructors to raise learner's awareness of collocation (Chen 2016).

Interestingly, while proctoring the pre-test, the author/instructor could

easily witness that a good many students did not bother consulting a dictionary or other online materials at their disposal. Instead, they at times filled in the blanks with deviant L2 collocates without hesitation, and did not check the correctness of their answers, possibly due to a lack of awareness that the collocations were in fact unfamiliar to them (Laufer 2011). Given that they produced collocations in such a manner during the collocation production *test*, it might go without saying that they do not consult dictionaries as often as they need to when they write. By and large, the observation seems to indicate that, consistent with Bahns (1993) and Bahns and Eldaw (1993), many EFL learners do not grasp the possible lexical incongruence between their L1 and the target language.

For Items 4A (*All students are required to [듣다] at least two writing classes*) and 6B (*These are core courses you have to [듣다] this year*), for instance, 46% of the students on average answered incorrectly and the wrong answers included *hear*, *listen*, and *listen to*, all of which are direct translation of the given Korean verb 듣다. Also, for Item 7A (*You need a student ID card to [연장하다] the library book*), the inaccuracy level logged 75% among the students who took the Type A as a pre-test, and the two most frequent incorrect answers were *extend* and *prolong* — English equivalents for the given Korean verb 연장하다. As amply shown in previous studies, it was evident that learners across proficiency levels negatively transfer their linguistic knowledge of the L1 to an L2 context as a common learning strategy (Selinker 1992, Fan 2009, Sadeghi 2009).

Drawing a stark contrast with the performance on the pre-test, the analysis of the post-test scores provided a compelling evidence for the effectiveness of the instruction targeting collocation dictionary skills development. In accordance with Laufer (2011) and Chen (2016), teaching collocation dictionary skills significantly increased the number of correct collocations on the post-test. A close look into the participants' performance on each of the three collocation types showed that all participants got a perfect score for the *verb + noun* and *adjective + noun* combinations. Of course, the fact that the participants improved their collocations production ability comfortably conforms to the initial expectations of the study; however, such dramatic progress resulting in *perfect* performance was far beyond what any researcher could possibly expect. Even though further research is needed to identify possible factors, one plausible explanation might be that, since the test was to fill in the blank with a verb for the *verb + noun* combination and an adjective for the *adjective + noun* combination — both of which are "content" words — it must have been doable for the participants to sensibly distinguish among a number of feasible collocate sets. For Item 4B (*It was so relaxing to be among [예전의] friends*), for example, after typing in *friend* as a node word, one is directed to the result page presented in Figure 3.

Among a list of possible collocates, one can easily tell that the third set of adjectives — *lifelong*, *long-standing*, *long-time*, *old* — means the closest meaning

to the provided Korean equivalent 예전의 ("old/past"). Therefore, once establishing a habit of consulting collocation dictionaries for collocational information, one may well find it easier to pick natural-sounding collocates for the node word than doing otherwise.

함께 사용되는 단어

[명사]로 friend이(가) 사용될 때

[형용사] best, bosom, close, dear, fast(AmE), good, great, intimate, real, special | faithful, loyal, real, true, trusted | lifelong, long-standing(esp. BrE), long-time, old | long-lost | new, new-found | female, guy(AmE,informal), male, woman | gay, straight | single | fair-weather | so-called, supposed | mutual | family, personal | childhood, college, school | imaginary | online

- Her best friend at school was called Anna.
- I'm inviting only my closest friends to the party.
- It was so relaxing to be among old friends.
- He was last seen leaving a restaurant with a female friend.
- Does your sister have any single friends?
- People he had trusted turned out to be only fair-weather friends.
- My so-called friends are making fun of me because of my weight.
- We met each other through a mutual friend.
- Do you keep in touch with any school friends?

Figure 3: Screenshot of the *Naver Dictionary* for the entry word *friend*, noun

For the *preposition + noun* collocation type, participants' post-test performance was not as impressive as on the other two collocation types, even though both groups achieved a statistically meaningful mean subscale score gains. Unlike Dziemianko's (2010) study, in which the Polish college students succeeded in completing gapped sentences with prepositions at an over 92% success rate, the average post-test score of the participants in this study was slightly over 65%. Even though further investigations into this matter are required to establish the validity, it stands to reason that such a result was at least partly attributable to difference in word classes. That is, for this collocation type, the students were asked to fill in the blank with a preposition — a "function" word — which has little lexical meaning in its own nature (Fries 1952). In the case of Item 3A (*Nowadays, we can buy Eiffel Tower tickets [에서] the web*), for example, after typing in *web* as a node word, one is directed to the result page that shows only one collocating preposition — *on* (see Figure 4).

함께 사용되는 단어

[명사]로 web이(가) 사용될 때

1. that a spider makes의 의미인 경우

[형용사] spider(esp. AmE), spider's(esp. BrE)

[동사 + web] build, make, spin, weave



Figure 4: Screenshot of the *Naver Dictionary* for the entry word *web*, noun

Expectedly enough, all participants filled in the blank with the correct preposition. Likewise, the items whose node words collocate with only one preposition (Items 3A, 12A, 2B, 5B, and 8B) also logged 100% correct answer percentage. For those with multiple possible collocates (Items 5A, 9A, 15A, 11B, and 14B), however, the average accuracy percentage was only barely over 31%. For Item 9A (*The school provides free Wi-Fi [에서] campus*), for instance, after typing in the node word *campus*, one is directed to the result page presented in Figure 5.

함께 사용되는 단어



Figure 5: Screenshot of the *Naver Dictionary* for the entry word *campus*, noun

To fill in the blank, one needs to choose from the three options of *at a/the ~*, *off ~*, and *on ~*. The correct collocate is the last one as there is no article in front of *campus* in the given sentence. Despite the fact that it was heavily stressed during the instruction that they need to check whether or not a nominal node word is preceded by an article, a majority of the participants wrongly chose *at*. Since EFL learners tend to lack linguistic intuitions about naturalness of word combinations, it must have been challenging even for the advanced group to pick the correct collocating preposition. Similarly, for Item 14B (*Her novel is [에] the list of recommended readings*), after typing in *list* as a node word, the participants must have been confronted with the choice between *in* and *on* (see Figure 6).

함께 사용되는 단어

[명사]로 list이(가) 사용될 때

[전치사] in a/the~ | on a/the~ | ~for | ~of

- Names of past members are not included in the list.
- I can't see your name on the list.
- We acquired a waiting list for works by these artists.
- a list of 2 names

Figure 6: Screenshot of the *Naver Dictionary* for the entry word *list*, noun

Of course, one might have taken the hint from the example *Names of past members are not included in the list* that *in the list*, as opposed to *on the list*, is used after the verb *include*; but, honestly speaking, that sounds almost entirely unlikely. The fact that the accuracy percentage of the item recorded a scant 12% bears out the theory, suggesting that embedding additional information within an example is not salient enough to catch attention. Therefore, as suggested by Gouws (2015: 179), it seems necessary that collocation dictionaries provide "additional semantic guidance to ensure a successful comprehension of the collocation."

An equally significant, and encouraging, finding from this study is that the teaching of collocation dictionary skills changed the participants' dictionary consultation behavior. While proctoring the post-test, the author could easily observe the difference: The students navigated to the online collocation dictionary and then took on the collocation production task. As described earlier, quite many of them filled in the blanks on the pre-test without consulting a dictionary or other materials, which often led to inaccurate output. After receiving the instruction, however, the students began to check for collocational word pairings in L2, which they used to take for granted if those word combinations made sense in their L1. Since they were provided with written feedback on collocation errors in their six writing assignment drafts throughout the semester and were encouraged to correct them by themselves with the help of a collocation dictionary, all participants must have become familiar with the collocation dictionary by the time they took the post-test. Given that colloca-

tions cannot be possibly predicted, the fact that the students began to develop the habit of consulting collocation dictionaries — albeit not necessarily voluntarily — might be the single most meaningful outcome of this study.

All in all, the improvement the participants made in producing correct collocations after receiving two 75-minute instructions was dramatic, indeed. The attempt at teaching EFL students "how to fish" seems to have suitably equipped them with an autonomous learning tool with which they can take charge of their own learning, free from teacher guidance or direction (Woolard 2000, Smith 2008). In addition to the improved reference skills, the participants' newly formed habit to consult a dictionary for collocation decoding seems to equally contribute to the positive results of the current study. As Zimmerman (2009) pointed out, non-native students need to turn to collocation dictionaries to achieve accurate and fluent collocation production, and it is a language instructor's job to teach them the necessary skill to do so (Ranalli 2013).

References

- Ackermann, K. and Y.H. Chen. 2013. Developing the Academic Collocation List (ACL): A Corpus-Driven and Expert-judged Approach. *Journal of English for Academic Purposes* 12(4): 235-247.
- Bahns, J. 1993. Lexical Collocations: A Contrastive View. *ELT Journal* 47(1): 56-63.
- Bahns, J. and M. Eldaw. 1993. Should We Teach EFL Students Collocations? *System* 21(1): 101-114.
- Bartsch, S. 2004. *Structural and Functional Properties of Collocations in English: A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence*. Tübingen: Gunter Narr.
- Benson, M., E. Benson and R.F. Ilson (Eds.). 2010. *The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar*. Third edition. Amsterdam: John Benjamins.
- Biskup, D. 1992. L1 Influence on Learner's Renderings of English Collocations: A Polish/German Empirical Study. Arnaud, P.J.L. and H. Béjoint (Eds.). 1992. *Vocabulary and Applied Linguistics*: 85-93. London: Macmillan.
- Chen, Y. 2016. Dictionary Use for Collocation Production and Retention: A CALL-Based Study. *International Journal of Lexicography*. doi:10.1093/ijl/ecw005.
- Conzett, J. 2000. Integrating Collocation into a Reading and Writing Course. Lewis, M. (Ed.). 2000. *Teaching Collocation: Further Developments in the Lexical Approach*: 70-87. Hove, UK: Language Teaching.
- Cowie, A.P. 1994. Phraseology. Asher, R.E. (Ed.). 1994. *The Encyclopedia of Language and Linguistics*: 3168-3171. Oxford: Pergamon.
- Cowie, A.P. (Ed.). 1998. *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.
- Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Durrant, P. 2009. Investigating the Viability of a Collocation List for Students of English for Academic Purposes. *English for Specific Purposes* 28(3): 157-169.
- Dziemianko, A. 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.
- Ellis, R. 2008. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Fan, M. 2009. An Exploratory Study of Collocational Use by ESL Students — A Task Based Approach. *System* 37(1): 110-123.

- Firth, J.R.** 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.
- Fries, C.C.** 1952. *The Structure of English*. New York: Harcourt Brace.
- Gass, S.M. and L. Selinker.** 2008. *Second Language Acquisition: An Introductory Course*. New York: Routledge.
- Gitsaki, C.** 1999. *Second Language Acquisition: A Study of the Development of Collocational Knowledge*. San Francisco: International Scholars.
- Gouws, R.H.** 2015. The Presentation and Treatment of Collocations as Secondary Guiding Elements in Dictionaries. *Lexikos* 25: 170-190.
- Granger, S.** 1998. Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. Cowie, A.P. (Ed.). 1998. *Phraseology: Theory, Analysis and Applications*: 145-160. Oxford: Oxford University Press.
- Granger, S. and F. Meunier (Eds.)**. 2008. *Phraseology. An Interdisciplinary Perspective*. Amsterdam/Philadelphia: John Benjamins.
- Groom, N.** 2009. Effects of Second Language Immersion on Second Language Collocational Development. Barfield, A. and H. Gyllstad (Eds.). 2009. *Researching Collocations in Another Language: Multiple Interpretations*: 21-33. Basingstoke, UK: Palgrave Macmillan.
- Halliday, M.A.K.** 1966. Lexis as a Linguistic Level. Bazell, C.E., J.C. Catford, M.A.K. Halliday and R.H. Robins (Eds.). 1966. *In Memory of J.R. Firth*: 148–162. London: Longman.
- Halliday, M.A.K., A. McIntosh and P. Stevens.** 1964. *The Linguistic Sciences and Language Teaching*. London: Longmans.
- Henriksen, B.** 2013. Research on L2 Learners' Collocational Competence and Development — A Progress Report. Bardel, C., C. Lindqvist and B. Laufer (Eds.). 2013. *L2 Vocabulary Acquisition, Knowledge and Use. New Perspectives on Assessment and Corpus Analysis*. Eurosla Monographs Series 2: 29-56. European Second Language Association.
- Howarth, P.** 1996. *Phraseology in English Academic Writing. Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer.
- Howarth, P.** 1998. Phraseology and Second Language Proficiency. *Applied Linguistics* 19(1): 24-44.
- Hunston, S. and G. Francis.** 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Jiang, J.** 2009. Designing Pedagogic Materials to Improve Awareness and Productive Use of L2 Collocations. Barfield, A. and H. Gyllstad (Eds.). 2009. *Researching Collocations in Another Language*: 99-113. New York: Palgrave Macmillan.
- Kennedy, G.D.** 1990. Collocations: Where Grammar and Vocabulary Teaching Meet. Anivan, S. (Ed.). 1990. *Language Teaching Methodology for the Nineties*: 215-219. Singapore: SEAMEO Regional Language Centre.
- Koya, T.** 2003. A Study of Collocation in English and Japanese Noun–Verb Combinations. *Intercultural Communication Studies* 12(1): 125-141.
- Laufer, B.** 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, B. and T. Waldman.** 2011. Verb–Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61(2): 647-672.
- Lewis, M. (Ed.)**. 2000. *Teaching Collocations: Further Developments in the Lexical Approach*. Hove, UK: Language Teaching Publications.
- Martyńska, M.** 2004. Do English Language Learners Know Collocations? *Investigationes Linguisticae* 11: 1-12.

- McCarthy, M. and F. O'Dell.** 2005. *English Collocations in Use*. Cambridge: Cambridge University Press.
- Nation, P. and S.A. Webb.** 2011. *Researching and Analyzing Vocabulary*. Boston: Heinle Cengage Learning.
- Nattinger, J.R. and J.S. DeCarrico.** 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesselhauf, N.** 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24(2): 223-242.
- Nesselhauf, N.** 2005. *Collocations in a Learner Corpus*. Amsterdam/Philadelphia: John Benjamins.
- Ranalli, J.** 2013. Designing Online Strategy Instruction for Integrated Vocabulary Depth of Knowledge and Web-based Dictionary Skills. *CALICO Journal* 30(1): 16-43.
- Sadeghi, K.** 2009. Collocational Differences Between L1 and L2: Implications for EFL Learners and Teachers. *TESL Canada Journal* 26(2): 100-124.
- Selinker, L.** 1992. *Rediscovering Interlanguage*. London: Longman.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Smith, R.** 2008. Learner Autonomy. *ELT Journal* 62(4): 395-397.
- Wible, D., C-H. Kuo, M-C. Chen, N-L. Tsao and T-F. Hung.** June 2006. A Computational Approach to the Discovery and Representation of Lexical Chunks. *Proceedings of the 13th Conference on Natural Language Processing (TALN 2006), 10-13 April, 2006, Leuven, Belgium*: 868-875.
- Woolard, G.** 2000. Collocation: Encouraging Learner Independence. Lewis, M. (Ed.). 2000. *Teaching Collocation: Further Developments in the Lexical Approach*: 28-46. Hove, UK: Language Teaching Publications.
- Wray, A.** 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Ying, Y. and M. O'Neill.** 2009. Collocation Learning Through an 'AWARE' Approach: Learner Perspectives and Learning Process. Barfield, A. and H. Gyllstad (Eds.). 2009. *Researching Collocations in Another Language*: 181-193. New York: Palgrave Macmillan.
- Zimmerman, C.B.** 2009. *Word Knowledge: A Vocabulary Teacher's Handbook*. Oxford: Oxford University Press.

Appendix 1: Collocation production tests

Type A

Fill in the blank with the most appropriate word.

- 1A. An unbalanced diet can (야기하다) serious problems.
- 2A. (과중한) workloads are not uncommon in today's workplace.
- 3A. Nowadays, we can buy Eiffel Tower tickets (에서) the web.
- 4A. All students are required to (듣다) at least two writing classes.
- 5A. She was filled with many emotions (에) her 20th birthday.
- 6A. Students drink energy drinks or (진한) coffee to stay up late.
- 7A. You need a student ID card to (연장하다) the library book.
- 8A. The school cafeteria has a(n) (다양한) menu for lunch.
- 9A. The school provides free Wi-Fi (에서) campus.
- 10A. With this timeline, we can (맞추다) the schedule.
- 11A. There will be (많은) snow tomorrow.
(Note: Do not use "a lot of" or "lots of" to fill in the blank.)
- 12A. There are many historic buildings (에는) this university.
- 13A. Teachers often (내다) assignments far beyond students' abilities.
- 14A. There was a(n) (강한) smell of burning tires.
- 15A. Students can check almost all information (에서) the Internet.

Type B

Fill in the blank with the most appropriate word.

- 1B. The shop offers a(n) (다양한) selection of wines.
- 2B. I am going to a concert (에) Christmas Eve.
- 3B. Many countries (입다) damage every year from the typhoon.
- 4B. It was so relaxing to be among (예전의) friends.
- 5B. We can easily find free online chatting sites (에서) the website.
- 6B. These are core courses you have to (듣다) this year.
- 7B. I like the (진한) scent of Indian cooking.
- 8B. I think I did badly (에서) the history test.
- 9B. We must (하다) an effort to stay healthy.
- 10B. Below are some of the most (흔한) English mistakes.

- 11B. The event was canceled (때문에) safety reasons.
- 12B. You can (채우다) three credits by taking this course.
- 13B. Many health problems are associated with (과도한) drinking.
- 14B. Her novel is (에) the list of recommended readings.
- 15B. Most journalists work under pressure to (맞추다) deadlines.

Multimodal Exemplification: The Expansion of Meaning in Electronic Dictionaries

Xiqin Liu, *School of Foreign Languages,
South China University of Technology, China and
Guangdong University of Foreign Studies, China*
(flxqliu@scut.edu.cn)

Abstract: This article investigates electronic dictionaries under the framework of Systemic-Functional Multimodal Discourse Analysis (SF-MDA) and argues for improving their exemplification multimodally. Multimodal devices, if well coordinated, can help optimize e-dictionary examples in informativity, diversity, dynamicity and interactivity. The term multimodal exemplification is tentatively proposed under the umbrella of multimodal lexicography (Lew 2010), and defined as the selection and presentation of examples with multimodal devices for achieving greater effectiveness in exemplifying than language does alone, especially in an e-dictionary. Evidence shows that multimodal exemplification can expand the three metafunctional meanings of the e-dictionary discourse: ideational, interpersonal and textual. Ideational meaning can be enriched by not only multimodal examples *per se* but also cross-modal example–definition ties, and hyperlinks facilitate meaning flow in the semantic network. Interpersonal meaning can be expanded by user participation and design options, including those for page layout (spatial mode) and example genre style (verbal mode). Textual meaning can be reinforced by information value, composition, salience and framing. This article makes a first attempt to explore the intermodal relationship between a definition and the examples under the same sense, and to present a diagram illustrating a typical design of visual space in e-dictionaries. By exploring the special features of multimodal example texts, it may deepen our understanding of the emerging multimodal lexicography and complement multimodal discourse studies from a lexicographical perspective.

Keywords: E-DICTIONARY, EXAMPLE, METAFUNCTIONAL MEANING, MULTIMODAL DISCOURSE ANALYSIS, MULTIMODAL EXEMPLIFICATION, MULTIMODAL LEXICOGRAPHY

Opsomming: Multimodale toelgting: Betekenisuitbreiding in elektroniese woordeboeke. In hierdie artikel word elektroniese woordeboeke binne die raamwerk van Siste-mies-Funksionele Multimodale Diskoersanalise (SF-MDA) ondersoek en geargumenteer ten gunste van die verbetering van hul multimodale toelgtingshulpmiddels. Indien multimodale hulpmid-dels goed gekoördineer word, kan hulle help om die e-woordeboekvoorbeelde te optimaliseer ten opsigte van informatiwiteit, diversiteit, dinamiek en interaktiwiteit. Die term multimodale toelg-tig word tentatief voorgestel onder die oorkoepelende multimodale leksikografie (Lew 2010), en gedefinieer as die seleksie en aanbieding van voorbeelde met behulp van multimodale hulpmid-dels ten einde groter effektiwiteit in toelgting te verkry as wat met slegs taal gedoen kan word,

veral in 'n e-woordeboek. Daar is bewyse dat multimodale toeligting die drie metafunksionele betekenis van die e-woordeboekdiskoers kan uitbrei: begripvormend, interpersoonlik en teksueel. Begripvormende betekenis kan uitgebrei word nie alleen deur multimodale voorbeelde *per se* nie, maar ook deur tussenmodale voorbeeld-definisie verbintnisse, en hiperskakels vergemaklik betekenisvloei in die semantiese netwerk. Interpersoonlike betekenis kan uitgebrei word deur gebruikersdeelname en ontwerppopsies, insluitend dié vir bladuitleg (ruimtelike modus) en voorbeeldgenrestyl (verbale modus). Teksuele betekenis kan versterk word deur inligtingswaarde, samestelling, treffendheid en raming. In hierdie artikel word 'n eerste poging aangewend om die intermodale verhouding tussen 'n definisie en die voorbeelde onder dieselfde betekenis te verken, en om 'n diagram aan te bied wat 'n ontwerp tipies aan die visuele ruimte in e-woordeboeke illustreer. Deur die spesiale kenmerke van multimodale voorbeeldtekste te verken, kan ons begrip van die groeiende multimodale leksikografie dalk verbeter word en kan die multimodale diskoersstudies vanuit 'n leksikografiese perspektief dalk aangevul word.

Sleutelwoorde: E-WOORDEBOEK, VOORBEELD, METAFUNKSIONELE BETEKENIS, MULTIMODALE DISKOERSANALISE, MULTIMODALE TOELIGTING, MULTIMODALE LEKSIKOGRAFIE

1. Introduction

Electronic dictionaries (e-dictionaries) have become increasingly popular as a tool of language learning and use in the digital era. In 2012, Macmillan announced that henceforth its dictionaries would only be published online. As "we are liberated from the straitjacket of the printed page and alphabetical order" (Atkins 1996: 516), the use of multimodal resources has become an important topic for lexicography. The combination of different semiotic modes (textual, visual, audio, spatial, etc.) is critical to the information processing of an audience because it can expand meaning by creating new space for interpretations (Lemke 2002). So multimodality plays an important role in meaning representation in (e-)dictionaries.

Examples serve as a basic constituent of the dictionary microstructure, reinforcing meaning explanations, illustrating collocations and colligations, and contextualizing for cultural, stylistic and pragmatic implications (Xu 2009: 12, 26-29; see also Nielsen 2014; Prinsloo 2015; Taljard 2015). In e-dictionaries, the different semiotic modes can be employed to develop this potential to a much greater extent than in print dictionaries. In fact, e-dictionary multimodality has changed exemplification profoundly, enriching and expanding meaning in various dimensions. To deepen our understanding of e-dictionary exemplification in the digital transformation, the term multimodal exemplification is proposed under the umbrella of multimodal lexicography (Lew 2010), following my previous work on multimodal definition¹ (Liu 2015). It is tentatively defined as the selection and presentation of examples with multimodal devices for achieving greater effectiveness in exemplifying than language does alone, especially in an e-dictionary.

The present study is guided by the framework of Multimodal Discourse Analysis from a Systemic-Functional perspective, SF-MDA (cf. O'Halloran 2008). MDA is an emerging paradigm in discourse studies which extends the study of language per se to the study of language in combination with other resources, such as images, scientific symbolism, gesture, action, music and sound (O'Halloran 2011: 120). After all, dictionary examples constitute a special genre of discourse, and contribute to the formation of a dictionary text, achieving cohesion with it (cf. Szende 1999; Xu 2009: 8, 26-29). The (multimodal) discourse perspective can be traced back to a holistic view of meaning in a dictionary.

Following the theory of "meaning as use" (Wittgenstein 1953), meaning and usage are inseparable, and "in a living language vocabulary and grammar do not have their own independent existences" (Tarp 2008: 135). Cognitive findings tell us that linguistic information of a word or phrase is merged with encyclopedic information. Meaning is of multiple facets blended into a whole, and "multifaceted meaning descriptions can improve the usability of a dictionary" (Kremer and Abel 2010). If properly employed, multimodal resources for e-dictionary examples may facilitate this process of blending. With space restrictions removed, it is likely for e-dictionaries to have an optimum number of examples showing a great variety of usage. When necessary, e-dictionaries can have options for longer examples than their paper counterparts, providing a complete co-text for users² (cf. Wojciechowska 2015). Meaning can be enriched in such a way as to improve the informativity and diversity of examples.

In addition, the meaning network in a dictionary is dynamic and interactive (cf. Prinsloo et al. 2012). For one thing, meaning tends to be co-constructed by the dictionary writers and users in the digital revolution. Users' participation in exemplification and their interaction with the dictionary have increased, and user-friendliness and example acceptability can be improved. Users can contribute examples, comment on them or make customized use of them. For another, the hypermodal affordances, like hyperlinks in examples, enable dictionary users to expand the meaning network outside the microstructure of a dictionary, thus meaning "flows". E-dictionary examples are characterized by hypermodality and hypertextuality³, allowing for more reading paths.

The paper argues that multimodal exemplification has the potential of expanding meaning on demand. Expansion means not only enrichment (informativity and diversity) but also flow (dynamicity or interactivity) in the semantic network. In previous literature on MDA, meaning expansion may only refer to the former. Meaning refers to Halliday's (1985) three metafunctions: ideational, interpersonal and textual⁴.

Drawing evidence mainly from online English learner's dictionaries, the article addresses the following research questions:

- (1) Is it necessary to optimize meaning representation by e-dictionary examples from a multimodal perspective?

- (2) Can multimodal exemplification expand the ideational, interpersonal and textual meanings in the e-dictionary discourse? If yes, how?

2. Background

2.1 Exemplification limitations of paper dictionaries

Paper dictionaries are often subject to space restrictions, with their layout confined by the alphabetic order. Unlike the hypertext, the printed text only has a single conventional sequence. As a result, limitations of exemplification may arise in terms of number, variety and the way of presentation.

One contentious problem about the exemplification in paper dictionaries is what words should be exemplified and how many examples should be given in specific cases, with a seeming focus on collocations. Xu (2009: 156-166) argued that some words should be better exemplified in the Big Five dictionaries⁵, with illustrations of collocational and syntactic complexities of verbs. A survey by Zhang (2015: 89-92) showed there were not enough collocational examples in the mainstream English dictionaries. COBUILD decided on a policy of exemplifying as many words as possible (Fox 1987: 137), but this still can't guarantee consistency in exemplification, and Stein (2002: 206) pointed out the need of examples for some concrete nouns to show "morphological restriction on specific word-formation structures".

A second limitation involves the deixis (person, time or space). "The dictionary example cannot usually afford to look outside itself for complete elucidation" (Chen and Chung 2008) and in the fragmental context exhibited in an example sentence, deixis poses a problem for lexicographers (Xu 2009: 120-137). Pronouns and "neutralized" atemporal statements are used (Szende 1999). But in such cases, the cultural, stylistic and pragmatic information may be dispensed with even when it is necessary. Accordingly, authenticity, naturalness and meaning completeness of examples may decline.

A third limitation may lie in the way of displaying information and the users' access to examples. In such densely printed pages of text, reading is often linear and strictly coded (cf. Van Leeuwen 2005: 204). The one-time displaying in a fixed order might leave the users in a passive state of reception. There is a lack of customized presentations of information which are likely to encourage users' active participation in meaning making and improve the usability of dictionaries.

Exemplification limitations of paper dictionaries highlight some advantages of e-dictionaries: hybridization, more and better data, efficiency of access, customization, corpus integration, and user input (cf. Granger 2012). Freedom from the traditional paper format has removed constraints on size and format, paving the way for multi-faceted, flexible and rich representations of word meaning and use that have been unfeasible for print dictionaries (Fellbaum 2014). With the potential of displaying information in a more usable and user-friendly way, multimodal devices in e-dictionaries, if properly employed, can

optimize the meaning representation by examples.

2.2 Potential problems with current e-dictionaries

Previous literature about e-dictionaries shows that problems still exist in at least two areas, examples and multimodal features.

First, there is room for improvement in examples. Many online dictionaries, like *Wordnik*, show contemporary web-derived example sentences on the right side of the screen, supported, on the left, by definitions from a range of traditional dictionaries which don't reflect the recently-coined meaning in the examples (Rundell 2015). As noted by Frankenberg-Garcia (2012), there seem to be few dictionary examples that give clear contextual clues about meaning and help with grammar at the same time. "Language users seem to appreciate examples from dictionaries, but there is not much proof that the examples in dictionaries as they are today suffice" (Frankenberg-Garcia 2015).

Second, surveys showed that users were not satisfied with the multimodal features of e-dictionaries. According to Müller-Spitzer et al. (2012), features typical of digital dictionaries, specifically adaptability and multimedia, received the lowest ratings, and "we wanted to know how users assess innovative features, such as the use of multimedia data or the option of user-adaptive adjustment to an online dictionary" (Müller-Spitzer 2014: 6). After all, it is not easy for lexicographers to understand how to make proper use of them in specific e-dictionaries for specific users since related empirical research has provided an unclear picture.

Kaneta (2011) found dictionary interfaces influenced both the amount and the length of reference to illustrative examples, and users' choice of interfaces "depends on the quality of information available outside the folded elements" (Lew 2015), but Klosa et al. (2014) showed most users chose the expanded view as a default. Dziemianko (2015) proved that color influenced the speed and effectiveness of dictionary search as well as the retention of the retrieved information. Lew and Doroszewska (2009) found a negative effect of viewing animated images on vocabulary retention, but Kemmer (2014: 251-278) indicated a slight majority of users preferred pictorial information to verbal information, especially the younger generation (cf. Lew 2015). Lew et al. (2017) identified a surprising degree of balance in whether the picture or (verbal) definition was viewed first, and the fact that dictionary users normally viewed both. With mixed and even inconsistent findings, these studies highlighted the importance of using multimodal devices appropriately.

According to Gouws (2014), many lexicographic e-products were developed without any influence from innovative theoretical suggestions, and as a result current e-dictionaries often do not live up to the expectations of their users. It is necessary to devise theoretical models for e-dictionaries that focus on critical areas like the data to be included in these dictionaries, the structures to present and accommodate the data, and the way they should respond to the

needs of their target users (Gouws 2014). E-dictionaries should be characterized by customized presentations of lexicographical data, and the display of data must be redefined (cf. Gouws 2014; L'Homme and Cormier 2014).

The potential of multimodal resources has not been well exploited for exemplification. For the sake of dictionary users, it is necessary to optimize the examples in e-dictionaries from a multimodal perspective. Exploring how multimodal exemplification influences meaning representation in the e-dictionary discourse may serve as a preliminary step for the construction of theoretical models for e-dictionaries.

3. Expansion of ideational meaning

From this part on, the article will explore how multimodal exemplification can expand the three metafunctional meanings of the semiotic system, ideational, interpersonal and textual (Halliday 1985). Ideational semiotic structures construct the nature of events, the objects and participants, and the circumstances; interpersonal structures construct the nature of relationships among addressers and addressees, and viewers and the viewed; and textual structures concern the distribution of the information value or relative emphasis among different elements (cf. Kress and Van Leeuwen 2006: 15).

The inherent relationship between the three metafunctions and the criteria for examples may help us understand multimodal exemplification better. According to Xu (2009: 10), there are four major criteria, namely, informativity, intentionality, acceptability and cohesion⁶. The ideational metafunction concerns informativity, the interpersonal has much to do with intentionality and acceptability, and the textual involves cohesion.

It is found that ideational meaning can be expanded by multimodal exemplification in three major ways: meaning enrichment realized by multimodal examples (i.e. multimodal resources used for examples), meaning enrichment mediated by example-definition ties, and meaning flow facilitated by example hyperlinks.

3.1 Meaning enrichment by multimodal examples

When necessary, the quantity and variety of examples can be increased in an e-dictionary, and more particulars or nuances of vocabulary usage can be shown than in print dictionaries, enriching the ideational meaning. With the technique of natural language processing, diversified classifications of examples, like collocations, quotations and proverbs, can be extracted from databases. And they can be tailored to user needs, for instance, encoding or decoding purposes. Many e-dictionaries, like Oxford and Longman, have an example bank with various corpus sentences for users to choose from. Multimodal resources are indispensable for enriching the ideational meaning of examples, and audio and visual modes are mostly used.

A lot of e-dictionaries offer the recordings of examples. They can be human voice and synthesized speech. The sound speed can be adjusted by dictionary users, or sometimes there are two speed alternatives (slow or fast). There are also sound choices of gender (male or female) and language variety (e.g. British English or American English). Some e-dictionaries even provide authentic recordings clipped from movies, speeches or news programs, like *Youdao English Dictionaries* (YED, <http://dict.youdao.com/>) while others may give those from well-known literary works and mainstream textbooks, like *Fayu Zhushou French Dictionaries* (FZFD, <https://www.frdic.com/>).

The visual mode, including color, font, symbol and notation, can also be properly used to reinforce the ideational meaning. Figure 1 shows a part of the *love* entry in *Longman Dictionary of Contemporary English* (LDOCE). Several semiotic resources are used to reinforce the representation of collocation and idiomatic usage, like highlighting such an expression in its examples with an underline and boldface font when the mouse cursor hovers over it (see "love at first sight" in Figure 1). During this multimodal human-computer interaction, the solid line emerges, which may imply a lexical bundle as cross-modal metaphor. Its emerging process might give some readers a sense of formation of the bundle.

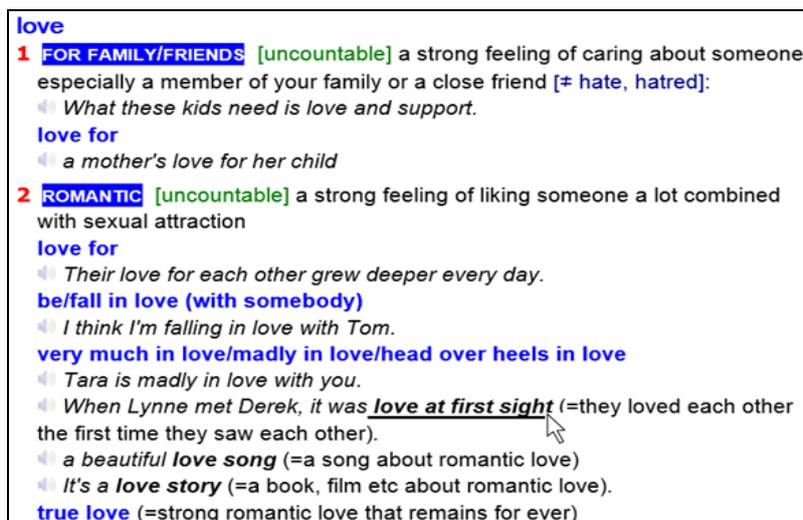


Figure 1: *Love* in LDOCE

Furthermore, there is simultaneous highlighting of parallel lexical units. For instance, in the bilingual examples of YED, any word and its Chinese/English counterpart will be concurrently highlighted with the mouse cursor hovering over it, as indicated by the match between "exploit" and "利用" in the second sentence in Figure 2 (screenshot from the entry of *learn*, <http://dict.youdao.com/search?le=eng&q=learn&keyfrom=dict.top>).



Figure 2: Learn in YED

The visual information can also be synchronized with the audio. A word in a sentence changes its color when its recording is played, achieving a kind of real-time cross-modal alignment like karaoke singing or movie subtitles. By representing the ideational meaning of example sentences in great detail, such visual-visual or visual-audio synchronization may facilitate the understanding of specific words or collocations, which is likely to improve exemplification effectiveness.

3.2 Meaning enrichment by example-definition ties

The enrichment of ideational meaning within the microstructure is investigated by looking at the cross-modal/intermodal ties between a definition and an example under the same sense. Identifying cohesive links and examining the logical relations in ideational meaning that extend across semiotic modes may help us gain an in-depth understanding of intermodal synergy, the cooperation and interaction among different modes (Royce 1998, 2007; Chan 2011: 144-167). According to the model of Chan (2011: 144-165), there are two types of intermodal relations, concurrence and complementarity. The (intermodal) concurrence or agreement denotes that meanings across modes are similar while (intermodal) complementarity suggests that meanings across modes are different but complement each other (ibid.). An online children's dictionary, *Wordsmyth Kid's English Dictionary (WKED)*, is used to illustrate the example-definition concurrence and complementarity.

The example-definition concurrence can be illustrated by the entry of *swan* in Figure 3 (screenshot from <http://kids.wordsmyth.net/wild/>). The picture is a non-verbal definition of the headword according to the concept of multimodal definition (Liu 2015). The example sentence describes what the picture is, indicating intermodal concurrence. So meanings across modes are similar and

in this way the example enriches the meaning of the verbal definition, indicating the difference in size between a swan and a duck or goose. At first sight, the dictionary users may see only two salient white swans in the foreground of the picture, and the example sentence can remind them of the small ducks and geese in the background (cf. Kemmer 2014: 251-278). That may constitute a two-step cognition process from a definition to examples.

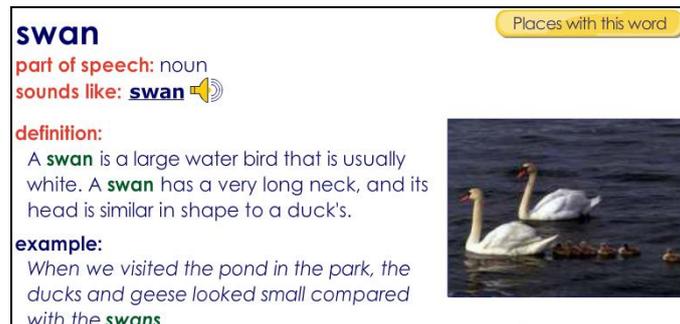


Figure 3: *Swan* in WKED

Furthermore, I'd like to show how the example-definition complementarity can be achieved with the entry of *owl* in Figure 4 (<http://kids.wordsmyth.net/wild/>). The verbal mode (language) and the visual mode (picture) of its definition provide visual information about the owl by depicting its physical appearance. Its example sentence offers auditory information by describing its sound verbally. So the definition and the example complement each other. However, if an audio file of the owl's crying or the example sentence were offered, that would achieve example-definition complementarity to a greater extent. In this case, the inclusion of pertinent sound and images as devices optimizing meaning representation is grounded on the principle of synaesthesia.

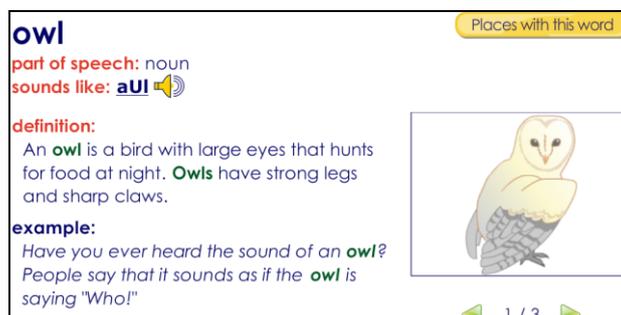


Figure 4: *Owl* in WKED

3.3 Meaning flow through example hyperlinks

Hyperlinks in e-dictionaries help users find and build semantic connections between words. They can be used on each word in an example to direct the user to another entry in or outside the dictionary. Such exemplification problems as collocations and deixis can be greatly eased. As the smartphone application (app) of *OALD (Oxford Advanced Learner's Dictionary)* says, "you can tap on any word within an entry to jump to the entry for that word". That enhances hypertextuality of the dictionary and creates meaning flow. So meaning is expanded from the microstructure to other parts of the dictionary, or even outside the dictionary.

Furthermore, there is a trend of linking and fusing dictionaries with other resources (Chen 2010), i.e. hybridization (cf. Granger 2012). A local app of *Fayu Zhushou French Dictionaries (FZFD)* in China offers example sentences from novels, movies and textbooks, with hyperlinks directing users to the source materials. Figure 5 shows its entry of *dans* with sentences respectively from *The Three Musketeers*, *Intermediate French Listening Textbook* and *Comprehensive French Textbook 2* in its smartphone app. The "traversals" between the semantic world within the dictionary and external learning resources are semiotically realized with the design of the page or screen (cf. Lemke 2005).



Figure 5: *Dans* in FZFD

4. Expansion of interpersonal meaning

Interpersonal meaning concerns two major parties: dictionary users and makers/writers. By user participation and interaction with the dictionary (makers),

interpersonal meaning flows. Since there are more design options for dictionary makers, interpersonal meaning can be enriched.

4.1 User participation and interaction as meaning flow

Compared with paper dictionaries, e-dictionaries provide users with more chances of participation and interaction, facilitating greater flexibility in example use, which can expand interpersonal meaning and improve the acceptability of examples. Information can be presented in an array of interlinked web pages and media networks, enabling e-dictionary users to navigate and choose their own pathways through this semantic "universe". Users can tailor the use according to their own needs. They can change dictionary settings (like interface style), make bookmarks to examples, and store favourite sentences in their own example bank. Taking smartphone or tablet dictionary apps as an example, due to their portability, they can work as flash cards for learners to carry with them. And they have cameras, so a user can add a picture to certain words or examples to aid memorization and retrieval if such apps are open-ended.

Many e-dictionaries invite users to contribute examples or make comments on them (see "user input" by Granger 2012; and "user-generated content" by Lew 2014). In the sentence example bank interface of *Jinshan Ciba English Dictionaries (iCIBA)*, there are user interaction buttons for reporting a mistaken example, praising a good one and saving a useful one. Figure 6 is a screenshot of examples of *love* (<http://dj.iciba.com/love1.html#>).



Figure 6: User interaction buttons in *iCIBA*

In Figure 6, the "Learn" button at the end of each example sentence indicates a

timed activity of memorizing an example. When triggering it, the user will be invited to put in order the words in an example sentence in 40 seconds (see Figure 7). The time counter urges the user to finish the task quickly, arousing a strong sense of engagement. Furthermore, in contrast to the cool and rational blue which may represent authority in the example page (see Figure 6), bright orange is used for the learning activity interface, which may connote such cultural meanings or personal emotions as enthusiasm, fascination, creativity, determination, and stimulation (Gage 1993).



Figure 7: Learning activity in *iCIBA*

4.2 Design options as meaning enrichment

With less pressure to save space, there are more design choices for e-dictionaries. This diversification is likely to reinforce the interpersonal metafunction, improving the intentionality of examples. Integrated with customization, it may enhance their acceptability. The page/screen layout (spatial mode) and the example genre style (verbal mode) are taken as examples for illustration.

The page layout for examples usually looks less "crowded" and cluttered in e-dictionaries. Individual examples often start on a new line while they tend to be run-on in traditional print dictionaries (Lew and De Schryver 2014). Contemporary designs may have large bands of space simply to suggest lightness of the reading experience and also create a sense of "room to think" or "contemplation" (Zhang et al. 2015). A case in point is online *Merriam-Webster Learner's Dictionary (MWLD)*. We find lots of white space for its example texts as the entry of *love* in Figure 8 shows (screenshot from <http://www.learnersdictionary.com/definition/love>). In contrast, the print dictionary design is filled with dense texts and communicates the opposite, providing comprehensive amounts of detailed information fulfilling its role as an informer.

2 [noncount] : attraction that includes sexual desire : the strong affection felt by people who have a romantic relationship

- a declaration of *love*
- He was just a lonely man looking for *love*.
- romantic *love*

[+] [more examples](#)

— often used before another noun

- a *love* poem/song/letter
- a *love* potion
- a *love* scene/story

— see also [FREE LOVE](#), [PUPPY LOVE](#)

◇ When people begin to feel romantic love for each other, they **fall in love**.

- They *fell* (madly/passionately) *in love* (with each other).

◇ This phrase is also used figuratively.

- She *fell in love* with sailing the first time she tried it.

Figure 8: *Love* in *MWLD*

The verbal mode of examples in e-dictionaries can have more choices for genre styles. *MWLD*, for instance, employs a conversational style. It makes extensive use of instructions and explanations about word usage for different examples. As Figure 8 shows, such instructions as "often used before another noun" and "when people begin to feel..." help enrich interpersonal meaning because they sound like tips from a thoughtful guide or a patient helper. This communicative style suggests a sense of accessibility on behalf of the speaker and of engagement with a reader presumed to have needs (Zhang et al. 2015).

Similar to the conversational style, there is a story style for *WKED*. The examples for each definition constitute a story in a continuous text unit (see Figures 3 and 4). The story style, together with the image on its right, may be very suitable for children's dictionaries for having advantages of readability and user-friendliness. Besides the interpersonal metafunction, this style may also help reinforce the textual metafunction, both coherence in content and cohesion in form. In contrast, the discrete examples in other dictionaries are relatively independent semantic units.

In a word, the digital transformation has changed the relations between dictionary writers and users, from one where authoritative experts presented information to a passive public to one of engagement and participation. This shift to user-friendliness might be driven by the social force of marketization (Zhang et al. 2015), attracting funding from advertisements to maintain the e-dictionary.

5. Expansion of textual meaning

This part analyzes how textual meaning is expanded in the multimodal example discourse in relation to its neighbouring elements. Information value, salience and framing are three interrelated systems in exploring textual meaning (Kress and Van Leeuwen 2006: 177). Information value refers to how certain elements are placed in a space to give them a particular value, which is closely related to composition types (ibid.). Salience means that elements are made to attract the viewer's attention to different degrees, as realized by such factors as placement in the foreground or background, relative size, contrasts in tonal value (or color), and differences in sharpness, etc (ibid.). Framing is the use of boundaries connecting or disconnecting spaces (cf. ibid.).

5.1 Information value and composition as meaning flow

According to Kress and Van Leeuwen (2006: 177-201), there are three main types of composition: vertical (top-bottom, Ideal-Real), horizontal (left-right, Given-New) and circular (Center-Margin, nucleus-subservient). They are used for different purposes on the e-dictionary screen.

On a vertical axis, elements at the top are often described as ideal, and the generalized essence of information and therefore usually most salient; elements placed at the bottom are often represented as real, and present more specific, factual information (cf. Leijon 2016). In the microstructure, a definition is usually placed above the examples under the same sense, forming a typical vertical composition. The definition could be ideal information while examples could be real. In the example text, the vertical composition is most often used. While consulting an e-dictionary, the user usually scrolls down from a definition to the examples, forming a top-bottom reading path. The example text can be further divided into several vertical compositions. In many dictionaries, examples are classified into groups of collocation patterns. Each pattern is followed by specific phrase or sentence examples (see Figure 1). The same rule of information value applies here.

There is a distinctive subtype of the vertical composition for examples in smartphone or tablet dictionary apps. In *FZFD* (see Figure 5), for instance, elements are mostly placed into equally sized tiles which could be swiped across to see more. This form of organization allows contents to be textually linked as choices of the same order since tiles of the same size also create textual linking or rhyming, alongside that accomplished by color and fonts (Zhang et al. 2015). In this sense, the textual meaning flows rhythmically and this kind of harmony may be important for small-screen reading.

In a horizontal composition, elements on the left could be given information, and those on the right could be new as we tend to read from left to right. This type of composition is found in many e-dictionaries. For instance, in *WKED* (see Figures 3 and 4), the verbal text of the definition and examples is

placed on the left while the picture is on the right. Compared with the picture, the verbal text is given information. The image plays a distinctive role as bearing witness. As a non-verbal definition device (cf. Liu 2015), the picture is mainly aligned with the verbal definition. But the bottom of the picture is aligned with the top of the example text. Such a nuance in composition symbolizes that examples, as a complement to the definition, will describe the picture. So the position of the picture, if carefully chosen, can embody reading tips for dictionary users.

The circular composition is often used if we take a look outside the micro-structure. Elements in a space could be located in a circular relation along the dimensions of Center (the nucleus) or Margin (more subservient) (Van Leeuwen 2005: 208). As far as a whole page of an e-dictionary is concerned, the definitions and examples are usually placed in the center, with additional information (usually hyperlinks) put around them. A triptych, a three-part structure, is often seen in such a circular layout.

On the basis of Kress and Van Leeuwen (2006: 177-201), Figure 9 is drawn to illustrate the above three composition types with their information value rules, which constitutes a typical design of visual space in e-dictionaries.

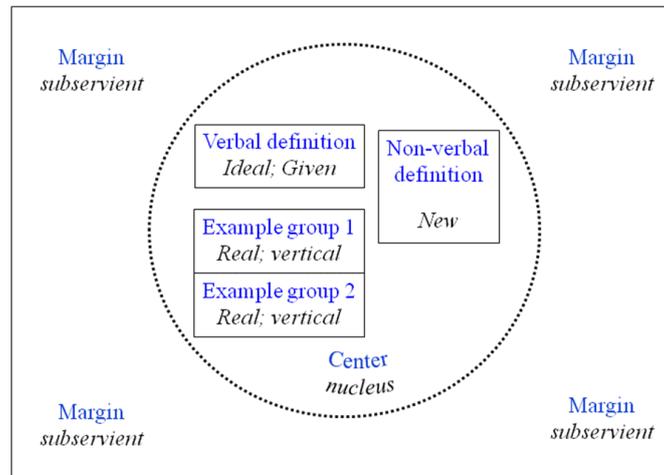


Figure 9: A typical design of visual space in e-dictionaries

In terms of information value, the textual meaning usually flows from Ideal to Real, from Given to New, and from nucleus to subservient, forming an organic network of textual cohesion in e-dictionaries.

5.2 Salience and framing as meaning enrichment

According to Kress and Van Leeuwen (2006: 201), salience can create a hierar-

chy of importance among the elements, selecting some as more important, more worthy of attention than others. The Given may be more salient than the New, for instance, or the New more salient than the Given, or both may be equally salient. And the same applies to Ideal and Real and to Center and Margin. As illustrated above, in a typical visual design of e-dictionaries, more salience may be intended for the Given (verbal), Ideal (definition or collocation heading) and Center (definition and example) although psychological salience depends on individual users in specific cases.

For salience and framing, various semiotic devices can be coordinated such as color, font and line (cf. Liu and Qu 2014). *LDOCE* (see Figure 1) is characterized by the use of striking blue with high saturation and boldface font for the collocation heading, which helps framing, symbolizing more salience and more information value on collocation patterns (i.e. fixed phrase examples standing for generality) than sentence examples (embodying specificity). Instead, *MWLD* (see Figure 8) extensively uses icons (squares, addition marks, dashes and diamonds etc.) for framing. *FZFD* (see Figure 5) uses lines instead. The different styles may cater for different cognitive styles of dictionary users.

In addition, foregrounding and backgrounding are used on an e-dictionary screen. For instance, there is a practical function of hiding examples which may help optimize textual cohesion in e-dictionaries. A case in point is the offline desktop version of *Jinshan Ciba English Dictionaries (JCED)* where the examples following each definition can be hidden. The "hide" (top of Figure 10) and "unhide" (bottom of Figure 10) functions give users chances of arranging the interface layout for their own sake. This customization device may facilitate their cognitive process and also enhance the interpersonal metafunction as well.

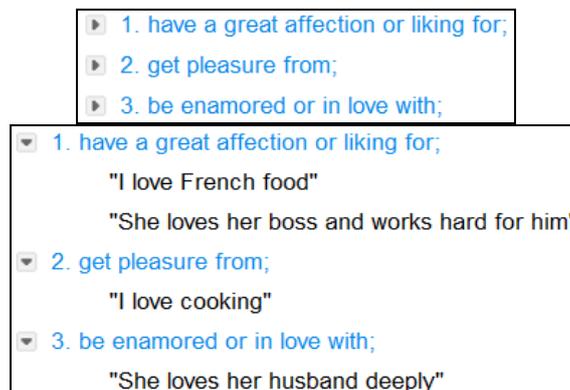


Figure 10: Hide and unhide functions in *JCED*

With hierarchies of salience and devices of framing, the example texts in e-dictionaries allow for more reading paths than paper dictionaries, and hence for the heterogeneity and diversity of their large readership (cf. Van Leeuwen

2005: 205). Unlike movies or exhibitions, e-dictionary example texts are non-linear. Non-linear texts often impose a paradigmatics (instead of syntagmatics), leaving the reader to sequence and connect the elements presented (Kress and Van Leeuwen 2006: 208). In the design of such texts there will be pressure to put more of the meaning in the individual elements of the composition, and to use more self-contained items of information (ibid.).

6. Discussion and conclusion

6.1 Main findings and implications

This study was inspired by Lew (2010). Following my previous work on multimodal definition, the article tentatively proposes a new term, multimodal exemplification. A holistic view of meaning and a systemic-functional approach to the semiotic system are adopted for multimodal discourse analysis. Complementing previous literature, the study has the following main findings.

First, the concept of meaning expansion in the field of MDA is extended, with evidence that multimodal exemplification can expand meaning by enriching it and making it "flow" in the semantic network. In this process of expansion, the three metafunctions of the semiotic system can be reinforced: ideational, interpersonal and textual.

Second, meaning can be enriched not only by multimodal examples per se but also by example-definition ties. To the best of my knowledge, this is the first attempt to explore the intermodal relationship between a definition and the examples under the same sense.

Third, the dynamicity and interactivity of e-dictionary example texts enable user participation, and freedom from the traditional paper format allows for more design options, expanding the interpersonal meaning and improving the acceptability and intentionality of examples. Users can enjoy flexibility and customization while creating their own reading paths. Regarding the verbal mode of examples, I identified two distinctive genre styles (the conversational style and the story style), supporting the claim that dictionaries are "discourses in their own right" and they are "divergent and intertwining" (Chen 2015).

Fourth, concerning the example texts, I analyzed the three main composition types with corresponding information value rules, presenting a diagram illustrating a typical design of visual space (see Figure 9). This may be the first endeavor to do so in lexicography.

Last, smartphone dictionary app pages often have a vertical composition of equally-sized tiles, which creates choices of the same order, and textual linking or rhyming. With such findings, the current study echoes with research on other types of multimodal texts, such as science communication websites (Zhang et al. 2015).

The current study has two implications for lexicography and multimodal

discourse studies. One implication is that the uniqueness and diversity of e-dictionary example texts are central to the critical analysis and understanding of the multimodal discourse. Different types of e-dictionaries have distinctive features, and each needs a close look separately (cf. L'Homme and Cormier 2014).

The other implication is that an integrative and panoramic perspective should be taken to broaden the horizons of multimodal lexicography. For one thing, different modes and their semiotic devices should be coordinated to make new meanings. For another, the three metafunctions are intrinsically interrelated and can be seamlessly integrated in example texts. Color, for instance, is multi-functional for it can express ideational meaning, arouse personal emotions, and also serve as a tool for separation and framing (cf. Kress and Van Leeuwen 2006: 230).

6.2 Limitations and future research

Largely relying on introspective analysis, this study has some limitations. Empirical research into user needs should be conducted to support the arguments stated in the article. More investigations are needed into the mode selection and intermodal synergy for e-dictionary examples. More data should be collected from smartphone and tablet dictionary apps in the global wave of mobile assisted language learning. This leaves much room for future work.

There are two main directions for future research, user research and (multimodal) lexical database, which have become the foci of lexicographical research in the digital revolution (cf. Thieberger 2011; Fellbaum 2014; Lew 2015; Lew et al. 2017). First and foremost, analysis should be made to uncover how such multimodality impacts upon users' experiential and cognitive encounters with these examples. User research should focus on user needs in specific types of situations, which may lay a foundation for customized example classifications (cf. Bergenholtz and Bothma 2011; Potgieter 2012). Furthermore, multimodal lexical databases should be built on the basis of multimodal corpora and be finely tuned to a variety of parameters or facets of user customization (cf. Kipp et al. 2009; Bergenholtz and Nielsen 2013). These two directions of future research may help build a solid empirical base for a theoretic model of multimodal lexicography.

In summary, the research questions proposed at the beginning of this article have been answered. As an important part of multimodal lexicography, multimodal exemplification deserves scholarly attention since it can bring both chances and challenges for meaning representation.

Endnotes

1. Multimodal definition is (the act of) meaning explanation of a word or phrase with multimodal devices for achieving better defining effect than language does alone, especially in an e-dictionary (Liu 2015).

2. The number and complexity of examples should be increased with precautions against information overload (cf. Gouws and Tarp 2016), though. Dictionary users want information fast.
3. Hypermodality is more than multimodality in just the way that hypertext is more than plain text, and it is not simply that we juxtapose image, text, and sound; we design multiple interconnections among them, both potential and explicit (Lemke 2002). In the simplest form of hypertext, we might have a web of "pages" (or paragraphs, sentences, or even single words) in which the whole or some part of the page was linked to the whole or some part of another page in some way other than by the default sequential convention of ordinary reading (ibid.).
4. A communicative system simultaneously fulfills three metafunctions: the ideational, the function of constructing representations of the world; the interpersonal, the function of enacting (or helping to enact) interactions characterized by specific social purposes and specific social relations; and the textual, the function of marshalling communicative acts into larger wholes, into the communicative events or texts (Kress and Van Leeuwen 2006: 228).
5. The Big Five dictionaries refer to the following English learner's dictionaries: Oxford, Longman, Collins COBUILD, Cambridge and Macmillan.
6. Xu (2009: 9-10) tailored De Beaugrande's (1980: 6) seven criteria for textuality to be four criteria for illustrative examples, viz. intentionality, informativity, cohesion and acceptability. Intentionality concerns lexicographic intentions behind a dictionary example; informativity means the types and amount of linguistic information shown; cohesion involves the ties between the dictionary example and its source text, and the links between the example and the dictionary text; and acceptability refers to the degree to which an example satisfies its target users (ibid.).

Acknowledgements

This work was supported by the China Scholarship Council [201706155083], the Philosophy and Social Science Fund of the 12th Five-year Plan of Guangdong Province [GD15XWW23], and the Philosophy and Social Science Fund of the 13th Five-year Plan of Guangzhou City [2016GZGJ68]. I would like to express my deep gratitude to Prof. Robert Lew of Adam Mickiewicz University in Poznań for his discussion about multimodal lexicography during his visit to China in 2015. Furthermore, I am obliged to Prof. Paul van den Hoven of Utrecht University for analyzing the problems in an early draft of this paper. Special thanks are also due to Prof. Danie J. Prinsloo and the anonymous reviewers for their constructive comments.

References

A. Dictionaries

- COBUILD.** 1995. Sinclair, John M. (Ed.). 1995. *Collins COBUILD English Language Dictionary*. Second edition. London/Glasgow: HarperCollins.
- FZFD.** *Fayu Zhushou French Dictionaries* (smartphone app from <http://rj.baidu.com/soft/detail/11669.html?ald>). Accessed on February 13, 2016.

- iCIBA.** *iCIBA Sentence Example Bank*. <http://dj.iciba.com>. Accessed on February 8, 2016.
- JCED.** *Jinshan Ciba English Dictionaries* (offline desktop version downloaded from <http://tj.baidu.com/soft/detail/11908.html?ald>). Accessed on February 15, 2016.
- LDOCE.** *Longman Dictionary of Contemporary English*. Accessed on February 8, 2016. <http://www.ldoconline.com/dictionary>.
- MWLD.** *Merriam-Webster Learner's Dictionary*. Accessed on February 15, 2016. <http://www.learnersdictionary.com/>.
- OALD.** *Oxford Advanced Learner's Dictionary* (smartphone app). Eighth edition. Accessed on February 15, 2016.
- WKED.** *Wordsmyth Kid's English Dictionary*. Accessed on February 10, 2016. <http://www.wordsmyth.net/wild>.
- Wordnik.** <https://www.wordnik.com/>.
- YED.** *Youdao English Dictionaries*. Accessed on February 10, 2016. <http://dict.so.163.com/>. <http://dict.youdao.com>.

B. Other literature

- Atkins, B.T. Sue.** 1996. *Bilingual Dictionaries: Past, Present and Future*. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Rödger Pappmehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.
- Bergenholtz, Henning and Theo J.D. Bothma.** 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.
- Bergenholtz, Henning and Jesper Skovgård Nielsen.** 2013. What is a Lexicographical Database? *Lexikos* 23: 77-87.
- Chan, Eveline.** 2011. Integrating Visual and Verbal Meaning in Multimodal Text Comprehension: Towards a Model of Intermodal Relations. Dreyfus, Shoshana, Susan Hood and Maree Stenglin (Eds.). 2011. *Semiotic Margins: Meaning in Multimodalities*: 144-167. London: Continuum.
- Chen, Chih-Ming and Ching-Ju Chung.** 2008. Personalized Mobile English Vocabulary Learning System Based on Item Response Theory and Learning Memory Cycle. *Computers & Education* 51(2): 624-645.
- Chen, Wenge.** 2015. Bilingual Lexicography as Recontextualization: A Case Study of Illustrative Examples in a New English-Chinese Dictionary. *Australian Journal of Linguistics* 35(4): 311-333.
- 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.
- De Beaugrande, Robert.** 1980. *Text, Discourse, and Process: Toward a Multidisciplinary Science of Texts*. London: Longman.
- Dziemianko, Anna.** 2015. Colours in Online Dictionaries: A Case of Functional Labels. *International Journal of Lexicography* 28(1): 27-61.
- Fellbaum, Christiane.** 2014. Large-scale Lexicography in the Digital Age. *International Journal of Lexicography* 27(4): 378-395.
- Fox, Gwyneth.** 1987. The Case for Examples. Sinclair, John M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London/Glasgow: Collins ELT.

- Frankenberg-Garcia, Ana.** 2012. Learners' Use of Corpus Examples. *International Journal of Lexicography* 25(3): 273-296.
- Frankenberg-Garcia, Ana.** 2015. Dictionaries and Encoding Examples to Support Language Production. *International Journal of Lexicography* 28(4): 490-512.
- Gage, John.** 1993. *Color and Culture: Practice and Meaning from Antiquity to Abstraction*. Boston: Bulfinch Press.
- Gouws, Rufus H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Gouws, Rufus H. and Sven Tarp.** 2016. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* (Advance access). <https://doi.org/10.1093/ijl/ecw030>.
- Granger, Sylviane.** 2012. Electronic Lexicography: From Challenge to Opportunity. Granger, Sylviane and Magali Paquot (Eds.). 2012. *Electronic Lexicography*: 1-11. Oxford: Oxford University Press.
- Halliday, Michael A.K.** 1985. *An Introduction to Functional Grammar*. London: Edward Arnold.
- Kaneta, Taku.** 2011. Folded or Unfolded: Eye-tracking Analysis of L2 Learners' Reference Behaviour with Different Types of Dictionary Interfaces. Akasu, Kaoru and Satoru Uchida (Eds.). 2011. *Asialex2011 Proceedings Lexicography: Theoretical and Practical Perspectives*: 219-224. Kyoto: Asian Association for Lexicography.
- Kemmer, Katharina.** 2014. Rezeption der Illustration, jedoch Vernachlässigung der Paraphrase? Müller-Spitzer, Carolin (Ed.). 2014. *Using Online Dictionaries*: 251-278. Berlin/Boston: Walter de Gruyter.
- Kipp, Michael, Jean-Claude Martin, Patrizia Paggio and Dirk Heylen (Eds.).** 2009. *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*. Berlin/Heidelberg: Springer.
- Klosa, Annette, Alexander Koplenig and Antje Töpel.** 2014. Benutzerwünsche und -Meinungen zu dem monolingualen deutschen Onlinewörterbuch Elexiko. Müller-Spitzer, Carolin (Ed.). 2014. *Using Online Dictionaries*: 281-384. Berlin/New York: Walter de Gruyter.
- Kremer, Gerhard and Andrea Abel.** 2010. Semantic Relations in Cognitive eLexicography. Dykstra, Anne and Tanneke Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6-10 July 2010*: 380-388. Leeuwarden: Afûk.
- Kress, Gunther and Theo van Leeuwen.** 2006. *Reading Images: The Grammar of Visual Design*. Second edition. London: Routledge.
- Leijon, Marie.** 2016. Space as Designs for and in Learning: Investigating the Interplay between Space, Interaction and Learning Sequences in Higher Education. *Visual Communication* 15(1): 93-124.
- Lemke, Jay L.** 2002. Travels in Hypermodality. *Visual Communication* 1(3): 299-325.
- Lemke, Jay L.** 2005. Multimedia Genres and Traversals. *Folia Linguistica* 39(1-2): 45-56.
- Lew, Robert.** 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.
- Lew, Robert.** 2014. User-generated Content (UGC) in English Online Dictionaries. Abel, Andrea and Annette Klosa (Eds.). 2014. *Der Nutzerbeitrag im Wörterbuchprozess 3. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*. (Opal — Online publizierte Arbeiten zur Linguistik): 8-26. Mannheim: Institut für Deutsche Sprache.
- Lew, Robert.** 2015. Research into the Use of Online Dictionaries. *International Journal of Lexicography* 28(2): 232-253.

- Lew, Robert and Gilles-Maurice de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
- Lew, Robert and Joanna Doroszewska.** 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.
- Lew, Robert, Rafał Kaźmierczak, Ewa Tomczak and Mateusz Leszkowicz.** 2017. Competition of Definition and Pictorial Illustration for Dictionary Users' Attention: An Eye-tracking Study *International Journal of Lexicography* (Advance access). <https://doi.org/10.1093/ijl/ecx002>.
- L'Homme, Marie-Claude and Monique C. Cormier.** 2014. Dictionaries and the Digital Revolution: A Focus on Users and Lexical Databases. *International Journal of Lexicography* 27(4): 331-340.
- Liu, Xiqin.** 2015. Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos* 25: 210-232.
- Liu, Xiqin and Dianning Qu.** 2014. Exploring the Multimodality of EFL Textbooks for Chinese College Students: A Comparative Study. *RELC Journal* 45(2): 135-150.
- Müller-Spitzer, Carolin (Ed.).** 2014. *Using Online Dictionaries*. Lexicographica Series Maior 145. Berlin: Walter de Gruyter.
- Müller-Spitzer, Carolin, Alexander Kopenig and Antje Töpel.** 2012. Online Dictionary Use: Key Findings from an Empirical Research Project. Granger, Sylviane and Magali Paquot (Eds.). 2012. *Electronic Lexicography*: 425-457. Oxford: Oxford University Press.
- Nielsen, Sandro.** 2014. Example Sentences in Bilingual Specialised Dictionaries Assisting Communication in a Foreign Language. *Lexikos* 24: 198-213.
- O'Halloran, Kay L.** 2008. Systemic-functional Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning Using Language and Visual Imagery. *Visual Communication* 7(4): 443-475.
- O'Halloran, Kay L.** 2011. Multimodal Discourse Analysis. Hyland, Ken and Brian Paltridge (Eds.). 2011. *The Continuum Companion to Discourse Analysis*: 120-137. London: Continuum.
- Potgieter, Liezl.** 2012. Example Sentences in Bilingual School Dictionaries. *Lexikos* 22: 261-271.
- Prinsloo, Danie J.** 2015. Corpus-based Lexicography for Lesser-resourced Languages — Maximizing the Limited Corpus. *Lexikos* 25: 285-300.
- Prinsloo, Danie J., Ulrich Heid, Theo Bothma and Gertrud Faaß.** 2012. Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22: 290-320.
- Royce, Terry.** 1998. Synergy on the Page: Exploring Intersemiotic Complementarity in Page-Based Multimodal Text. *JASFL Occasional Papers* 1(1): 25-49.
- Royce, Terry.** 2007. Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis. Royce, Terry and Wendy Bowcher (Eds.). 2007. *New Directions in the Analysis of Multimodal Discourse*: 63-109. Mahwah, NJ: Lawrence Erlbaum.
- Rundell, Michael.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Stein, Gabrielle.** 2002. *Better Words: Evaluating EFL Dictionaries*. Exeter: Exeter University Press.
- Szende, Thomas.** 1999. Problems of Exemplification in Bilingual Dictionaries. *Lexicographica* 15: 198-228.
- Taljard, Elsabé.** 2015. Collocations and Grammatical Patterns in a Multilingual Online Term Bank. *Lexikos* 25: 387-402.
- 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.

- Thieberger, Nick.** 2011. Building a Lexical Database with Multiple Outputs: Examples from Legacy Data and from Multimodal Fieldwork. *International Journal of Lexicography* 24(4): 463-472.
- Van Leeuwen, Theo.** 2005. *Introducing Social Semiotics*. London/New York: Routledge.
- Wittgenstein, Ludwig.** 1953. *Philosophical Investigations*. Oxford: Blackwell.
- Wojciechowska, Sylwia.** 2015. Collocations and Colligations of the Metonymic *Shipment*: Exemplification in Learners' Dictionaries vs. Corpus Evidence. *International Journal of Lexicography* 28(4): 465-489.
- Xu, Hai.** 2009. *Towards Prototypical Exemplification in English Dictionaries for Chinese EFL Learners*. Beijing: The Science Press.
- Zhang, Yihua.** 2015. *Second Language Acquisition and Learner's Dictionaries*. Beijing: The Commercial Press.
- Zhang, Yiqiong, David Machin and Tao Song.** 2015. Visual Forms of Address in Social Media Discourse: The Case of a Science Communication Website. *Journal of Multicultural Discourses* 10(2): 236-252.

Koalas, Kiwis and Kangaroos: The Challenges of Creating an Online Australian Cultural Dictionary for Learners of English as an Additional Language

Julia Miller, *School of Education, The University of Adelaide,
Australia (julia.miller@adelaide.edu.au)*

Deny A. Kwary, *Department of English Literature, Airlangga University,
Indonesia (kwary@yahoo.com)*

and

Ardian W. Setiawan, *The State Polytechnic Manufacture Bangka Belitung,
Sungai Liat, Indonesia (ardian.setiawan@gmail.com)*

Abstract: This article reports on an online cultural dictionary for learners of English as an Additional Language (EAL) in Australia. Potential users studying English for academic purposes in an Australian university pre-entry program informed each stage of the dictionary's creation. Consideration was given to the need for such a dictionary; terms to be included; information necessary for each entry (including audio and visual material); use of a limited defining vocabulary; example sentences; notes on each term's usage; and evaluation of user feedback once the dictionary had been launched online. Survey data indicate that users particularly value the dictionary's ease of use, example sentences, and specifically Australian content (including pronunciation given in an Australian accent). It is suggested that more entries be added, and that cultural dictionaries be created for other varieties of English, as well as for other languages.

Keywords: AUSTRALIAN, CULTURE, DICTIONARY, ENGLISH AS AN ADDITIONAL LANGUAGE, LEARNER'S DICTIONARY, ONLINE

Opsomming: Koalas, kiwi's en kangaroes: Die uitdagings in die skep van 'n aanlyn Australiese kulturele woordeboek vir aanleerders van Engels as 'n addisionele taal. In hierdie artikel word verslag gedoen oor 'n aanlyn kulturele woordeboek vir aanleerders van Engels as 'n Addisionele Taal (EAT) in Australië. Potensiële gebruikers wat Engels vir akademiese doeleindes in 'n Australiese universiteitstoelatingsprogram studeer, het die inligting vir elke fase in die skep van die woordeboek verskaf. Daar is oorweging geskenk aan die behoefte aan so 'n woordeboek; terme wat ingesluit moet word; inligting wat benodig word vir elke inskrywing (insluitend oudio- en visuele materiaal); die gebruik van 'n beperkte definiëringswoordeboek; voorbeeldsinne; notas oor die gebruik van elke term; en die evaluering van gebruikstersterugvoer nadat die woordeboek aanlyn verskyn het. Data verkry uit die vraelyste dui daarop

dat gebruikers spesifiek waarde heg aan die gebruikersvriendelikheid, voorbeeldsinne, en spesifiek Australiese inhoud (insluitend uitspraak gegee in 'n Australiese aksent). Daar word voorgestel dat meer inskrywings bygevoeg word, en dat kulturele woordeboeke geskep word vir ander variëteite van Engels, sowel as vir ander tale.

Sleutelwoorde: AUSTRALIES, KULTUUR, WOORDEBOEK, ENGELS AS ADDISIONELE TAAL, AANLEERDERSWOORDEBOEK, AANLYN

Background

Studying in another country usually requires a good command of that country's language. For example, a speaker of English as an additional language (EAL) who wants to study in an Australian university will usually need to obtain a minimum International English Language Testing System (IELTS) score of 6. However, this level of English does not prevent the student from having problems when encountering English words in daily life. For example, on arriving in Australia they might look for a meal and see the word *brekkie* on a board outside a café. Not knowing the meaning of *brekkie*, they consult a dictionary but cannot find an entry. Indeed, *brekkie* appears in only one of the major English learner's dictionaries (the *Oxford Advanced Learner's Dictionary*). They would also be unlikely to find the word in a bilingual dictionary. *Brekkie* in fact is the Australian slang word for *breakfast*. Even on deciphering this meaning, however, further problems arise: what constitutes breakfast in Australia, and when do people eat this meal?

In these days of easy Internet access (in many countries, though not necessarily throughout the vastness of outback Australia), it is relatively simple to search for a word online, using either a computer, a tablet or a mobile phone. Even then, however, the student may run into problems. *Urban Dictionary* is likely to show up first in a search for slang words such as *brekkie* (though of course the student does not initially know that *brekkie* is slang). The definitions in *Urban Dictionary* may include accurate explanations, but these can be obscured by misleading information, inaccurate spelling, or highly colourful language. *Brekkie*, for example, has three *Urban Dictionary* entries, each with an example sentence incorporating its use (given here in brackets):

1. Abbreviation of breakfast. (*Couldn't be arsed to eat brekkie this morning.*)
 2. An Australian slang term for breakfast. (*I had eggs and tomato on toast for brekkie today.*)
 3. A person who is obsessed with the breakfast club. Kinda like a trekkie. But insted a brekkie. (*Ashley is obsessed with the breakfast club i consider her as a brekkie.*)
- (*Urban Dictionary*, accessed on 3 December 2015)

The second definition here is accurate, and uses a helpful example which elaborates on the content of a possible Australian breakfast. The first definition is adequate, although the example sentence is confusing and contains a taboo

word. The third definition is confusing, badly spelled, and uninformative. Other common Australian words fare even worse on this site, with the innocent *koala* receiving 27 definitions, including the alarming and misspelt:

A very dangerous mammal that is prone to attack. Also poisonous, these are deadly beast which we must protect ourselves from. Even now many officials are considering bombing all koala habits to destroy these dangerous beasts. Alternate definiton: a whore.

It is evident, then, that students should be guided towards trustworthy dictionaries, and in fact all the six major English learner's dictionaries are freely available online: *Cambridge Advanced Learner's Dictionary (CALD)*, *Collins COBUILD Advanced Dictionary (COBUILD)*, *Longman Dictionary of Contemporary English (LDOCE)*, *Macmillan English Dictionary for Advanced Learners (MEDAL)*, *Merriam-Webster Learner's Dictionary (MWLD)*, and *Oxford Advanced Learner's Dictionary (OALD)*. Unfortunately, however, these dictionaries are mainly aimed at the British or American English markets, and fail to include many terms used in other English-speaking countries. Australia used to have a dictionary for advanced learners, the excellent *Macquarie Learner's Dictionary* (1999), but this is no longer published. The *Australian National Dictionary* (2016) and *Macquarie Dictionary* (2013) are designed for native speakers of English and do not contain all the information a learner may need. Moreover, although they are available online, they are not free to use. There is thus a need for a reliable dictionary for EAL learners in Australia, since Australian cultural references are frequently missing from the existing learner's dictionaries. A cultural dictionary of Australian terms for EAL learners would help to fill this gap, and a freely available online version would make such a dictionary easily accessible.

Literature review

In compiling a learner's dictionary, at least eleven factors need to be considered: dictionary medium; definition style; spelling variations; grammatical information; pronunciation guide; defining vocabulary; usage labels; example sentences; audiovisual material; language variety; and cultural context. All of these must relate to the needs of the proposed dictionary user, who in this case is an adult EAL learner in Australia.

— *Dictionary medium*

Online dictionaries are superseding paper dictionaries, and in some cases (e.g. Macmillan) publishers are no longer producing hard copy dictionaries and are only publishing online. This is unfortunate for those people who do not have Internet access. Nevertheless, for those producing a dictionary under a limited budget, the advantages of an online dictionary are manifold. There is no limit as to space, because there are no printing costs; audiovisual material can be

incorporated more cheaply; and it is easier to update the material in the dictionary. Despite these advantages, however, it is still important that users' needs be considered and that the lexicographer should not be carried away by the potential of the new online medium (Gouws 2011: 21).

There is debate over the terminology to describe online dictionaries. Lew and De Schryver (2014: 342-344) examine the terms 'electronic dictionary', 'e dictionary', 'digital media dictionary', 'digital dictionary', and 'online dictionary', and predict that the term 'online dictionary' may well become the term favoured by lexicographers and metalexicographers. Since the dictionary described in this paper is purely online, and this term is favoured by many writers, we have chosen the term 'online dictionary' to describe our work.

— *Definition style*

The style of definition in the leading English learner's dictionaries varies. While *CALD*, *LDOCE*, *MEDAL*, *MWLD*, and *OALD* all use an analytical format (Lew and Dziemianko 2006: 229) based around a phrase (e.g. *emphasis*: special importance that is given to something (*OALD* online 7 April 2015)), *COBUILD* is distinctive in using a sentence definition (*emphasis*: **Emphasis** is special or extra importance that is given to an activity or to a part or aspect of something (*COBUILD* online 7 April 2015)). The sentence definition has been criticized for its uneconomical length (Cowie 1999: 160). However, this is no longer a problem in an online dictionary.

— *Spelling variations*

One of the most common reasons for consulting a dictionary is to verify a word's spelling (e.g. Harvey and Yuill 1997: 259), and it is standard practice in dictionaries to provide spelling variations. These variations may reflect differences within the same variety of a language (e.g. *hello/hallo/hullo* in UK English) or differences between language varieties (e.g. *litre* in UK English and *liter* in American English).

— *Grammatical information*

Information on a word's part of speech and other grammatical features is an important element in a learner's dictionary (Zgusta 2006), although it is not clear how much users refer to this information (Bogaards 2001: 105), especially if it is given in codified form (Lew and Dziemianko 2006: 226). For example, the symbol 'U' might represent the fact that a noun is uncountable, but users may be unaware of this because they fail to consult the dictionary's user guide. In fact, research suggests that users gain grammatical information more by seeing a word used in an example sentence (Dziemianko 2006) than by following a dictionary code. These reservations notwithstanding, grammatical information

can be used to good effect by learners and teachers if they are aware of its existence and usefulness.

— *Pronunciation guide*

Learners' needs, and changes in English pronunciation according to the variety of English, have led to the inclusion of IPA characters in learner's dictionaries (Häcker 2012), compared to other systems of pronunciation, such as the use of diacritics and ordinary Roman alphabetic symbols (Fraser 1997). Online dictionaries can also provide audio files, and recent research (reviewed in Lew 2015) indicates that users highly value this feature, although Lew (2015) recommends that phonemic transcriptions also be included to raise awareness of phonemic contrasts with users' first languages.

— *Defining vocabulary*

One key pedagogical factor in the design of learner's dictionaries is the use of a limited defining vocabulary (DV), so that learners are not faced with incomprehensible words in definitions. The idea of such a vocabulary was first created by West and Endicott in *The New Method English Dictionary* of 1935 (Cowie 1999: 24). In that dictionary, 1490 words were used to define 23,898 entries (Cowie 1999: 24). Even such a limited list of defining words can pose problems for learners, however, and each DV word may itself need to appear as a headword (Cowie 1999: 24). Of the current advanced learner's dictionaries, *LDOCE* was the first to use a DV of around 2000 words (Procter et al. 1978: viii-ix), based on West's (1953) *General Service List of English Words*. Most of the other learner's dictionaries also use a controlled DV. *CALD* (2008) has 2000 words; *COBUILD* (2009) has 3197; *LDOCE* (2003: 1943) has 'around 2000 common words'; *MEDAL* (2002: 1677) has 'under 2500 words'; and *OALD* (2010: R43) uses 'keywords' from the Oxford 3000 list. *MWLD* does not have a defining vocabulary, but includes a list of 3000 core vocabulary words. There is a tendency to understate the number of words used in a DV (Cowie 1999: 110), and to list only one word (the lemma) in the DV when in fact three words are used in definitions within the dictionary. For example, *COBUILD* (online, 6 March 2015) lists only *accident* in its DV, but the words *accidental* and *accidentally* are used in the definitions of *collateral damage* and *bump*, respectively.

Another problem arises when a word is needed for a definition but falls outside the DV because it is of a technical nature. Three British learner's dictionaries (*MEDAL*, *LDOCE*, and *OALD*) use small capital letters for these extra words, with either an explanation in brackets immediately after the word or a link to the extra word's entry in the main dictionary. For example, *LDOCE* (2003: 1943) defines *kangaroo* as 'an Australian animal that moves by jumping and carries its babies in a POUCH (= a special pocket of skin) on its stomach'. Similarly, *MEDAL* (2002) defines *kangaroo* as 'a large Australian animal that moves by jumping, has strong back legs, and carries its baby in a POUCH (= pocket on

the front of its body)'. *Pouch* is a technical word that is explained in the *LDOCE* and *MEDAL* definitions because it is not commonly used and so is not part of the DV. *OALD* (2010: 844) reverses the explanation and use of the word *pouch* and defines *kangaroo* as 'a large Australian animal with a strong tail and back legs that moves by jumping. The female carries its young in a pocket of skin (called a POUCH) on the front of its body.' One might question the need for the word *pouch* in this definition, since it appears after its explanation ('a pocket of skin'). *CALD* omits any reference to a pouch. *COBUILD* (2009) and *MWLD* use the word *pouch* in their definitions, but do not highlight it or explain it in any way. They do, however, define *pouch* under its own separate headword, although this is not hyperlinked from *kangaroo*. These varied examples indicate the problem of using technical or more unusual words in definitions. However, there is a limit to how simple a definition needs to be. Xu (2012: 369), for example, criticizes the *LDOCE* definition of *tabasco* as 'a very hot-tasting liquid' for being 'unnatural'. Furthermore, there are some words, such as *marsupial*, which are essential to a definition and yet hard to explain. None of the *kangaroo* definitions above includes this word, and yet, to be accurate, it is important to distinguish a mammal from a marsupial, particularly in a country such as Australia which has many marsupials.

— *Usage labels*

Labels in dictionary entries enable users to know the part of speech, as well as the register, frequency, and context of a word in everyday discourse. Words that are not labeled are considered to represent standard usage (Kipfer 1984: 140), but the application of labels to entries is often problematic. Currency and frequency of use, for example, may change according to the age of the user. Furthermore, labels such as *old-fashioned* are used inconsistently in the advanced learner's dictionaries (Miller 2011), sometimes indicating that a word is used mainly by older people and at other times indicating that the word is passing out of use altogether. At the opposite end of the age spectrum, many new words used by younger age groups may prove to be ephemeral, and it is hard to keep track of these in a dictionary, although online versions make the updating process easier. Context is also important: is a word used by everyone in all circumstances, or is its use restricted in some way? It is important for learners to be aware of such details and to know which words are used in which circumstances by a certain age group (Atkins and Rundell 2008: 229). Currency and usage can be indicated in a dictionary by means of a usage label and further portrayed in the example sentences. Frequency of use information requires corpus data in order to be accurate.

— *Example sentences*

Sentences which exemplify usage are vital for learners, in that they not only explain but also model native speaker use of a language (Sinclair et al. 1987: xv).

For this reason, authentic examples of use, taken from native speaker corpora, are preferable to invented sentences. Such example sentences must, however, complement the definition. In the third *Urban Dictionary* entry quoted in the introduction to this paper, the example sentence (*Ashley is obsessed with the breakfast club i consider her as a brekkie*) adds nothing helpful to the definition. The first example (*Couldn't be arsed to eat brekkie this morning*) tells the reader that *brekkie* is something edible and that it is eaten in the morning, which is helpful. The second example sentence, however (*I had eggs and tomato on toast for brekkie today*), shows the reader that *have* collocates with *breakfast*, and that *eggs*, *tomato* and *toast* are examples of breakfast food. It is therefore much more informative, although it still does not tell the reader at what time of day the meal is eaten. Example sentences thus need to be both authentic and informative so that they illustrate the headword and show its use in a real life context. In our dictionary, example sentences were taken wherever possible from the Australian version of the VOLE corpus in Sketch Engine. Where words did not appear in the VOLE corpus, an Australian Internet search was conducted to find suitable examples. All example sentences were chosen because they not only showed the word in use in a sentence but also added a dimension of understanding to the term. As in *LDOCE* (2003), words used in the example sentences were not restricted to the defining vocabulary, as to do so would have placed unnatural limits on the examples and excluded the use of most of the authentic corpus illustrations.

— *Audiovisual material*

Dictionaries should be user friendly, with data that match what users need and that are presented 'in a convivial, pedagogical and easy-to-use way' suitable for a particular user situation (Heid 2011: 289, 290). This may entail factors such as the appearance of the dictionary interface, the layout of the contents, and the principles upon which the dictionary is based. Online dictionaries make it easier to include images, audio clips, and video clips, and to hyperlink to other entries within the dictionary. These features are not without potential problems, however. Lew (2011: 246) indicates that online dictionaries of English have not yet used video imagery to a great extent and that, in any case, animated images do not aid vocabulary retention. Other researchers (for example, Chun and Plass 1996) have indicated that static images are more effective than videos in helping users to remember vocabulary. However, it should be noted that while retention is a desirable outcome of dictionary use, initial comprehension of a term is equally important, and both static and video images may have a role to play here (cf. Atkins and Rundell 2008: 210-211; Svensén 2009: 298; Ogilvie 2011: 393). Audio files can add to the richness of an online dictionary, but so far all the main learner's dictionaries contain only British or US English pronunciation. This fails to cater for the decoding and encoding needs of language learners in other English speaking countries.

— *Language variety*

Among the countries in which a language is spoken as either a first or official language, it is inevitable that variations of that language will occur. For example, Australia has its own variety of English, originating from the children of the first British settlers who arrived in Australia early in the nineteenth century (Delbridge 1983: 36). However, Australian English only came to be recognized in its own right a few decades ago (Peters 2001). Apart from its own pronunciation, Australian English is often marked by colloquial terms and informality of style (Peters 2007: 251) with a tendency to shorten words, so that *biscuit* becomes *bikkie* and *breakfast* becomes *brekkie*. Informality is seen in words such as *Kiwi*, used to refer to inhabitants of New Zealand, and *Pom/Pommy* used (not always flatteringly) to refer to a British person.

— *Cultural context*

In addition to style and pronunciation, each variety of a language reflects the culture in which it is used. The word 'culture' refers to the values, beliefs and customs held or practised by a social group over the generations (Bolaffi et al. 2003: 61). Culture may be made explicit through language, and, in fact, Zgusta (2006) emphasizes that every word of a language is 'embedded in culture' (p. 114). A cultural dictionary may include terms that are not unique to a particular culture, but may have special significance within it or be used frequently by people in that country (Béjoint 2011).

Because a cultural dictionary could include references to events, such as Australian Rules football matches, and other items, such as road signs, the phrase 'culturally bound term' is useful. A culturally bound term is 'a cultural entity that is unique to a particular language and culture in a country, or has a unique meaning in that country among a certain cultural group' (Kwary and Miller 2013).

A dictionary of Australian culturally bound terms could benefit international students, migrants, tourists, and business people visiting the country, as well as anyone who is simply curious to see a list of terms commonly associated with Australia. Because Australia is a vast country, with six states and two territories, culturally bound terms may differ from place to place. Nevertheless, many terms are commonly used throughout the country.

Study

Aim

This action research study focuses on the needs of international students in Adelaide, the capital city of the state of South Australia. It details the process by which an online Australian cultural dictionary was created and provides guidelines for those wishing to construct a similar resource.

Ethics approval for the study was obtained from the University of Adelaide's ethics committee, and all students involved in the study received information and complaints procedure sheets, and signed a consent form. There were nine stages in the dictionary's development.

Stage 1

In order to establish which words needed to be included in an Australian cultural dictionary (ACD) for EAL learners, a paper questionnaire was administered to 269 international students on a 20 week pre-enrolment English program (PEP) which prepared them for entry to a South Australian University. (Please see Appendix 1 for details of the questionnaire.) The questionnaire requested details of each student's age, gender, home country, and first language. Students were then asked what information they would like to find in a dictionary (e.g. definition, part of speech, pronunciation). Finally, they were asked to list three Australian words or expressions that they found hard to understand, and requested to say where they had found out the meaning of that word, whether they had used a dictionary, and, if so, how helpful that dictionary had been. The three words or expressions most frequently elicited from all the students were then incorporated in a second written questionnaire.

Stage 2

The second questionnaire (see Appendix 2) presented a new group of international students on the PEP course ($n = 337$; similar demographics to the first group) with three terms (*koala*, *brekkie*, and *Royal Adelaide Show*) highlighted as problematic by students responding to the first questionnaire. These terms were presented in three ways:

Version A: existing online dictionary entries for *koala* (OALD online), *brekkie* (Macquarie Dictionary), and *Royal Adelaide Show* (based on the entry for *show* in CALD online)

Version B: our own definition, together with IPA pronunciation, pictures, parts of speech, and example sentences

Version C: our own definition, together with IPA pronunciation, pictures, parts of speech, and example sentences, amplified by historical and more encyclopaedic information, together with a word history

The majority of the students (75%) preferred version B. In other words, they wanted a definition that was more complete than that found in a normal learner's dictionary, but they did not want to be swamped by encyclopaedic information or etymology. This is in line with Szczepaniak and Lew (2011)'s suggestion that users may pay little attention to etymological information in a dictionary entry.

Stage 3

A third questionnaire (see Appendix 3) was later presented online to a third cohort of students from the same course (n = 97), in order to cross check the suggestions in questionnaire 1 and elicit more terms necessary for an Australian cultural dictionary.

Stage 4

The terms suggested by students in questionnaires 1 and 3 were conflated, and any words or expressions mentioned at least twice were included in a database, leading to the following entries: *Aboriginal; the Adelaide Festival; the Adelaide Fringe; the Royal Adelaide Show; Anzac day; Anzac biscuit; arvo; Aussie; Aussie Rules; barbie; barrack for; bbq; beaut; bickie/bikkie; big/small bickies/bikkies; bokie; billy; bloke; bludge; bludger; bogan; Bottoms up!; brekkie; budgie smugglers; bush tucker; byo; Centrelink; chip; chook; Christmas Pageant; cricket; Dagwood dog; deli; dingo; (duckbill) platypus; dob in; dummy; dummy run; dunny; echidna; eftpos/EFTPOS; emu; esky; fair enough; footy; g'day; goanna; Good on ya!; heaps; Hills hoist; Hockey-roo; hoon; hotel; How are you going?; How's it going?; It's your call; kangaroo; kiwi; Kiwi; kiwifruit; koala; lamington; lolly; marsupial; mate; Milo; No worries; O-Bahn; outback; pie floater; pokies; Pom; Pommy; (I) reckon; rort; sanga; See ya; She'll be right; showbag; Socceroo; Strine; spill; spruik; stoush; stubby holder; sunnies; sweet as; swag (=collection); swag (=bedding roll); Tasmanian devil; tax file number; tea (= evening meal); thongs; tracky dacks; tragic; Ugg boots; Uluru; ute; Vegemite; wallaby; Wallaby; What are you after?; What's up?; wombat; yakka; Zombie Walk.*

Some terms not unique to Australia (*alpaca; the Ashes; Brussels sprout; cheers; cuppa; Long time no see; spread oneself too thin; Sugar!; ta; tea; tram; vertical garden; yonks; and yummy*) were salient to the participants and so were also included in the dictionary. Other terms (e.g. *chip, hotel, lolly*) may appear in British dictionaries, but have a different meaning in Australia. Only 28 of the terms appeared in all the 'Big 6' English learner's dictionaries, while 31 did not appear in any of them.

Stage 5

An online Australian cultural dictionary for English language learners was then developed, based on the terms in the database. For each word, the part of speech, IPA pronunciation, and grammatical information (e.g. countable or uncountable noun) were provided. A simple definition was then written. The resulting definitions were circulated in an online survey to 20 adult native speakers of English in Australia who had lived for most of their lives in that country, with the request that they check the definitions for adequacy and accuracy. Changes were then made to the definitions based on the feedback of these speakers.

Stage 6

All the words used in the definitions, including information on part of speech and usage, were made into a corpus and analysed for frequency using Adelaide Text Analysis Tool (AdTAT) concordancing software (The University of Adelaide 2013). Numerals were removed from the resultant word list, as were words which were explained in the definition (e.g. *backyard* (*garden*)) or had their own definition (e.g. *marsupial* is used in various definitions, but has its own definition). Equivalents in other languages (such as *Salut* for *Cheers*) were also removed, and so were proper nouns relating to inventors (e.g. the Hills hoist: *named after its inventor, Lance Hill*). Another list was made in which only root words (such as *quick* or *sandwich*) were included in the defining vocabulary, and related parts of speech and plural forms (such as *quickly*, *sandwiches*) were removed. This resulted in a defining vocabulary of 845 words, 146 of which do not appear in any form in the *Academic Word List* (Coxhead 1998) or the *New General Service List* (Browne et al. 2013). Some of these, such as *Australia*, were not thought to present any difficulty to learners. However, there were still words in the defining vocabulary which occurred only once and which might not be entirely simple for learners. For example, *desiccated*, though it is the correct name of a type of coconut product, was thought to be too complicated, so the explanation *dried* was added in brackets afterwards. After all explanations had been added and simplifications made where possible, 33 words remained which might have been problematic for learners.

Stage 7

A fourth questionnaire was formed around these 33 defining vocabulary items to check student understanding of the terms. This online questionnaire was completed by 35 students on the PEP course (25 male, 10 female, average age 22, with 66% from China). This was a convenience sample, as all the students on the course were invited to complete the survey, but only 35 responded. Most of the students (89%) had been in Australia for one to four months, and their global IELTS scores ranged from 5 to 7, with many (71%) in the 5.5 to 6.5 brackets.

This defining vocabulary questionnaire gave the participants four possible meanings for a word and a 'don't know' option. For example, to test the understanding of the phrase *arcade game* the first question asked 'Which one of these is the best example of an arcade game?', with the answer options of 'darts', 'baseball', 'a poker machine', 'snooker' and 'I don't know'. Owing to the nature of the online survey program (SurveyMonkey) at that time, it was not possible to include a picture for each possible meaning. Twenty-four words were understood by 80% or more of participants. The remaining words were *arcade game*, *fairground*, *frankfurter*, *lizard*, *malted*, *oats*, *overhead*, *eucalyptus* and *spikes*. These words were either removed from the defining vocabulary or explained further. For instance, *lizard* is essential to the definition of *goanna*, but the defi-

dition was extended to 'Any type of Australian monitor lizard (a reptile with legs) from 20 centimetres to 2 metres long'. The word *oats* was retained in the definition of *ANZAC biscuit*, as oats are a central ingredient, and the more generic word *grain* might also not be recognized by students. *Eucalyptus* (with its more common alternative *gum*) was retained in the definition of *koala* because koalas are most commonly found in eucalyptus trees and eat eucalyptus leaves. The phrase *long soft spikes* was retained in the definition of *echidna* because the more suitable alternatives — *quills* or *spines* — were thought to be even less recognizable by students. Even the more established learner's dictionaries have trouble with defining *quills*: 'any of the long sharp pointed hairs on the body of a porcupine' (*CALD*); 'one of the long pointed things that grow on the back of a porcupine' (*LDOCE*); 'a long thin sharp object like a stick that grows from the body of porcupines and some other animals' (*MEDAL*); and 'one of the long sharp stiff spines on a porcupine' (*OALD*). We rejected *spines* because of possible confusion with *vertebrae*.

Stage 8

After all words and multimedia files had been collected, a new dictionary database was created using TLex software. For the sake of data standardization, the database was saved in TLex format (tldict file), meaning that all features of the software could be used easily. Next, the collected words were put into the database as new entries. Details were given for each word, including lemma sign, pronunciation/phonetic symbols, part of speech label, definition, and examples. Multimedia files (pictures, sounds, and videos) were assigned to the entries. The database was then exported into an HTML file. One HTML file was generated from each entry/word; all labels for the entries were included.

As TLex exports HTML files in plain text, an interface template for the dictionary website had to be designed (see Figure 1). In making the interface template, we considered consistency and predictability (Lynch and Horton 2008) in terms of the layout of the website and its paths or navigational links. Different pages on the site used the same template, and this was designed to be interesting for users and easily accessible. Links or elements were created and put on all HTML pages, which were designed to be accessible through devices such as desktops, tablets, and mobile phones. Therefore, the dimension of the interface template was set to be dynamic, having the ability to adapt to the height and width of the device screen.

In order to make a website easily accessible, HTML files and their components such as photos and audios must be small in size. Therefore, all multimedia files were compressed and resized. A domain name/hosting, www.culturaldictionary.org, was created, and all HTML files and the multimedia components were uploaded. HTML5 audio player was set as the player for the audio files because it is small, can be opened using most browsers and is routinely used for playing audio files on the Internet. However, we found that

only Firefox, Chrome, Safari, Opera, and Internet Explorer 9 and above supported the HTML5 audio standard, and the audio files did not work on Internet Explorer 8 or earlier. We therefore added this caveat to the website. Rather than hosting video files on the website, the videos were uploaded to YouTube with links embedded on the HTML pages of the dictionary website. To track the traffic data of the website, Google analytics code was put on the home page.



Figure 1: A section of the main page of the *Australian Cultural Dictionary* website

Features of the *Australian Cultural Dictionary*

The ACD is a freely accessible online dictionary with 119 entries in its initial phase. Figure 2 shows an example of the dictionary entry for the headword *kangaroo*.

As can be seen in Figure 2, an entry consists of a headword, pronunciation, grammatical information, usage note, definition, example, and a picture. Some entries, such as the entry for *kangaroo*, also contain a video and hyperlinks to other entries (in this case, to the word *marsupial*). The headword and pronunciation are placed in the first line. Each headword has its own entry. The headwords with spelling variations, i.e. *bickies/bikkies* and *big/small bickies/bikkies*, as well as *EFTPOS/eftpos*, are presented in the same entry. The pronunciation is given in IPA characters and audio files that use either a male or female voice with an Australian accent. Therefore, users who cannot read IPA characters can listen to the pronunciation from the audio file.



Figure 2: The entry for the headword *kangaroo* on the *Australian Cultural Dictionary* website

The second line consists of the grammatical information and usage note. As shown in Figure 2, the grammatical information is not abbreviated (for example, C or U/NC), but spelled out (i.e. 'countable noun') to make it easier for users to understand. The usage notes are included so that dictionary users are aware of the register of each term, and of occasions for its use. Common terms like *the Ashes* (relating to cricket) are marked 'This is the usual term used by all speakers'. More restricted forms are marked 'This term is used by many speakers' (e.g. *barrack for*); 'This term is used by some speakers' (e.g. *arvo*); 'This term is used by some speakers, often in the older age groups' (e.g. *sanga*); or 'This term is used by some speakers, often from the younger age groups' (e.g. *What's up?*). Common words such as *Aussie*, which are not official terms but are frequently used, are marked 'This term is used by speakers of all ages'. Special circumstances are noted, such as 'This term is used by many people when writing an invitation' (e.g. *bbq, byo*).

The definitions present an initial phrase followed by an explanatory sentence where necessary. As explained previously, the definitions are written using a limited defining vocabulary to make it easier for the users to comprehend them. Hyperlinks to other entries in the ACD are also provided. After the definition, there is an example sentence. Example sentences were taken wherever possible from the Australian version of the VOLE (Varieties of Learner English) corpus in Sketch Engine. Where words did not appear in the VOLE corpus, an Australian Internet search was conducted to find suitable examples. All example sentences were chosen because they not only show the word in use in a sentence but also add a dimension of understanding to the term. As in *LDOCE* (2003), words used in the example sentences are not restricted to the defining vocabulary, as to do so would have placed unnatural limits on the

examples and excluded the use of most of the authentic corpus illustrations.

All entries include a photograph. Although there is debate in the literature about the use of photographs rather than drawings (Szczepaniak and Lew 2011: 330), it was more economical for us to take our own photographs or to buy stock photographs, rather than to commission drawings from an artist. Videos are also provided in those cases where a moving image would add useful information to an entry. For example, a video of a kangaroo jumping provides a vivid demonstration of its distinctive motion which it is hard to capture in words or even in a photograph. The number of videos was also restricted due to time limitations in filming and editing.

Evaluation and discussion

Two consecutive new cohorts of PEP students were informed of the ACD website in 2015, and a notice was put on the dictionary's homepage inviting participants to complete an online evaluation in return for a \$10 book voucher. By the end of November 2015 there were 53 responses in total (55% male; 45% female). Most respondents (89%) were aged between 20 and 39 and all came from non-English speaking countries, with nearly three quarters (74%) from China. All but one of the respondents were living in Australia at the time of undertaking the survey. The survey asked detailed questions about the dictionary website (see Table 1 for questions and responses). In addition, there were two open-ended questions asking about the best features of the dictionary and suggestions for improvement (summarised in Table 2).

Table 1: Respondents' answers (n = 53) to questions about the *Australian Cultural Dictionary*

	Strongly agree	Agree	Partly agree	Neither agree nor disagree	Partly disagree	Disagree	Strongly disagree
1. I prefer online dictionaries to paper dictionaries.	28%	33%	13%	17%	4%	4%	0%
2. I often look in a dictionary to help me know how to spell a word.	15%	38%	31%	9%	7%	0%	0%
3. This dictionary has enough detail in the definitions	11%	69%	16%	2%	2%	0%	0%
4. The grammar information (e.g. countable/uncountable noun) in this dictionary is helpful	28%	35%	26%	9%	2%	0%	0%

5. The definitions in this dictionary are easy to understand	20%	69%	7%	2%	2%	0%	0%
6. The spoken (audio) pronunciation in this dictionary is helpful.	39%	37%	17%	7%	0%	0%	0%
7. The details about how to use the words in this dictionary are useful (e.g. "This term is used by many speakers").	31%	39%	24%	4%	2%	0%	0%
8. The example sentences help me to understand how to use the words in the dictionary.	33%	50%	15%	0%	2%	0%	0%
9. The photographs in this dictionary are helpful	31%	39%	28%	2%	0%	0%	0%
10. The videos (used for some terms in this dictionary) are helpful.	31%	51%	7%	11%	0%	0%	0%
11. I would like to see more videos on the website for this dictionary.	27%	44%	11%	18%	0%	0%	0%
12. I like the fact that this dictionary can help me with Australian (rather than British or American English) terms.	44%	42%	12%	2%	0%	0%	0%
13. This dictionary is more useful than my bilingual dictionary for helping me to understand Australian terms.	25%	53%	16%	4%	2%	0%	0%
14. This dictionary is more useful than an English learners' dictionary for helping me to understand Australian terms.	16%	44%	27%	11%	2%	0%	0%
15. This dictionary stimulates my interest in learning Australian terms.	16%	42%	24%	16%	2%	0%	0%

16. This dictionary will help me to use these terms correctly when I speak or write English in future.	20%	50%	22%	6%	2%	0%	0%
--	-----	-----	-----	----	----	----	----

Table 2: Summary of respondents' (n = 53) qualitative feedback on the *Australian Cultural Dictionary*

17. The best feature(s) of this dictionary was/were:	Audio Easy to understand Easy to use Example sentences Free Pictures
18. This dictionary might be improved in future by:	Larger font More entries More example sentences Offline platform

The percentages below, based on the data in Table 1, mainly refer to the total number of respondents who agreed or strongly agreed with a statement.

With regard to general dictionary use, fewer students than expected (61%) preferred online to paper dictionaries. There were also only 53% who used a dictionary to check spelling (although this figure is much higher than Harvey and Yuill's (1997: 259) finding of 24%).

The restricted DV was obviously effective, as 89% of users found the definitions easy to understand. Most of them (80%) also agreed that the ACD provides enough detail in its definitions. Grammatical information was useful for 64% of respondents, indicating that users do actually use (or claim to use) this information (cf. Bogaards 2001). The audio pronunciation was well received (76%), and 70% found the usage information helpful. Example sentences were also thought to be beneficial (83%). Both photographs and videos were considered helpful, but the videos were appreciated more than the photographs (i.e. 82% for videos and 70% for photographs). In addition, 71% requested more videos in a future edition.

Participants particularly liked the Australian focus (86%), and 78% found the ACD more useful than their usual bilingual dictionary for understanding Australian terms. However, only 60% found it more useful than their English learner's dictionary. Nevertheless, only 28 of our terms appeared in all the Big 6, and 31 terms were not in any of the Big 6.

More than half the respondents (56%) said the dictionary stimulated their interest in learning Australian terms. This shows that the ACD could be used as a tool to introduce EAL learners to Australian terms and to encourage better understanding of Australian culture.

Answers to the open ended questions suggested that participants appreciated the fact that the dictionary was free, easy to use and easy to understand. They also particularly liked the audio files and pictures. For the further development of the ACD, they wanted more entries and more example sentences, and would have liked offline availability and a larger font. At the moment, the default font type and size of the definition in the ACD is Verdana 10.5, which is actually similar to the font size of the other online learner's dictionaries. In addition, users can use the Ctrl + buttons on their keyboard if they want a larger font size.

Conclusion

The ACD was an experiment in creating a specialized dictionary to meet a perceived need. Its reception indicates that the need was largely met, particularly in relation to the dictionary's Australian focus, its use of videos and spoken pronunciation, its example sentences and the clarity of its definitions. Useful information was also gained about certain dictionary use habits, the main findings here being that only 61% of participants preferred online to paper dictionaries and 53% of participants used dictionaries to help with spelling.

The data indicate that students do indeed appreciate cultural information in a dictionary, and that this need may not be met in current learner's dictionaries or bilingual dictionaries, particularly if the target culture is not that of the UK or the US. It is therefore suggested that the current ACD be expanded by using words found in the indices of books on Australian culture or by collecting additional data using the method explained in this article. We also suggest that similar dictionaries be created for other varieties of English and for other languages.

References

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Béjoint, H. 2011. Review of Patrick J. Cummings and Hans-Georg Wolf: *A Dictionary of Hong Kong English. Words from the Fragrant Harbour*. *International Journal of Lexicography* 24(4): 476-480.
- Bogaards, P. 2001. The Use of Grammatical Information in Learners' Dictionaries. *International Journal of Lexicography* 14(2): 97-121.
- Bolaffi, G., R. Bracalenti, P. Braham and S. Gindro (Eds.). 2003. *Dictionary of Race, Ethnicity and Culture*. London: Sage.
- Browne, C., B. Culligan and J. Phillips. 2013. *A New General Service List (1.01)*. Available: <http://www.newgeneralservicelist.org/> [17 December 2014].
- Butler, S. (Ed.). 2013. *Macquarie Dictionary*. Sixth edition. Sydney: Pan Macmillan Australia.
- Chun, D.M. and J.L. Plass. 1996. Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal* 80(2): 183-198.

- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press/Clarendon Press.
- Coxhead, A.J.** 1998. *An Academic Word List* (English Language Institute Occasional Publication 18). Wellington, New Zealand: Victoria University of Wellington.
- Delbridge, A.** 1983. On National Variants of the English Dictionary. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 23-40. London: Academic Press.
- Dziemianko, A.** 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Lexicographica Series Maior 130. Tübingen: Max Niemeyer.
- Fraser, H.** 1997. Dictionary Pronunciation Guides for English. *International Journal of Lexicography* 10(3): 181-208.
- Gouws, R.** 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*: 17-29. London/New York: Continuum.
- Häcker, M.** 2012. The Change from Respelling to IPA in English Dictionaries: Increase in Accuracy or Increase in Prescriptivism? *Language and History* 55(1): 47-62.
- Harvey, K. and D. Yuill.** 1997. A Study of the Use of a Monolingual Pedagogical Dictionary by Learners of English Engaged in Writing. *Applied Linguistics* 18(3): 253-278.
- Heid, U.** 2011. Electronic Dictionaries as Tools: Towards an Assessment of Usability. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*: 287-304. London/New York: Continuum.
- Kipfer, B.A.** 1984. *Workbook on Lexicography: A Course for Dictionary Users with a Glossary of English Lexicographical Terms*. Exeter Linguistic Studies 8. Exeter: University of Exeter.
- Kwary, D. and J. Miller.** 2013. A Model for an Online Australian English Cultural Dictionary Database. *Terminology* 19(2): 258-276.
- Lew, R.** 2011. Online Dictionaries of English. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*: 230-250. London/New York: Continuum.
- Lew, R.** 2015. Research into the Use of Online Dictionaries. *International Journal of Lexicography* 28(2): 232-253.
- Lew, R. and G.-M. de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
- Lew, R. and A. Dziemianko.** 2006. 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.
- Lynch, P.J. and S. Horton.** 2008. *Web Style Guide: Basic Design Principles for Creating Web Sites*. New Haven/London: Yale University Press.
- Miller, J.** 2011. *Never Judge a Wolf by its Cover: An Investigation into the Relevance of Phrasemes included in Advanced Learners' Dictionaries for Learners of English as an Additional Language in Australia*. Ph.D. Thesis. Adelaide, Australia: Flinders University.
- Moore, B. (Ed.).** 2016. *The Australian National Dictionary*. Second edition. Melbourne: Oxford University Press.
- Ogilvie, S.** 2011. Linguistics, Lexicography, and the Revitalization of Endangered Languages. *International Journal of Lexicography* 24(4): 389-404.
- Peters, P.** 2001. Corpus Evidence on Australian Style and Usage. Blair, D. and P. Collins (Eds.). 2001. *English in Australia*: 163-178. Amsterdam: John Benjamins.

- Peters, P.** 2007. Similes and Other Evaluative Idioms in Australian English. Skandera, P. (Ed.). 2007. *Phraseology and Culture in English*: 235-255. Berlin: Mouton de Gruyter.
- Procter, P. et al. (Eds.).** 1978. *Longman Dictionary of Contemporary English*. London/Harlow: Longman.
- Rundell, M. (Ed.).** 2002. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan Education. (MEDAL)
- Sinclair, J.M., P. Hanks, G. Fox, R. Moon and P. Stock (Eds.).** 1987. *Collins Cobuild English Language Dictionary*. London/Glasgow: William Collins Sons & Co.
- Sinclair, J. (Ed.).** 2009. *Collins Cobuild Advanced Dictionary of English*. Boston, MA: Heinle Cengage Learning/Harper Collins Publishers. (COBUILD)
- Summers, D. (Ed.).** 2003. *Longman Dictionary of Contemporary English*. Fourth edition. Harlow, Essex: Pearson Education. (LDOCE)
- Svensén, B.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Szczepaniak, R. and R. Lew.** 2011. The Role of Imagery in Dictionaries of Idioms. *Applied Linguistics* 32(3): 323-347.
- The University of Adelaide.** 2013. *AdTAT: The Adelaide Text Analysis Tool*. Available: <http://www.adelaide.edu.au/red/adtat> [17 December 2014].
- Urban Dictionary*. Available: <http://www.urbandictionary.com/> [3 December 2015].
- Walter, E. (Ed.).** 2008. *Cambridge Advanced Learner's Dictionary*. Third edition. Cambridge: Cambridge University Press. (CALD)
- Wehmeier, S. (Ed.).** 2010. *Oxford Advanced Learner's Dictionary*. Eighth edition. Oxford: Oxford University Press. (OALD)
- West, M. (Ed.).** 1953. *A General Service List of English Words: With Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology*. London/New York: Longmans/Green.
- Xu, H.** 2012. A Critique of the Controlled Defining Vocabulary in *Longman Dictionary of Contemporary English*. *Lexikos* 22: 367-381.
- Zgusta, L.** 2006. *Lexicography Then and Now: Selected Essays*. Edited by Dolezal, F.S.F. and T.B.I. Creamer. Lexicographica. Series Maior 129. Tübingen: Max Niemeyer.

Appendix 1: Cultural dictionary questionnaire 1

Please fill out the following information and answer the questions. Your answers will remain entirely confidential.

Gender Male Female **Age**

Home country

First language

We would like to know what Australian words (e.g. *kangaroo*), expressions (e.g. *no worries*) or events (e.g. *an Aussie Rules football game*) you find hard to understand.

When you look for an Australian word, expression or event, what information would you like to find? (You can tick as many answers as you like.)

- A definition
- The pronunciation
- The part of speech (Noun, Verb, etc.)
- Example sentences
- A picture
- An audio file
- A video file
- The history of the word, expression or event
- The register of the word or expression (formal, taboo, etc.)
- Other (please specify)

Please give information about 3 Australian words, expressions or events below.

Question 1

1a. Australian word, expression or event

1b. How did you find the meaning of this Australian word, expression or event?

- From a friend
- From a book about Australian culture

- From the Internet (not an online dictionary)
- From a dictionary (paper or online)
- Other source (please specify)

1c. If you used a dictionary, what was the title of the dictionary? (Please write what you can remember. e.g. dictionary.com, Oxford, etc.)

1d. If you used a dictionary, how helpful was the information provided in the dictionary?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| Very much | A bit | Undecided | Not at all |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

1e. If you did not use a dictionary, please give your reasons.

Question 2

2a. Australian word, expression or event

2b. How did you find the meaning of this Australian word, expression or event?

- From a friend
- From a book about Australian culture
- From the Internet (not an online dictionary)
- From a dictionary (paper or online)
- Other source (please specify)

2c. If you used a dictionary, what was the title of the dictionary? (Please write what you can remember. e.g. dictionary.com, Oxford, etc.)

2d. If you used a dictionary, how helpful was the information provided in the dictionary?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| Very much | A bit | Undecided | Not at all |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

2e. If you did not use a dictionary, please give your reasons.

Question 3

3a. Australian word, expression or event

3b. How did you find the meaning of this Australian word, expression or event?

- From a friend
- From a book about Australian culture
- From the Internet (not an online dictionary)
- From a dictionary (paper or online)
- Other source (please specify)

3c. If you used a dictionary, what was the title of the dictionary? (Please write what you can remember. e.g. dictionary.com, Oxford, etc.)

3d. If you used a dictionary, how helpful was the information provided in the dictionary?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| Very much | A bit | Undecided | Not at all |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

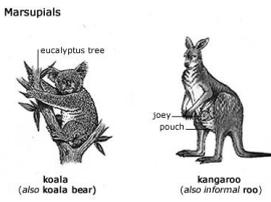
3e. If you did not use a dictionary, please give your reasons.

Appendix 2: Cultural dictionary questionnaire 2

The following are dictionary entries for 3 different words or expressions. For each one, please indicate which version you prefer and state your reason in the box at the end of each page. On the last page there is a question about your first language. Please complete all four pages of the questionnaire.

Your answers will remain entirely confidential.

Question 1

<p style="text-align: center;">Version A koala <i>BrE</i> /kəʊˈɑ:lə/ <i>NAme</i> /kouˈɑ:lə/ (also koala 'bear') noun an Australian animal with thick grey fur, large ears and no tail. Koalas live in trees and eat leaves.</p> <p>Marsupials</p>  <p>koala (also koala bear)</p> <p>kangaroo (also informal roo)</p>	<p style="text-align: center;">Version B koala /kəʊˈɑ:lə/ noun (countable) A koala is an Australian animal about 75cm long with thick grey fur, which lives in trees and feeds on gum (eucalyptus) leaves. They are sometimes called 'koala bears', but they are not bears.</p> <p>e.g. <i>We went to a wildlife park and saw a koala sleeping in a tree.</i></p>  <p style="text-align: center;"><u>Link to a Video File about Koalas</u></p>	<p style="text-align: center;">Version C koala /kəʊˈɑ:lə/ noun (countable) A koala is an Australian animal about 75cm long with thick grey fur, which lives in trees and feeds on gum (eucalyptus) leaves. Koalas sleep up to 18 hours a day. They are sometimes called 'koala bears', but they are not bears.</p> <p>e.g. <i>We went to a wildlife park and saw a koala sleeping in a tree.</i> Koalas are marsupials, so they carry their young in a pouch. The word 'koala' is based on an Australian Aboriginal word.</p>  <p style="text-align: center;"><u>Link to a Video File about Koalas</u></p>
<p>The version I prefer is (write A, B, or C):</p> <p>The reason:</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>		

Question 2

<p>Version A brekkie /'breki/ noun (Colloquial) breakfast</p>	<p>Version B brekkie /'breki/ Countable noun (colloquial) breakfast The first meal of the day. Typical Australian breakfast dishes are cereal with milk; toast; and coffee, tea or fruit juice. e.g. "What did you have for brekkie?" "Oh, not much — just toast and Vegemite."</p>  <p><u>Link to a Video File about brekkie</u></p>	<p>Version C brekkie /'breki/ Countable noun (colloquial) breakfast The first meal of the day. Typical Australian breakfast dishes are cereal with milk; toast; and coffee, tea or fruit juice. e.g. "What did you have for brekkie?" "Oh, not much — just toast and Vegemite." 'Brekkie' is short for 'breakfast'. Many Australian words are shortened in this way. They are informal, but not rude. You would use them when talking to friends.</p>  <p><u>Link to a Video File about brekkie</u></p>
<p>The version I prefer is (write A, B, or C):</p> <p>The reason:</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>		

Question 3

<p>Version A Royal Adelaide Show /'rɔɪəl 'æd.leɪd 'ʃəʊ/ Noun [C] an agricultural event at which a group of related things are available for the public to look at <i>There were some amazing animals and cakes at the Show.</i></p>	<p>Version B Royal Adelaide Show /'rɔɪəl 'æd.leɪd 'ʃəʊ/ Countable noun a week-long annual agricultural show, at which you can see animals, birds, plants, cakes and other produce. There are also fair-ground rides, and children in particular enjoy buying 'showbags', which contain many different items ranging from lollies (sweets) to football paraphernalia.</p> <p>e.g. <i>We went to the Show and saw some beautiful cows and horses. I really liked the decorated cakes. The kids loved the rides, and of course they liked the showbags too!</i></p> <p> <u>Link to a Video File about the Show</u></p>	<p>Version C Royal Adelaide Show /'rɔɪəl 'æd.leɪd 'ʃəʊ/ Countable noun a week-long annual agricultural show, at which you can see animals, birds, plants, cakes and other produce. There are also fair-ground rides, and children in particular enjoy buying 'showbags', which contain many different items ranging from lollies (sweets) to football paraphernalia.</p> <p>e.g. <i>We went to the Show and saw some beautiful cows and horses. I really liked the decorated cakes. The kids loved the rides, and of course they liked the showbags too!</i></p> <p>The Royal Adelaide Show dates back to the Adelaide Produce Show in 1840. It is now held in a special location at Wayville. Showbags were originally bags with free samples of produce from the show. The word 'showbag' is sometimes used ironically in other contexts where you are offered a bag with free information or samples.</p> <p> <u>Link to a Video File about the Show</u></p>
<p>The version I prefer is (write A, B, or C):</p> <p>The reason:</p> <p>.....</p>		

Question 4

Do you find the following things helpful when you use a dictionary?

- | | | | |
|---|-------------|---------------|--------------|
| 1. Definition | Not helpful | A bit helpful | Very helpful |
| 2. Part of speech (noun, adjective, etc.) | Not helpful | A bit helpful | Very helpful |
| 3. Information on whether a noun is countable | Not helpful | A bit helpful | Very helpful |
| 4. Word history (where the word comes from) | Not helpful | A bit helpful | Very helpful |
| 5. Picture | Not helpful | A bit helpful | Very helpful |
| 6. Pronunciation | Not helpful | A bit helpful | Very helpful |
| 7. Example sentence | Not helpful | A bit helpful | Very helpful |
| 8. Register (when to use the word) | Not helpful | A bit helpful | Very helpful |
| 9. Would you like to see a video of someone using the words? | No | Undecided | Yes |
| 10. Would you like to hear the words spoken in an audio file? | No | Undecided | Yes |

Please fill out the following information. Your answers will remain entirely confidential.

Gender Male Female **Age**

Home country

First language

English dictionary I usually use.....

Thank you very much for completing the questionnaire.

Appendix 3: Cultural dictionary questionnaire 3

Please fill out the following information and answer the questions. Your answers will remain entirely confidential.

Gender Male Female Age

Home country

First language

How often do you use a dictionary?

Never Once a week Several times a week Once a day More than once a day

If you use a dictionary, what kind do you prefer? Please tick only one box.

- Mostly paper dictionary
- Mostly online dictionary (via computer, tablet or iPad)
- Mostly online dictionary (via mobile phone)
- Mostly electronic dictionary (installed in a computer or tablet)
- Mostly mobile dictionary (installed in a mobile phone)
- Mostly pocket electronic dictionary

Why do you prefer this kind of dictionary?

.....

We would also like to know what Australian words (e.g. *kangaroo*), expressions (e.g. *no worries*) or events (e.g. *an Aussie Rules football game*) you find hard to understand.

Please list up to 5 Australian words, expressions or events that are/were difficult to understand.

1. Australian word, expression or event

.....

2. Australian word, expression or event

.....

3. Australian word, expression or event

.....

4. Australian word, expression or event

.....

5. Australian word, expression or event

.....

Appendix 4: Cultural dictionary questionnaire 4

Thank you for taking part in this survey.

We are trying to develop a list of words to define other words in a dictionary. We therefore want to see which words people know and which words are too hard to use in a dictionary definition.

Please match each word with the best meaning from the choices below. We are not testing your knowledge; we are testing whether our definitions will work. Therefore, if you don't know an answer, please tick 'don't know' instead of guessing.

1. Which one of these is the best example of **an arcade game**?
 - a. darts
 - b. baseball
 - c. a poker machine
 - d. snooker
 - e. I don't know

2. Which one of these is the best example of **artwork**?
 - a. a painting
 - b. a cake
 - c. a tractor
 - d. a tablecloth
 - e. I don't know

3. What is a **bikini**?
 - a. a short dress
 - b. a two piece swimsuit
 - c. a blouse
 - d. a type of biscuit
 - e. I don't know

4. What is a **cabbage**?
 - a. a gun
 - b. a vegetable
 - c. a taxi
 - d. a drink
 - e. I don't know

5. What is **charcoal**?
 - a. something you put in a gun
 - b. something you use to wash clothes

- c. something you burn to cook food
 - d. something you mix with fruit juice
 - e. I don't know
6. What is a **circus**?
- a. a place full of machines
 - b. a place with clowns, animals and acrobats
 - c. a place like a factory
 - d. a place where they make ice cream
 - e. I don't know
7. What is a **coconut**?
- a. a hard brown nut with a white inside
 - b. a vegetable with a purple skin
 - c. a bird with long feathers
 - d. a type of dog
 - e. I don't know
8. What is a **cookie**?
- a. Someone who cooks (another word for 'chef')
 - b. an animal
 - c. a type of pasta
 - d. Something sweet that you eat (another word for 'biscuit')
 - e. I don't know
9. Which one of these is usually **crispy**?
- a. a banana
 - b. a piece of cheese
 - c. bubblegum
 - d. a potato chip
 - e. I don't know
10. Which one of these do people usually do at a **fairground**?
- a. go on exciting rides
 - b. practise cooking
 - c. do their homework
 - d. sunbathe
 - e. I don't know
11. What is a **frankfurter**?
- a. a type of vegetable
 - b. a type of food used in a hot dog
 - c. a piece of furniture
 - d. another word for an essay
 - e. I don't know

12. What is **gambling**?
- dancing at a disco until you fall over
 - drinking too much alcohol
 - risking money to try and win more money
 - walking in the country until you are very tired
 - I don't know
13. What is a **grill**?
- a metal frame to cook meat on
 - a kind of potato found in Australia
 - a person who is new to an area
 - a type of insect that lives in cities
 - I don't know
14. What is a **gum**?
- a type of Australian animal
 - a type of Australian tree
 - a type of Australian beer
 - a type of Australian aeroplane
 - I don't know
15. Where would you find **lenses**?
- in a forest
 - in the clouds
 - in a pair of glasses/spectacles
 - in a kitchen
 - I don't know
16. What is a **lizard**?
- a reptile
 - a rock
 - a crack in the ground
 - a kind of elephant
 - I don't know
17. What is a **lollipop**?
- a kind of green vegetable (e.g. a bean)
 - a hard sweet on a stick (e.g. a Chupa Chup)
 - a kind of song you hear at bedtime
 - an Australian bird
 - I don't know
18. Which one of these drinks is **malted**?
- tea
 - coffee

- c. soda water
 - d. Ovaltine/Horlicks
 - e. I don't know
19. What is a **motorcycle**?
- a. a fast car
 - b. a boat with an engine
 - c. a two wheeled vehicle with an engine
 - d. a fast bird like an eagle
 - e. I don't know
20. What are **oats**?
- a. a way of measuring speed
 - b. a type of grain
 - c. a type of beer
 - d. a way of planting vegetables
 - e. I don't know
21. Which one of these is **oval**?
- a. a tree
 - b. a bottle
 - c. an orange
 - d. an egg
 - e. I don't know
22. Where would you see **overhead** wires?
- a. under you
 - b. above you
 - c. inside a table
 - d. on your feet
 - e. I don't know
23. What is a **pea**?
- a. a brown bird
 - b. an orange fruit
 - c. a green vegetable
 - d. a purple fish
 - e. I don't know
24. What does '**p.m.**' mean?
- a. in the afternoon
 - b. in the morning
 - c. at dawn
 - d. just after midnight
 - e. I don't know

25. What is **poker**?
- a drink
 - a kind of mobile phone
 - a book
 - a card game
 - I don't know
26. Which one of these is usually **rectangular**?
- a cat
 - a pearl
 - a book
 - a flower
 - I don't know
27. What do you do with **sandals**?
- Wear them on your head
 - Eat them with a spoon
 - Wear them on your feet
 - Cook them in the oven
 - I don't know
28. Which one of these is often made of **sheepskin**?
- a cake
 - a drink
 - a rug
 - a pencil
 - I don't know
29. What is a **snack**?
- an animal with no legs
 - something quick to eat
 - something to put on the wall
 - a zoo animal
 - I don't know
30. Which one of these has **webbed** feet?
- an elephant
 - a lion
 - a cat
 - a duck
 - I don't know
31. What is a **eucalyptus**?
- a type of tree
 - a type of bird

- c. a type of game
 - d. a type of song
 - e. I don't know
32. What do you usually find at the **beach**?
- a. machines and factories
 - b. trees and sheep
 - c. sea and sand
 - d. pandas and cats
 - e. I don't know
33. Which one of these animals has **spikes**?
- a. a tiger
 - b. a cow
 - c. a rabbit
 - d. a porcupine
 - e. I don't know

Thank you for completing the questionnaire.

Please click 'done' to submit your answers and exit the survey.

The Dictionary in Examinations at a South African University: A Linguistic or a Pedagogic Intervention?*

Dion Nkomo, *School of Languages, Rhodes University,
Grahamstown, South Africa (d.nkomo@ru.ac.za)*

Abstract: This paper interrogates students' use of dictionaries for examination purposes at Rhodes University in South Africa. The practice, which is provided for by the university's language policy, is widely seen as a linguistic intervention particularly aimed at assisting English additional language students, the majority of whom speak African languages, with purely linguistic information. Such a view is misconceived as it ignores the fact that the practice predates the present institutional language policy which was adopted in 2006. Although it was difficult to establish the real motivation prior to the language policy, this study indicates that both English mother-tongue and English additional language students use the dictionary in examinations for assistance that may be considered to be broadly pedagogic rather than purely linguistic. This then invites academics to reconsider the manner in which they teach and assess, cognisant of the pedagogic value of the dictionary which transcends linguistic assistance.

Keywords: DICTIONARY USE, DICTIONARY ASSISTANCE, PEDAGOGICAL LEXICOGRAPHY, OPEN-BOOK EXAMINATIONS, LINGUISTIC INTERVENTION, PEDAGOGIC INTERVENTION

Opsomming: Die woordeboek in eksamens by 'n Suid-Afrikaanse universiteit: 'n taalkundige of 'n opvoedkundige ingryping? Hierdie artikel ondersoek studente se gebruik van woordeboeke vir eksamendoeleindes by die Rhodes Universiteit in Suid-Afrika. Die praktyk, waarvoor daar voorsiening gemaak word in die universiteit se taalbeleid, word wyd beskou as 'n taalkundige ingryping met suiwer taalkundige inligting wat in die besonder daarop toegespits is om Engels Addisionele Taal-studente te help, waarvan die meerderheid Afrikatale praat. So 'n siening is 'n mistasting, aangesien dit die feit ignoreer dat die praktyk van voor die huidige institusionele taalbeleid, wat in 2006 aanvaar is, dateer. Alhoewel dit moeilik was om die werklike motivering daarvoor vóór die huidige taalbeleid vas te stel, toon hierdie navorsing dat sowel Engels Moedertaal- as Engels Addisionele Taal-studente die woordeboek tydens eksa-

* This article was presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1-3 July 2014.

mens gebruik vir hulp wat oor die algemeen beskou kan word as opvoedkundig eerder as suiwer taalkundig. Dit dring akademië om die manier waarop hulle onderrig gee en assessering doen, te heroorweeg, en kennis te neem van die opvoedkundige waarde van die woordeboek, wat taalkundige hulp oortref.

Sleutelwoorde: WOORDEBOEKGEBRUIK, WOORDEBOEKHULP, OPVOEDKUNDIGE LEKSIKOGRAFIE, OOPBOEKEKSAMENS, TAALKUNDIGE INGRYPING, OPVOEDKUNDIGE INGRYPING

1. Introduction

This paper interrogates students' use of dictionaries for examination purposes at Rhodes University in South Africa. The practice is provided for by the language policy of the university while the institution's assessment policy is silent about it. The language policy requires academic departments to "provide access to a wider range of dictionaries in examinations" where appropriate (Rhodes University Language Policy 2006; 2014). Accordingly, the African Language Studies and the German Studies sections of the School of Languages and Literatures recommend the use of certain dictionaries in translation and language learning studies examinations. Such dictionary use during examinations is common in South Africa and abroad (Atkins 1998; Barnes et al. 1999; Bensoussan 1983; Bishop 2000; Nesi and Haill 2002). However, it is the use of monolingual English dictionaries during examinations, almost across the board, that concerns this paper. As far as could be established, this practice is rare although it would not be surprising to see it spreading with the on-going efforts of implementing university language policies in South Africa. Besides implementing multilingualism, the language policies encapsulate interventions that are hoped to help students with challenges posed by the language of teaching and learning. It is important that such interventions are investigated in terms of their efficacy.

The rationale for dictionary use in examinations at Rhodes University is not explicitly articulated but rather inferred only in terms of the language policy as an intervention that seeks to eliminate linguistic barriers for English additional language students, the majority of whom speak African languages. This suggests that the dictionaries, which are discussed later in the article, are consulted for purely linguistic problems during examinations, especially by English additional language students. For this reason, the practice is supported by academics who regard it as integral in facilitating students' success irrespective of linguistic backgrounds. Unfortunately, such claims are not based on concrete knowledge of what happens during examinations in relation to dictionary use. Even more remarkable, the senior citizens of the institution only recall that the practice pre-dates the language policy which came into effect in 2006. Institutional language policies became a legislative imperative following the recommendation of the *Language Policy for Higher Education* in 2002 in order

to outline how universities would, among other issues, address the language of teaching and the role of indigenous African languages as part of transformation. The argument for dictionary use during examinations as a transformative intervention that facilitates epistemic access for English additional language students therefore becomes an unsustainable case of convenience where the language policy merely co-opted a practice whose original motivation might have been different from the current discourses of language in higher education.

This paper is based on data that was collected through three questionnaire surveys, observations and interviews. The first questionnaire (Appendix 1) was administered among first-year students during registration in February 2014 to get new students' views regarding a practice that would characterise their examinations. The second questionnaire (Appendix 2) was completed by returning students to establish the extent to which they have been using the dictionary in their previous examinations. The final questionnaire (Appendix 3) was completed during the June 2014 examinations by students who were observed using the dictionary in particular examinations mainly to establish specific issues that instigated dictionary consultation. Observations were conducted over two weeks during the June 2014 examinations to get a sense of the prevalence and frequency of dictionary use, as well as the information needed and obtained by students in the process. Interviews were conducted with the key institutional stakeholders, including the Deputy Vice-Chancellor for Teaching and Learning, University Registrar, the Dean of Students, the Chairperson of the institutional Language Committee, academics and students in order to solicit perceptions towards the practice. Appropriate ethical procedures were followed, including seeking institutional permission from the Registrar, the Examinations Officer and applying for ethical clearance from the Rhodes University Ethical Standards Committee. Students who participated in the research were also fully informed about it and advised about the voluntary nature of their participation.

Using the collected data, this article challenges the purported language-oriented motivation for dictionary use in examinations and projects a pedagogic argument as an alternative. The data displays various perceptions regarding dictionary use at Rhodes University, both in general and specifically in connection with examinations, and presents an empirical picture of what actually happens with the dictionary in examination venues. For this purpose, the notion of open-book examinations, complemented by insights from pedagogical lexicography, informs the discussion. In the context of rapidly developing educational technologies, including computer-based assessments, issues of open-book examinations and access to electronic dictionaries and other tools installed in such technologies during examinations become relevant. The adoption of take-home examinations as a strategy of saving the academic year following student protests in 2016 by Rhodes University and other South African institutions has probably catalysed the implementation of open-book

examinations, which will have implications for dictionary use during examinations.

2. Pedagogical lexicography

Although lexicography was originally regarded as a branch of linguistics (Zgusta 1971), and dictionaries as largely linguistic instruments (Atkins and Rundel 2008; Béjoint 2010), McArthur (1986) aptly regards them as containers of knowledge, with other modern scholars recognising that such knowledge transcends language (Wiegand 1984; Tarp 2008). The branch of lexicography that provides relevant theoretical lenses for understanding the needs of students leading to dictionary consultation in examinations and the assistance obtained from dictionaries would be pedagogical lexicography, also called learner's lexicography (Cowie 2002; Tarp 2008). It developed in connection with the teaching of English as a foreign language in Asia in the 1970s through groundbreaking lexicographic approaches of Albert Snell-Hornby, Harold Palmer and Michael West, who were experienced English teachers. The term *learner's dictionary* has been extensively used as part of titles of dictionaries produced for English additional language learners. However, its interpretation has been restricted to regard a learner only as somebody who is learning an additional language, as opposed to a native speaker, who may also enrol in a programme to learn his/her native language (Gouws 2010; Tarp and Gouws 2012). Such an interpretation concurs with the generally problematic perception in this study that the dictionary in examinations supports English additional language students. In a sense that resonates with the South African education system, the learner in pedagogical lexicography refers to any person who is learning any language, native or additional, or any academic subject (Gouws 2010; Tarp and Gouws 2012). This means that users who consult dictionaries in relation to their learning, qualify to be learners, as the dictionary offers them assistance in connection with their education. In this paper, principles drawn from pedagogical lexicography are used to characterise the needs of students who consult dictionaries during examinations and the type of information provided by the dictionaries to satisfy those needs.

3. Open-book examinations

The use of dictionaries under examination conditions evokes an old debate on open-book examinations. Open-book examinations entail the use of different pedagogic aids such as graphics calculators and computer algebra systems in mathematics, atlases in geography, set texts in literature or even dictionaries in second language acquisition or translation studies, textbooks and notes in any subject (Stalnaker and Stalnaker 1934; Boniface 1985; Barnes et al. 1999; Bensoussan 1983; Bishop 2000; Graham et al. 2003; Kemp et al. 1996; Monaghan

2000). The consensus seems to be that open-book examinations are not inherently inappropriate. Graham et al. (2003: 320) reckon that such examinations offer better opportunities of aligning assessment with curriculum objectives, teaching practices and even real life or work situations compared to some traditional closed-book examinations for which students have to prepare by cramming concepts that they may never need in their future professions. For example, the use of graphics calculators and computer algebraic systems in mathematics examinations has been considered helpful in as far as candidates have been taught how to use these tools appropriately for specific tasks in the course (Stalnaker and Stalnaker 1934: 117). The use of learning aids becomes part of the skills that are the focus of assessment alongside others in the course. The same could be said of dictionary use in the context of second language learning or translation studies examinations where looking up the meaning of words can be helpful for the candidate who would be able to choose appropriate translation equivalents, comprehend definitions and then apply them appropriately in the relevant context. Dictionary using skills become important for the translator's future practice. In this paper, the notion of open-book examinations is used in relation to the purposes for which dictionaries are used in examinations *vis-à-vis* how the questions have been formulated, and whether such use is consistent with the hypothesis that the dictionary is permitted only as a language intervention tool.

4. Perceptions of academics

Interviews with university personnel revealed perceptions ranging from support to opposition and indifference towards dictionary use in examinations. The contradictory perceptions indicate different attitudes towards the language question in higher education, as it is assumed to be the major issue addressed by the practice, as well as the pedagogic role of the dictionary in general. Some of the perceptions are unfortunately misguided. Academics who support the dictionary in examinations, as already indicated, evoke the institutional language policy and would recommend any intervention meant to eradicate linguistic barriers to access by and success for English additional language students. They particularly advance that mother-tongue English speakers enjoy an advantage over their non-mother tongue peers who have to switch from other languages whenever they move in and out of academic spaces. The dictionary is thus seen as a transformative intervention supporting English additional language students, just like opening spaces for indigenous languages in South African higher education and education broadly to level the field for students with diverse linguistic, cultural and educational backgrounds. While sensitivity towards the language issue is welcome, this argument falls short as it disregards the long history of the practice and takes for granted what students do with the dictionary and what dictionaries offer, especially considering the fact

that the dictionaries used employ English as the language of explicating the academic language with which English language additional students are struggling.

There are basically two arguments advanced by academics against the dictionary in examinations. One position simply opposes the language-policy oriented argument, positing that the language-problem in South African higher education is misconstrued if not overstated (Boughey 2002), given that academic language is nobody's mother-tongue. From this position, language-based problems should not be an excuse for students' academic struggles to the extent of needing the dictionary as an intervention in examinations. This argument is hugely flawed, if not insensitive; it disregards the notion of pedagogy of place, thereby decontextualising academic expectations of university students. It constructs universities as similar across space and time, while delinking the linguistic identities of students from the curriculum. Specifically, the argument fails to recognise that mother-tongue speakers of the language of teaching and learning already have stronger linguistic and cultural foundations for the development of academic language than non-mother tongue speakers. Although there are some irregularities regarding dictionary use in examinations, as the study shows, the practice may not simply be challenged by merely dismissing the language problem or saying 'not at university' without a pedagogic qualification of why not at university and particularly in examinations.

The limitation of the foregoing argument against dictionary use in university examinations leads to the second one, which appears to be a matter of greater consideration. This argument advances that dictionary use for examination purposes may not be embraced or dismissed as a question of right or wrong. Instead, the role of the dictionary must be considered in relation to the broader curriculum aims and in terms of its implications for assessment. Such discourse can be illustrated by the views of a Pharmacy lecturer who considers dictionary use for pharmacy examinations problematic in that the dictionary may explain not only the language but also the concepts that students need to learn and master during the course in order to apply them in future practice. The orientation is that discipline-specific terminology, which is also assessed in various subjects, is an integral part of the pharmaceutical knowledge. Thus, according to this academic, the technical acumen of a student who needs a dictionary in the exam would be suspect, i.e. the student may not have fully achieved the learning outcomes of the course. This view is shared by the majority of lecturers who are conscious of what they teach and how they assess. Another example is from the Law Faculty where the following instruction was given in a *Legal Pluralism* examination paper:

The *Oxford Concise English Dictionary* may be used during this examination for the purpose of looking up words that are not legal terms relevant to the course content. Invigilators may ask you which word you wish to look up to determine whether the use of the dictionary is appropriate.

This statement indicates that dictionary use may be considered appropriate or inappropriate, depending on what students look up and in relation to course content.

However, there are some academics who prohibit the dictionary in their examinations because they consider it unhelpful and a time-wasting exercise. The view is that dictionaries provide limited and unreliable definitions when it comes to discipline-specific language and students who consult a dictionary for this purpose would be misguided. A Journalism and Media Ethics lecturer prohibits dictionary use in his examinations mainly to protect his students from this danger. Unfortunately, this view is based on limited knowledge about dictionaries as knowledge tools, particularly around issues of dictionary typology and the kinds of assistance that users may get from different types of dictionaries. While some dictionaries may have limitations, as indicated by Nielsen (2015), such limitations may not be generalised to the extent of influencing a policy on dictionary use.

Finally, it is significant to note that some academics who are not against the use of dictionaries in examinations at Rhodes University do not clearly or strongly support the practice. The practice is now part of a tradition that new academics merely embrace without giving it sufficient thought or taking it into cognisance when setting examinations. Their attitude towards the practice is at best that of indifference and prevails even among the senior management of the university. Some thought-provoking questions during separate interviews with the Registrar and the Dean of Students saw a sudden change in the practice even before the findings of this research were presented. The Examinations Office, which supplies the dictionaries, suddenly called on academics to include a compulsory instruction regarding the permissibility of dictionary use on their examination papers since June 2014. The sudden change indicates that there was hitherto no well-conceived institutional position regarding dictionary use in examinations. Nevertheless, it has offered academics an opportunity to think about the dictionary and ways of formulating assessment tasks which may have implications for dictionary use in examinations. Whether such an opportunity is optimised is a question for the future.

5. Perceptions of students

The data obtained through Questionnaires 1 and 2 indicate that students generally support dictionary use under examination conditions. 69 of the 109 first-year students who completed the questionnaire thought they would need the dictionary during university examinations, a privilege they have never had at school level. This is most probably because of fear of the unknown as students expect a steep increase in academic expectations from high school to university. One student actually remarked, "I have never written an exam of university level and would not know what to expect" (sic). Such students seem to expect language use at university to increase in complexity together with the level of depth in academic engagement. Interestingly, 7 of the 69 first-year students

who indicated that they would take advantage of dictionary use in examinations consider the practice as cheating which may "give students answers". Such sentiments also emerge from Questionnaire 2 data. 48 out of 95 students considered dictionary use as an appropriate intervention in examinations.

However, only 26 of the 95 students indicated that they had resorted to this intervention, with 2 of them regarding it as bordering on cheating. The perception that dictionary use in examinations borders on cheating also surfaced in some responses of first-year students who anticipated not to need the dictionary, as well as the returning students who indicated that they have never experienced the need in their previous examinations. Some students made the following remarks:

- Examination is preparation. Dictionaries mean less prep is required.
- People should learn all the necessary language when studying for the exam.
- Students should understand all concepts before the exam.
- Learners should be able to understand questions without it. If they don't, either work was not studied ...
- I used it to find definitions I should have learnt. They were provided in class.

These sentiments suggest that dictionary use in examinations supports or rewards lack of effort on the part of students and generally concur with the arguments of the majority of academics typified by the Pharmacy lecturer and Law Faculty referred to above. However, such sentiments did not feature in Questionnaire 3 data, perhaps given that students completed the questionnaire immediately after using the dictionary in their examinations. The chances of these students being critical of the practice immediately after benefitting from it would be low.

Comprehending questions seems to be the most salient reason why new students anticipated the need for the dictionary in their examinations. This is captured below:

- ... so that I could give explicit answers and be sure that I understood the question.
- So that I may look up words that I do not fully understand and I may answer confidently.
- Using a dictionary would ensure that I am fully able to understand all the questions.
- To correctly answer questions.

Within the function theory of lexicography, which is based on a simple communication model, comprehension is an important aspect of the text reception function of dictionaries in communicative situations (Bergenholtz and Tarp 2003; Tarp 2008). According to this theory, comprehension problems arise when one is not familiar with the language that is used or some of its aspects. A combined total of 19 students who support dictionary use in examinations

indicate that the dictionary would provide assistance in relation to understanding English, the language of teaching at university, or understanding the specialised languages used in the different discipline-specific subjects. A link was made by some students between comprehension challenges and the proficiency of English additional language students, as can be seen below:

- ... language of study is English and not all students necessarily understand it and therefore need a dictionary.
- ... if English is not their home language.
- ... there are a lot of English words that will be new to me as I am not used to using English as my medium of instruction and communication.
- Because I want to make sure of the words that I will use or read. ... English is not my mother tongue.
- Afrikaans was my home language (at school) so I will need a dictionary for some explanations.

Such a link may not be dismissed. English mother-tongue speakers would have a broader vocabulary base compared to their non-mother tongue counterparts. This provides a stronger foundation for building the academic language and has implications for their communicative competence, including comprehension which is fundamental to how well students respond to examination questions and tasks (Rundell 1999: 35). However, although comprehension is largely a communicative process, it may also be affected by cognitive factors that are beyond issues of language. Language *per se* may never be the sole factor in the examination. One also needs to be cognisant of what the language is used in connection with, inasmuch as examination preparation and the pedagogies that precede it may not be limited to linguistic processes. Therefore, a dictionary that addresses only linguistic problems would be appropriate for use in examinations if language was the real and only issue that the intervention sought to deal with. Yet it remains unclear yet whether the problems are only linguistic and also whether the dictionary offers purely linguistic assistance in a manner that does not affect assessment. When asked why they would need a dictionary in their examinations, some first-year students made allusions to pedagogical issues that make the dictionary handy:

- Exams can be so overwhelming that you forget even the minor words used in questions. So having a dictionary will make me feel at ease.
- Exam questions are often vague and sometimes lecturers include words that may not be fully understood by myself or other students ...
- Learners should be able to understand questions without it. If they don't, either work was not studied or lecturer did not set appropriate papers.

From the above responses regarding circumstances that lead to students using the dictionary, issues of anxiety, poor examination preparation by students as well as poor formulation of examination questions and tasks by lecturers come to the fore (see also Stalnaker and Stalnaker 1934: 119). This clearly suggests

that the dictionary serves as more than a linguistic intervention. Whether the intervention offers sufficient assistance and whether such assistance is indeed necessary is another question. One student who considers dictionary use to provide room for cheating indicated that (s)he would use the dictionary for the convenience of having it: "If it is available why not?"

6. Dictionary users, their linguistic backgrounds and subjects of study

81 post-examination questionnaires were completed by 79 candidates who used the dictionary during the June 2014 examinations. Two candidates notably completed the questionnaire twice in connection with separate examinations. The 79 candidates are those who wrote their examinations at the observed venues (exams were written concurrently at several venues and not all of them could be observed) and consented to participate in the research as others had no energy, interest nor time to complete the questionnaire after energy-sapping examinations. This means that dictionary use during examinations at Rhodes University was more prevalent than reported here.

Observational and Questionnaire 3 data indicate that the dictionary was used in connection with various subjects that include Art History, Economics, Entomology, History, Information Systems, Law, Microbiology and Political Science during the June 2014 examinations. The subjects are offered in four of the five faculties whose examinations were observed, with the Faculty of Pharmacy having no dictionary usage instance. The non-use of the dictionary for Pharmacy examinations may be attributed to the fact that students had no choice as their examinations clearly prohibited dictionary use through the compulsory instruction that had just come into effect. On the other hand, the use of the dictionary by Law students is an interesting case where candidates either did not read the instruction regarding dictionary use or deliberately violated it. As may be seen from Appendix 4, legal terms and expressions were looked up while the instruction only permits students to look up non-legal words. This suggests that invigilators did not fulfil the responsibility of checking what students would look up, a difficult task indeed.

When looking at the users of dictionaries during the June 2014 examinations, observational data presented speakers of indigenous African languages as a slight majority, true to the anticipation of the language-policy oriented rationale of the intervention. However, mother-tongue English speakers also used the dictionary. A third-year English monolingual student writing an Entomology examination who requested the dictionary countless times is a good example. This suggests that the problems prompting dictionary consultation may not be exclusively linguistic. To some extent, they may have to do with the preparedness of the candidate for examinations, which includes how students have been taught and prepared for their examinations. For example, while 26 of the 51 candidates who wrote the *Introduction to Science Concepts & Methods Paper 1* examination used the dictionary (some of them repeatedly), they are on

the Extended Studies Programme (ESP), which offers an alternative access route to higher education for students with lower than the required entry level points. Most of them speak isiXhosa, the most dominant language in the Eastern Cape, the province where Rhodes University is located. Generally, ESP students have low levels of preparedness for university education, having attended under-resourced schools. The programme employs different intervention strategies to see these students through the course before joining mainstream programmes. When it comes to examinations, the dictionary seems to become very handy for ESP students.

What is disappointing from a lexicographic point of view is that a significant number (34 out of 79) examination candidates who used the dictionary responded negatively to the question which sought to establish whether or not they were generally regular dictionary users in their academic lives. Most of them indicated that they did not experience language problems that require regular dictionary support as their vocabulary was good enough. However, triangulating this explanation with what some students had looked up, e.g. *prolonged, impetus, government, power, etc.*, suggests otherwise. It highlights a generally poor dictionary culture among South African students (Gouws 2013; Nkomo 2015; 2016; Taljard et al. 2011) who overstate their knowledge while underestimating the potential benefits of dictionary use. There are others who benefit from a lexicographic intervention such as internet dictionaries but do not classify them as dictionaries because for them only printed dictionaries qualify to be called thus. Responses such as "I use the internet" were unsurprisingly prominent.

Although the number (40) of candidates who claimed to be regular dictionary users is higher than those who claimed not to be, most of them are ESP students. One of them indicated that they are required to buy two dictionaries, one general-purpose English dictionary and a science dictionary for the *Introduction to Scientific Concepts and Methods* module. One ESP lecturer who supports dictionary use in examinations confirmed that she encourages students to buy and use dictionaries in their studies in connection with language problems. Most of the ESP students indicated that they were regular dictionary users in their studies because they struggled with English. Some even emphasised the value of dictionaries in their disciplines such that dictionaries are an essential part of their educational toolkits.

7. Specific needs of examination candidates from the dictionary

The need for meaning emerged most prominently during examinations that were observed, as is the case of dictionary use in general (Atkins 1998; Chen 2012; Dziemianko 2010; Nesi and Hail 2002; Rundell 1999; Scholfield 1999). 74 out of a total of 79 students who completed Questionnaire 3 consulted the dictionary for meaning information. The only other information type that was reportedly sought was spelling, albeit by a few candidates. Unfortunately, it could not be

established whether spelling is such an emphasised skill in their studies that would also be marked in examinations.

The potential need for meaning was also high among new students who completed Questionnaire 1. Typical answers in response to why the respondents thought they would need a dictionary in examinations included the following:

- (To) understand certain terms
- Explaining difficult jargon
- Technical jargon might prove difficult
- So that we can have a better way of understanding concepts
- To know the real definition of the terms, especially in Biology/Chemistry

The same applies to returning students who responded to Question 12 of Questionnaire 2. They had to choose at least one of the five options, namely, understanding English, expressing oneself in English, understanding discipline-specific language, understanding one's own language and other reasons that had to be specified. From the 26 dictionary users in previous examinations, the need to understand the language of teaching a certain subject (terminology and register) featured most prominently (10 times), followed by the need to understand the language of teaching (English) generally. Again, it is the need for meaning information that is generally highlighted.

It is important to determine the nature of linguistic forms for which examination candidates need meaning information. Are the problems linguistic or pedagogic? Appendix 4 lists the items whose meaning was sought during the June examinations. Based on the context of usage in the respective examinations, only a few of those items, e.g. *socio-economic* (Sociology), *impetus* (Information Systems), *prolonged* (Introduction to Science Concepts and Methods), *disillusionment* (History), *disingenuous* (History), *legitimate* and *illegitimate* (History) may be considered to be general words without specialised disciplinary designations. There are others which were isolated from expressions in which they may have developed special meanings and looked up as separate words, e.g. *fundamental* instead of *fundamental contradiction* (Political Philosophy), *liminality* instead of *cultural liminality* (Art History and Visual Culture), *ubiquitous* instead of *ubiquitous intelligence* (Information Systems), among others. However, the majority of lexical items that were looked up appear to have discipline-specific designations and meanings; they fall within the terminological scope or registers of particular disciplines, as indicated by the context of usage of underlined items in the examination questions presented below:

- Which of the following does not apply to data?
 - (a) Is a fact of any kind
 - (b) Raw material
 - (c) Elementary
 - (d) Intangible (*Information Systems 201*).

- Part of the library's long term-vision is to strive towards securing unrestricted access to research outputs (i.e. supporting the Open Access movement). This is an example of:
 - (a) Democratization of knowledge
 - (b) Ubiquitous intelligence
 - (c) Secularization of knowledge
 - (d) Commoditization of knowledge (*Information Systems* 201).
- Discuss how Dictyoptera protect their eggs, and the defensive adaptations of Phasmida (*Entomology 202 — Theory Paper 1*).
- Define monophyletic and paraphyletic. Discuss the implication for understanding biodiversity, and give examples of their influence on modern classification of insect orders (*Entomology 202 — Theory Paper 1*).
- Discuss the concepts of ambivalence and cultural liminality found at the core of *mas* and *carnival* performances... (*Art History and Visual Culture 1*).
- Write an essay on constitutionalism as theory of limitation. Start by defining a) the state, b) government and authority (*Puzzles in Contemporary Political Philosophy and Introduction to Political Theory, Paper 1*).
- The Rwanda genocide can be understood as a massive failure to negotiate the fundamental contradiction. Write an essay in which you discuss the role of Hamitic hypothesis in this failure (*Puzzles in Contemporary Political Philosophy and Introduction to Political Theory, Paper 1*).
- Distinguish between matter and anti-matter, giving an example of each (*Introduction to Science Concepts and Methods, Paper 1*).
- What does the word zapping mean? (*Introduction to Science Concepts and Methods, Paper 1*).

The above examples illustrate that the lexical items that were looked up constitute vital aspects of the knowledge that was being assessed. As articulated in the *TermNet* motto that "There is no knowledge without terminology", disciplinary knowledge and terminology are inextricably linked. This suggests that the items that were looked up represent concepts that were studied in class, as indicated by some students. In that case, dictionary use in the respective examinations was not only a question of students having language problems but also problems related to what they have been taught and probably how they have (not) engaged with it beyond the lectures. It is in such cases where dictionary use *may* unduly benefit students, although the problem may be rooted in the manner of formulating the questions as indicated by Stalnaker and Stalnaker (1934: 119).

8. Dictionary assistance for examination candidates

In lexicography, a distinction has been made between general-purpose dictionaries and specialised dictionaries (Béjoint 2010; Gouws and Prinsloo 2005). The former deal with a particular language in a holistic and general way, whereas the latter deal with specialised languages and knowledge constituting specific disciplines. However, modern dictionaries may be considered to be typological hybrids as their efforts to address different user needs result in them combining features drawn from different types of dictionaries (Béjoint 2010). General-purpose dictionaries also include lexical items traditionally belonging to specialised dictionaries treating vocabulary of different specialised disciplines and subjects. Furthermore, when such lexical items are included, enough indications are given regarding the disciplinary associations of such vocabulary, including specialised definitions. However, publications that have analysed definitions of specialised and scientific terms in general dictionaries indicate that the definitions are not always helpful as they are sometimes littered with mistakes or completely false (cf. Nielsen 2015).

Different dictionaries, namely *The Concise Oxford Dictionary* (COD), *The South African Concise Oxford Dictionary* (SACOD) and its second edition entitled *The Oxford South African Concise Dictionary* (SACOD2) are used at Rhodes University during examinations. In terms of typology, they all fall under the category of general-purpose dictionaries. However, the university seems to have been under the impression that one dictionary was being used. Since June 2014, the instructions on examination question papers have specifically referred to the COD. This dictionary was originally published in 1911 based on the *Oxford English Dictionary* (OED). The idea was to produce a medium-sized dictionary offering a concise description of English out of the comprehensive OED. SACOD was adapted from the 10th edition of COD for South African users and therefore contains words and phrases considered to be mainly South African while its second edition is an update which captures more recent developments. The SACOD2 contains "a large number of scientific and technical terms, legal phrases and other specialist terms, especially those likely to be encountered by the general public and students" (Kavanagh 2002: vii). This also applies to the other two dictionaries. Accordingly, most words listed in Appendix 4 are available in these dictionaries, including *dictyoptera*, *monophyletic* and *paraphyletic* from a highly specialised subject such as entomology. The examples in Table 1 below illustrate that the idea that dictionaries do not provide explanations of meaning that are relevant to specialised disciplines is far from true. The dictionaries explain the terms that students looked up in ways that may provide assistance that addresses not only language but also subject knowledge gaps. However, it is the brevity of the definitions that would leave questions regarding the adequacy of assistance that students get in relation to complex questions about specialised subjects and topics.

Table 1: Examples of explanations of meaning from dictionaries used during Rhodes exams

Term	Dictionary	Explanation of meaning(s)
discourse	SACOD	1 <i>literary</i> (a) conversation; talk. (b) a dissertation or treatise on an academic subject. (c) a lecture or sermon. 2 <i>Linguistics</i> a connected series of utterances; a text.
government	SACOD2	1 the act or manner of governing. 2 system by which a state or community is governed. 3 (a) a body of persons governing the state. (b) (usu. Government) a particular ministry in office. 4 the state as an agent. 5 Gram. The relation between a governed and a governing word ...
liminal	SACOD	<i>technical.</i> 1 of or relating to a transitional or initial stage. 2 at a boundary or threshold
light year	SACOD2	1 Astron. the distance light travels in one year, nearly 6 million million miles. 2 colloq. a long distance or great mile
dictyoptera	SACOD	<i>Entomology.</i> an order of insects that comprises the cockroaches and mantises
epigenetic	SACOD2	1 Biol. (a) relating to epigenesis. (b) due to external not genetic influences. 2 Geol. formed later than the surrounding rock.
phagocytosis	SACOD	<i>Biology.</i> the ingestion of bacteria or other particles by phagocytes and amoeboid protozoans
monophyletic	SACOD2	Biol. (of a group of organisms) descended from a common evolutionary ancestor or ancestral group, esp. one not shared with any other group

Besides providing specialised explanations of meanings, the two editions of SACOD go further than the COED by including subject-field labels which indicate the disciplines, e.g. Astronomy, Biology, Linguistics, etc., in which the lexical items are used. In the case of polysemous lexical items doubling up as general words and subject-specific terms and registers, e.g. *government*, or even those that are used across disciplines with different meanings, e.g. *discourse*, the labels serve to disambiguate the senses by guiding the dictionary user to the appropriate disciplinary meaning. Not surprisingly, the majority of students who looked up discipline-specific terms during the June 2014 examinations indicated that their needs were successfully addressed. However, students' successful dictionary consultations and their claimed successes may not be nec-

essarily translated into successful examination question responses or performance of assessment tasks. Such success could not be determined without access to students' examination scripts, and this was beyond the scope of the present study.

However, there are other students who were unsuccessful in getting the needed information from the dictionaries they consulted. In part, this was a result of inappropriate dictionary use. For example, students who looked up *fundamental contradiction* in politics would have been unfortunate because what they looked for is a context-embedded expression rather than a disciplinary term, while those who looked up *fundamental* separately would not have benefitted either because they would have gotten only a general sense of the word. The same applies to *ubiquitous knowledge* and *ubiquitous* in the case of an *Information Systems* examination.

9. Efficacy of dictionary use under examination conditions

Notwithstanding its (possible) benefits to its users, under examination conditions, the dictionary should not be used at all costs. Certainly not at the inconvenience of other examination candidates. It is most probably along this reasoning that certain rules, albeit not documented as part of the Rhodes guidelines for invigilators, have to be applied when dictionary requests arise and are dealt with by invigilators. During an examination, when a candidate requests a dictionary, the invigilator is supposed to hand it over to the student, wait as the student consults the dictionary and collect it immediately to make it readily available for the next student who makes a similar request. The candidate should not overstay his/her consultation time with the dictionary.

Generally, students who used the dictionary during the 2014 examinations reported through Questionnaire 3 that the dictionary was generally handed on time and benefitted them. However, some candidates were concerned that they were not allowed to bring their own dictionaries and that the university supplies no more than two copies per venue. One candidate who identified herself as a regular dictionary user from her primary school days reckoned that being allowed to bring own dictionaries would benefit her more instead of having to regularly request one only for the invigilator to wait impatiently for her to finish. This actually links up with a concern that was expressed by first year students that they would not use a dictionary in an exam and the returning students who indicated that they had not previously used the dictionary because it could cost them valuable examination time, a concern raised in relation to other aids and notes used in open-book examinations (Boniface 1985: 201). However, this particular candidate even indicated that there were instances where she would do without the dictionary despite experiencing the need fearing that she would be deemed unintelligent because of her frequent dictionary use during examinations. Despite what was reported through ques-

tionnaires, observations also indicated some problems which result in the practice failing to adhere to the stipulated guidelines. That some examination candidates admitted that they were not regular dictionary users outside examinations manifested itself in the form of observed poor dictionary skills. Some students would page back and forth trying to find the words, with panic and frustration indicating that they were indeed struggling. This would result in them failing to get what they were looking for and even keeping the dictionary longer, with the invigilator or other students waiting.

Another problem pertains to students who seemed to be generally unknowledgeable about dictionary use. Instead of looking up meaning promptly and continue with the exam, some students seemed to regurgitate dictionary data, looking in the dictionary, writing a few words on the answer script, looking again, and writing again. Thus some students were not processing dictionary data to apply it to their answers. Instead, they seemed to take the data for answers. Yet no regulation empowers invigilators to monitor how exactly the dictionary is used. This makes the instruction stipulating that invigilators check what students need in Law examinations even difficult to apply. The management of dictionary requests and dictionary use is only one of the several responsibilities of the invigilators. Giving space to a current dictionary user and the need to also attend to other candidates sometimes means a prolonged hold of the dictionary by one candidate at the expense of others. During the *Introduction to Science Concepts and Methods* exam, where more than 60% of the candidates used the dictionary, there were instances when the need for the dictionary was so high that several students would wait long for the dictionary. The student with the dictionary would also be under extreme pressure, something that any intervention under examination conditions should be addressing rather than exacerbating.

Perhaps the most problematic issue about the dictionary use practice discussed in this article pertains to the use of different dictionaries. For the university, it probably does not matter, given the popular misconception that all dictionaries dealing with the same language are the same. Unfortunately, students may get different forms of assistance depending on the venue of their examination and the copy that would be available at the time of request. The SACOD adapted the OED to make it accessible for South Africans while the 2nd edition of the SACOD did not only seek to update the dictionary but also to increase its accessibility. The accessibility levels of the dictionaries vary. For example, the word *genocide*, which was looked up by students writing the Political Science examination is defined as "the mass extermination of human beings, esp. of a particular race or nation" by the OED, while the SACOD (2nd Ed.) defines it as "the deliberate killing of a very large number of people from a particular ethnic group or nation". Both definitions are factually correct, but the reference to "mass extermination of human beings" instead of "killing of a very large number of people" in the OED may require other students, especially

non-mother speakers of English, to further look up *extermination*. In other words, the COD definition may be more accessible to mother-tongue speakers of English than to their non-mother-tongue counterparts whom the practice is thought to support more. Such a flaw is fatal to the claim that the dictionary is a linguistic intervention for English additional language learners during examinations at Rhodes University.

10. Conclusion

This paper has presented an empirical picture of dictionary use under examination conditions at Rhodes University. This picture displays some irregularities regarding the practice. Unlike the lexicographer, "a harmless drudge" according to Samuel Johnson's (in)famous definition in his landmark *A Dictionary of the English Language*, the dictionary in examinations is not always harmless, depending on its type and how it is used. That students looked up and managed to obtain curriculum-relevant information from the dictionary for examination purposes is not surprising from the perspective of pedagogical lexicography. What is rather surprising is the naivety of academics who believe that dictionaries are merely linguistic tools with no relevance to the broader academy. Not only English additional language students use dictionaries in examinations, but also their mother-tongue counterparts. While a few students looked up words to confirm spelling and meaning of words that would impede their comprehension of examination questions, the majority needed subject-related information, including what should have been covered in their courses. More importantly, the dictionaries used provide such information, although the accessibility levels vary across the three dictionaries that are used, with mother-tongue English speakers having a better chance of benefiting more since the explanations are in their language.

Although dictionary use in examinations seems to be irregular according to the present study, this does not reduce the pedagogic value of dictionaries. Like the other aids used in open-book examinations, the dictionary may become harmful in the hands of a candidate who is not an experienced dictionary user or the one who has not prepared well for the examination, as observed in this study. As Stalnaker and Stalnaker (1934: 117) argue:

The case for open-book examinations involves the importance of knowing where to find information, how to evaluate it, and how to use it.

Academics also need to play a role in developing dictionary skills and think about how dictionaries can be used as resources in the curriculum, and formulating questions in a manner that may effectively assess rather than compromise what is being assessed. Not only will dictionaries prove to be useful linguistic tools, but also alternative pedagogical resources.

Acknowledgement

The author is indebted to Simthembele Matyobeni and Sisonke Mawonga who were the research assistants in this study. The two research assistants were responsible for administering Questionnaires 1 and 2, as well as data capturing.

References

Dictionaries

- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton/T. & T. Longman et al.
- Kavanagh, K. (Ed.)**. 2002. *South African Concise Oxford Dictionary*. Cape Town: Oxford University Press.
- Thompson, D. (Ed.)**. 1995. *The Concise Oxford Dictionary of Current English*. Ninth edition. Oxford: Clarendon Press.
- Van Niekerk, T. and J. Wolvaardt (Eds.)**. 2010. *The Oxford South African Concise Dictionary*. Second edition. Cape Town: Oxford University Press.

Other Literature

- Atkins, B.T.S. (Ed.)**. 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- Atkins, B.T.S. and M. Rundell**. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Barnes, A. et al.** 1999. Dictionary Use in the Teaching and Examining of MFLs at GCSE. *Language Learning Journal* 19(1): 19-27.
- Béjoint, H.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Bensoussan, M.** 1983. Dictionaries and Tests of EFL Reading Comprehension. *ELT Journal* 37(4): 341-345.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bishop, G.** 2000. Developing Learner Strategies in the Use of Dictionaries as a Productive Language Learning Tool. *Language Learning Journal* 22(1): 58-62.
- Boniface, D.** 1985. Candidates' Use of Notes and Textbooks during an Open-book Examination. *Educational Research* 27(3): 201-209.
- Boughey, C.** 2002. 'Naming' Students' Problems: An Analysis of Language-related Discourses at a South African University. *Teaching in Higher Education* 7(3): 295-307.
- Chen, Y.** 2012. Dictionary Use and Vocabulary Learning in the Context of Reading. *International Journal of Lexicography* 25(2): 216-247.
- Cowie, A.P.** 2002. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press.
- Dziemianko, A.** 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.

- Gouws, R.H.** 2010. The Monolingual Specialised Dictionary for Learners. Fuentres-Olivera, P.A. (Ed.). 2010. *Specialised Dictionaries for Learners*: 55-68. Berlin: De Gruyter.
- Gouws, R.H.** 2013. Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, V. (Ed.). 2013. *Specialised Lexicography*. Berlin: De Gruyter.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Graham, T. et al.** 2003. The Use of Graphics Calculators by Students in an Examination: What Do They Really Do? *International Journal of Mathematical Education in Science and Technology* 34(3): 319-334.
- Kemp, M. et al.** 1996. Graphics Calculator Use in Examinations: Accident or Design? *Australian Senior Mathematics Journal* 10(1): 36-50.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablets to the Computer*. Cambridge: Cambridge University Press.
- Ministry of Education.** 2002. *Language Policy for Higher Education*. Pretoria.
- Monaghan, J.** 2000. Some Issues Surrounding the Use of Algebraic Calculators in Traditional Examinations. *International Journal of Mathematical Education in Science and Technology* 31(3): 381-392.
- Nesi, H. and R. Haill.** 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.
- Nielsen, S.** 2015. Legal Terms in General Dictionaries of English: The Civil Procedure Mystery. *Lexikos* 25: 246-261.
- Nkomo, D.** 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.
- Nkomo, D.** 2016. An African User-Perspective on English Children's and School Dictionaries. *International Journal of Lexicography* 29(1): 31-54.
- Rhodes University.** 2006. *Language Policy*. Grahamstown: Rhodes University.
- Rhodes University.** 2014. *Revised Language Policy*. Grahamstown: Rhodes University.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-54.
- Scholfield, P.** 1999. Dictionary Use in Reception. *International Journal of Lexicography* 12(1): 13-34.
- Stalnaker, J.M. and R.C. Stalnaker.** 1934. Open-book Examinations. *The Journal of Higher Education* 5(3): 117-120.
- Taljard, E. et al.** 2011. The Use of LSP Dictionaries in Secondary Schools — A South African Case Study. *South African Journal of African Languages* 31(1): 87-109.
- 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. and R.H. Gouws.** 2012. School Dictionaries for First-Language Learners. *Lexikos* 22: 333-351.
- TermNeT.** <http://www.termnet.org/> (Accessed 23 March 2016).
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEX'eter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer Verlag.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

Examination papers cited in the article

Introduction to Science Concepts & Methods Paper 1. Examination. June 2014. Rhodes University.

Information Systems 201. Examination. June 2014. Rhodes University.

Entomology 202 — Theory Paper 1. Examination. June 2014. Rhodes University.

Art History and Visual Culture 1. Examination. June 2014. Rhodes University.

Puzzles in Contemporary Political Philosophy and Introduction to Political Theory, Paper 1. Examination.
June 2014. Rhodes University.

Introduction to Science Concepts and Methods, Paper 1. Examination. June 2014. Rhodes University.

Legal Pluralism. Examination. June 2014. Rhodes University.

Appendix 1

DICTIONARY USE IN EXAMINATIONS

As a student at Rhodes University, you will get a chance to use a dictionary during examinations. According to the language policy of the university, academic departments are required to provide you with "access to a wider range of dictionaries in examinations" (Rhodes University Language Policy 2006: 5). This survey is meant to find out what you think about this practice and about dictionaries in general. Please be as honest as possible, there are no right or wrong answers. All the material collected through this questionnaire will only be used for purely academic purposes.

1. Personal data
 - (a) Programme/Degree:
 - (b) Language spoken at home:
 - (c) Other languages:
 - (d) Type of school of matriculation (Mark with an X below)

2. How often did you use a dictionary (or dictionaries) at school?

Almost everyday	Every week	Once in a while	Never

3. **If you answered *never***, why did you never really use a dictionary (or dictionaries) at school?

We were not allowed to use dictionaries in class	
Dictionaries were not available at school	
I did not own a dictionary	
I did not know how to use a dictionary	
I was afraid that using a dictionary would make me look ignorant	
I did not see the need to use a dictionary	

4. If you did use a dictionary at school, what kind of dictionary did you use?

An English dictionary that explains English words in English	
A dictionary that explains English words in my own language	
A dictionary that explains words in my language using that language	
A dictionary that gives English words for words in my own language	
A dictionary that deals with a school subject and the language used in that subject, e.g. Mathematics	

5. If you did use a dictionary at school, what was the name(s) of the dictionary (or dictionaries) that you used? (If you **don't know** or **can't remember** the name please say).

6. Were you allowed or encouraged to use dictionaries at school?

YES	
NO	

Please explain the reason why you were allowed/encouraged or not allowed/discouraged to use dictionaries.

7. Were you ever taught how to use dictionaries at school?

YES	
NO	

8. In connection with which subjects did you use a dictionary? (**List them below**).

9. During school examinations, did you have challenges which you think you would have solved if you were allowed to use a dictionary in the examination room?

YES	
NO	

10. If you answered *yes* above, please identify the challenge(s).

Understanding the language used for teaching, e.g. English	
Saying what I wanted to say in the language used for teaching, e.g. English	
Understanding the special language used in teaching a certain subject, e.g. Science	
Understanding my own language	

11. As a university student, do you think you would need a dictionary?

YES	
NO	

12. If you answered *no* above, why do you think you will not need a dictionary?

13. If you answered *yes* above, why do you think you will need a dictionary?

14. As a university student, do you intend to buy a dictionary?

YES	
NO	

15. If you answered *no* above, why don't you intend to buy a dictionary?

16. As a university student, would you need a dictionary in an examination?

YES	
NO	

17. If you answered *yes* above, what do you think you will need a dictionary for in the exam?

18. If you answered *yes* above, what kind of dictionary would you need?

19. If you answered *no* above, why don't you think you will not need a dictionary in the exam?

20. As a university student, would you think it is appropriate for dictionaries to be used in an examination?

YES	It would help students understand the language used for asking questions	
	Students would get explanations of concepts in their areas of study	
NO	Students would get answers to questions asked in the examination.	
	It will take most of students' time	

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE

Appendix 2

DICTIONARY USE IN EXAMINATIONS

As a student at Rhodes University, you are allowed to use a dictionary during examinations. According to the language policy of the university, academic departments are required to provide you with "access to a wider range of dictionaries in examinations" (Rhodes University Language Policy 2006: 5). This survey is meant to find out whether you have utilised such an opportunity and how useful this has been to you. There are no right or wrong answers. The survey is just interested in your experiences and perceptions. All the material collected through this questionnaire will only be used for academic purposes.

1. Personal data
 - (a) Programme/Degree:
 - (b) Other degree obtained from this university:
 - (c) Language spoken at home:
 - (d) Other languages:
 - (e) Type of school of matriculation (Mark with an X below)

Ex-Model C	Private	Township	Rural	Foreign

2. As a university student, are you encouraged to use dictionaries in your studies?

YES	
NO	

3. As a university student, do you personally use a dictionary in your studies?

YES	
NO	

4. **If you answered *no* above, why don't you use a dictionary (or dictionaries)?**

I do not own a dictionary	
I do not know how to use a dictionary	
I did not see the need to use a dictionary in my studies	
Other (Please specify)	

5. If you answered *yes* in (3) above, why do you use a dictionary?

6. If you answered *yes* in (3) above, how often do you use a dictionary (or dictionaries) in relation to your studies?

Almost everyday	Every week	Once in a while

7. If you answered *yes* in (3) above, what kind of dictionary do you use?

An English dictionary that explains English words in English	
A dictionary that explains English words in my own language	
A dictionary that explains words in my language using that language	
A dictionary that gives English words for words in my own language	
A dictionary that deals with a school subject and explaining the language used in that subject, e.g. Mathematics	
Other (Please describe)	

8. If you answered *yes* in (3) above, do you own a dictionary (or dictionaries)?

YES	
NO	

9. If you answered *no* above, why don't you own a dictionary?

10. If you do use or own a dictionary, what is the name of that dictionary (You can list more than one and if you **don't know** or **can't remember** the name please say).

11. During your university examinations, have you ever encountered challenges which have prompted you to request and consult a dictionary?

YES	
NO	

12. If you answered *yes* above, please identify the challenge(s).

Understanding the language used for teaching, e.g. English	
Saying what I want to say in the language used in examinations, i.e. English	
Understanding the special language used in teaching a certain subject, e.g. Science	
Understanding my own language	
Other (Please specify)	

13. If you answered *yes* to Question 11 above, did the dictionary assist you?

YES	
NO	

14. If you answered *no* to Question 11 above, how did the dictionary fail to assist you?

15. As a university student, do you think it is appropriate for dictionaries to be used in an examination?

YES	It would help students understand the language used for asking questions	
	Students would get explanations of concepts in their areas of study	
	Other (Please specify)	
NO	Students would get answers to questions asked in the examination.	
	It will take most of students' time	
	Other (Please specify)	

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE

Appendix 3

DICTIONARY USE IN EXAMINATIONS

During a recent examination (June 2014), you requested a dictionary and consulted it in connection with a specific aspect of your examination. This short questionnaire is part of a research that seeks to establish the specific aspects that result in dictionary use in examinations at Rhodes University and its effectiveness. The research has obtained ethical clearance from the university (see Ethical Clearance). You are requested to complete the questionnaire (it should not take more than 5 minutes of your time). There are no right or wrong answers and there is no danger whatsoever associated with this research, neither are there any personal benefits that you will get from your direct participation. However, the information that you provide will give useful insights regarding the policy that permits students to use dictionaries in the exam. Some of the data will also be used in a paper that will be presented at an international conference and published in an academic journal. Your name and student number will never be used. Please refer to the consent form and complete it as well. Should you have any queries, please contact me through e-mail.

1. Personal data
 - (a) Degree for which you are studying:
 - (b) Year of study:
 - (c) Language spoken at home:
 - (d) Other languages:
 - (e) Type of school of matriculation (Mark with an X below)

Ex-Model C	Private	Township	Rural	Foreign

2. What specific aspect(s) of your examination required you to request a dictionary? (In the case of meaning of words on the question paper, please list those word(s))

Was the dictionary given to you on time?

YES	
NO	

3. Did you have any challenges when using the dictionary to solve the problem that you mentioned in 2 above? If *yes*, please describe the challenge(s) after ticking the appropriate answer in the box below:

YES	
NO	

4. Did the dictionary finally provide the assistance you needed?

YES	
NO	

5. If you answered *no* in (5) above, please explain a bit how your problem was not solved?
-

6. Do you have any suggestions that you think would improve the effectiveness of dictionary use in the examinations at this university? Please tick or explain as appropriate below.

Providing more copies of the acceptable dictionary	
Providing a wide range of dictionaries	
Allowing students to bring their own dictionaries	
Other	

7. Are you regular user of dictionaries in relation to your university studies?

YES	
NO	

8. If you answered *yes* in (8) above, why?
-

9. If you answered *no* in (8) above, why?
-

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE

Appendix 4

Some lexical items that were looked up during the examinations of certain subjects

Subject	Lexical item
Art History	Ambivalence
Art History	Liminality
Art History	Myth
Entomology	Monophyletic
Entomology	Dictyoptera
Entomology	Paraphyletic
History	Prerogative
History	Canonical
History	Disingenuous
History	Elusive
History	Disillusionment
Information Systems	Elementary
Information Systems	Ubiquitous
Information Systems	Impetus
Law	Attributive responsibility
Law	Maintenance
Law	Intangibility
Microbiology	Vector
Microbiology	Reservoir
Microbiology	Epigenetics
Microbiology	Natural killer cells
Microbiology	Phagocytosis
Microbiology	Tuberculosis
Microbiology	Black death
Political Science	Discourse
Political Science	State
Political Science	Government
Political Science	Power
Political Science	Authority
Political Science	Genocide
Political Science	Fundamental
Political Science	Constitutionalism
Political Science	Deployed
Political Science	Legitimate
Political Science	Illegitimate
Science	Zapping

Science	Prolonged
Science	Matter
Science	Anti-matter
Science	Distinguish
Science	Bout
Science	Shelf-life
Science	Light year
Sociology	Demographic
Sociology	Socio-economic

Planning a Dictionary for Mother Tongue Education: A Conceptual Framework for Gabonese Languages*

Blanche Nyangone Assam, *Department of Foreign Languages,
University of the Western Cape, Bellville, South Africa*
(bassam@uwc.ac.za)

Abstract: The present article is a plea for mother tongue education dictionary projects in Gabonese lexicography. The latter has been in a fast-developing process for the past twenty years and has experienced quite an important crop of dictionary products in such a short time. However, the production of dictionaries for mother tongue education, especially for Gabonese native languages, has been totally neglected. Only a very limited number of projects, which have ever since remained at the level of studies, do exist. This article provides a conceptual plan for mother tongue education dictionary production in Gabonese native languages. It successively suggests an organisation plan and a dictionary conceptualisation plan.

Keywords: CONCEPTUALISATION PLAN, DICTIONARIES, MOTHER TONGUE EDUCATION, GABON

Résumé: Planification d'un Dictionnaire pour l'Enseignement en Langue Maternelle: Un Cadre Conceptuel pour les Langues Gabonaises? Le présent article est un plaidoyer pour des projets de dictionnaires en vue de l'enseignement en langue maternelle dans la lexicographie gabonaise. Cette dernière connaît un développement rapide depuis ces vingt dernières années et a vu un nombre assez important de dictionnaires publiés en une période de temps limitée. Cependant, la production des dictionnaires pour l'enseignement en langue maternelle, particulièrement en langues locales gabonaises, est totalement négligée. Seul quelques projets restés depuis au stade d'études existent. Cet article propose un cadre conceptuel pour la production des dictionnaires comme outil d'enseignement en langue maternelle pour les langues locales gabonaises. L'article présente successivement un plan d'organisation et un plan de conceptualisation.

Mots-clés: CONCEPTUALISATION, DICTIONNAIRE, ENSEIGNEMENT EN LANGUE MATERNELLE, GABON

1. Introduction

This article in the form of a conceptual framework paper comes within the

* This article is based on a chapter of a D.Litt. Dissertation, *Dictionaries as Teaching Instruments for Mother-tongue Education: The Case of Fang in Gabon*.

overall plan and intention of introducing Gabonese native mother tongues into the education system of Gabon. The intention has been repeatedly stated by the Gabonese government during the General States Meeting for Education and Training held in Libreville in 1983 (Kwenzi Mikala 1990) and during the General States Meeting for Education, Research and Training-Employment Adequacy held in Libreville in 2010 (République Gabonaise 2010). Since then, a number of research works in the field of linguistics have been undertaken to fulfil such a plan and intention. In this regard, Nzang-Bie (2001: 17) states that one of the priorities of research in linguistics (and related disciplines) within the context of Gabon should be the "*introductions of local languages in national education curricula*". For Nzang-Bie (2001), who advocates for bilingual or multilingual education in Gabon, this process implies the compilation and production of textbooks and dictionaries not only for bi- or multilingual education but also for mother tongue education in rural areas.

In the field of lexicography, it has been stated that the development of lexicography and dictionary production would certainly contribute to mother tongue education implementation in Gabon (Emejulu 2001 and 2003; Mavoungou 2010). For this purpose, initiatives of pedagogical lexicography for Gabonese languages by Mabika Mbokou (2001 and 2006), Nyangone Assam (2001, 2002 and 2006) and Ella (2002 and 2007) constitute an important contribution to the strategic plan for Gabonese lexicography (Ndinga-Koumba-Binza 2005).

The current article is a follow-up to these enterprises by Mabika Mbokou and Nyangone Assam. Its relevancy is grounded on the fact that no lexicographic project related to mother tongue education has ever been undertaken in Gabon for the past ten years. Thus, the article is an attempt to revamp not only past projects, but also most importantly the whole focus on mother tongue education lexicography in Gabon. The content of the article mainly focuses on Fang as language for illustration and reference.

The planning phase in a lexicographic process can be divided into two distinctive plans, i.e. the organisational plan and the dictionary conceptualisation plan. The former will be presented in section 3 of this article, and the latter in section 4. Prior to these two sections, the article starts with a short rationale.

2. Bilingual dictionaries for mother tongue education: A case for Gabon

The making of a dictionary does not start with the defining of words. It is preceded by a great deal of planning. Gouws (1999: 41) argues that the planning of the compilation of a dictionary starts with the planning of the system to be utilised in that specific dictionary. Dictionaries must first be conceived and then written. Planning is the design phase for creating both the text specifications and the way the whole text will be written. Gouws (1999: 41) further maintains that before the lexicographer puts one word on paper in the compilation process, he/she has to be well aware of the precise structure each article should display so that the system can be applied with meticulous care.

The interest here is the sequence of work to be undertaken for a language, which does not have yet a fully developed standard form. The decision about what dictionary to plan depends largely upon the situation of the language itself and upon the state of its lexicographic treatment. Zgusta (1971) lists the steps that lexicographic planning has to follow: the first step is a very small dictionary concerned primarily with orthographic rules and their application in particular cases of single lexical units.

The next step is a bilingual dictionary in which the target language is the vernacular and the source language is the one used by the specific society in its contact with modern civilisation. In the specific case of Gabon, Gabonese native languages should actually be the target languages and French the source language. However, the fact is that apart from bidirectional dictionaries (e.g. Mickala-Manfoumbi 2004), all recently published Gabonese language dictionaries rather have French as the target language (e.g. Mavoungou and Plumel 2010, Van der Veen and Bodinga-bwa-Bodinga 2002, ILALOK 2008, etc.). This common trend stems from early Gabonese dictionaries compiled by missionaries and colonial administrators (e.g. Biton 1907, Biton and Adam 1969, Raponda-Walker 1934, etc.).

Meanwhile, the general situation in Gabon suggests that a bilingual dictionary would be more advantageous. Firstly, it brings a greater number of people into contact with the cultural patterns represented by the foreign language in question and thus increases the number of people for whose activities the development of a standard is necessary. Secondly, such a bilingual dictionary helps more effectively to remove the onomasiological gaps. Certainly, a monolingual dictionary can also try to introduce the necessary terminological problems, but it is difficult to find them there, because they are not generally known. On the contrary, an engineer or a teacher, for example, who knows the technical terms of his/her field of interest very well in French or English and tries to write in such subject in his/her own language, will easily find the necessary expressions in the bilingual dictionary.

Thus, the bilingual dictionary is a more effective tool for this purpose than the monolingual one. It should also not be forgotten that it is difficult and often impossible to couch the definition of lexical meaning, necessary in a monolingual dictionary, in a language without a fully stabilised standard form or that lacks many expressions. The compilation of a monolingual dictionary is meant to empower the speech community and that there is no sense in starting with a multi-volume dictionary similar to the *Oxford English dictionary* for languages without any dictionaries. They first need other dictionaries because a multi-volume dictionary will take many decades if not centuries to complete and, during that time, they will still be without any communicative assistance. Once again, the needs and the situation of a specific speech community must guide the process.

This article intends to make a case for developing bilingualized school dictionaries that will help in mother tongue education. The acquisition of both French and a Gabonese native language at this level of education is essential and

affect the rest of these learners' lives. It gives them a well-balanced perspective of both languages and cultures, they have the opportunity to develop into individuals who know their identity as Gabonese mother tongue speakers, but who are also exposed to the global world through an international language.

3. The organisational plan

The organisational plan takes into account all the planning around the project before anything has been done. It can be seen as the business plan of the lexicographic process. Gouws (2001: 65) indicates that the organisational plan is a basic and compulsory component of every lexicographic process. It is primarily directed at the logistics of the project and all the managerial aspects. This is essential for the success of any dictionary project, and the logistic and the managerial infrastructure must precede any editorial work.

Lexicography is a costly activity. It is rather infrequent that phenomenon lexicographic activities are turned into self-supporting undertakings. The usual situation is that funds must be found to finance the project. In any case, when negotiating financial support and when submitting plans, the lexicographer should be very careful in his/her estimate of both the necessary time and the money. In Gabon, language-related research projects receive funding from government public institutions, foreign diplomatic representations or individual benefactors. Most of the latter are politicians or former politicians who wish to have their language and culture in the process of development. The prominent illustration of this situation is given by the Pove language, which has had two dictionaries and a grammar published in the past ten years by Mickala-Manfoumbi (2004, 2012 and 2013). Mickala-Manfoumbi (2004: 9) states that his research on the Pove language was personally funded by Mr. Guy Nzouba Ndama¹, a mother tongue speaker of the language.

For mother tongue education lexicography projects, it is believed that the same current funding schemes will be followed. However, it is advisable that international funding should also be searched and applied for. In addition, most academic institutions receive public funding annually for their regular operations. These institutions should also consider supporting lexicography research projects that are undertaken by their respective research units.

Furthermore, the lexicographer's greatest struggle is the struggle against time. Zgusta (1971: 348) states: "I certainly do not know all the lexicographic projects past and present; but of those I know not a single one was finished in the time and for the money originally planned". Suggesting a definite proposal for the issue of time in the specific case of Gabon in terms of dictionary projects would not be an easy task. As a matter of fact, there would hardly be better suggestion other than to have an important number of workers in a single project. In addition, this implies an increase of funding, because not all staff in a dictionary project would agree to work for free intensively in a limited space of time.

Nowadays, lexicographic projects carried out by a single person result in

smaller dictionaries, apart from a few exceptional cases. The usual situation is that there is a staff, of which the most important members are the editors or the sub-editors grouped around one or two chief editors. In Gabon, where there is no research unit specifically devoted to lexicography, the need for lexicographic projects coincide with the need for establishing lexicography units (Ndinga-Koumba-Binza et al. 2017). For that purpose, this study will refer to the framework for the planning of a lexicographic project suggested by Van Schalkwyk (1999). In the strategic planning of the Bureau of the WAT, Van Schalkwyk stresses nine steps for good and efficient planning of a lexicographic project:

1. Mission statement
2. Areas of strategic focus
3. Description of the situation
 - 3.1 Internal environment
 - 3.2 External environment
4. Analysis of the environment
 - 4.1 Weak points (constraining factors) in the internal environment
 - 4.2 Strong points (supporting factors) in the internal environment
5. Environmental tendencies
 - 5.1 Threats (constraining factors) in the external environment
 - 5.2 Opportunities (facilitating factors) in the external environment
6. Scenario
7. Assumptions regarding the future strategic position (FSP)
8. Objectives
 - 8.1 Long-term objectives
 - 8.2 Medium term objectives
 - 8.3 Strategic aims
9. Plan of action

Once it is decided, what kind of dictionary it should be and what its scope should be, and especially once the target user of the dictionary and his/her needs are known, a mission statement can be formulated. The organisational plan of a dictionary project cannot start without first determining one aspect that has been neglected in so many dictionaries, but which forms an integral part of the organisational plan. This aspect is the identification and formulation of the *genuine purpose* of the intended dictionary. Gouws (2001: 65) emphasises that it is an exercise of extreme importance to ensure a sound theoretical point of departure for the compilation process.

3.1 The genuine purpose and target users group

In order to be a functional instrument, a planned dictionary has to fulfil the purpose of empowering the learners in their language acquisition endeavour. Wiegand (1998) refers to this purpose as the genuine purpose. Gouws (2001: 66) emphasises the importance of including a clear and unambiguous exposition of the genuine purpose of the dictionary to be compiled in the organisational plan of a dictionary project. According to Gouws (2001: 66), the genuine purpose of a dictionary is co-determined by, among others, its typological nature and its intended target user group.

The target user group of Gabonese pedagogical lexicography dictionaries will primarily be the learners in primary schools, most of which will be in the process of learning their mother tongues. Adults participating in literacy programmes may also find interest in such kind of dictionaries. After an assessment of the user group's needs and reference skills, the choice of the typological nature of an intended school dictionary of a Gabonese language is a bilingualized dictionary, which has features of both monolingual dictionary and a bilingual dictionary.

Gouws (2001: 66) also points out that the first dictionaries to be compiled for a specific language purpose are usually either desk or standard bilingual or monolingual dictionaries. The genuine purpose of such dictionaries is to transfer, by means of lexical data, information regarding the set of lexical items included as treatment units in order to ensure the linguistic empowerment of the target user.

The bilingualized school dictionary will be produced so that the learner who uses the dictionary in a context of learning his mother tongue will have an instrument to assist him in achieving a successful dictionary consultation procedure by reaching the goals that motivated the search. In this case, the learner finds linguistic information as well as knowledge information that enable him to learn about the cultural aspects of his/her mother tongue. This genuine purpose of the intended school dictionary has major implications for the dictionary conceptualisation plan because of its direct impact on the structure and contents of the dictionary articles and the data distribution pattern. In agreement with Gouws (2001), the formulation of the genuine purpose of a dictionary can be regarded as a response to the needs of the intended target users, the needs of the learners here are the motivation behind any decision regarding the typological choice, as well as the structure, contents and the presentation of the intended bilingualized school dictionary.

The organisational plan of the project has to make provision for managerial infrastructure able to cope with the compilation of this school dictionary as it fits the requirement implied and identified by its genuine purpose. The organisational plan of the future bilingualized school dictionary starts with the mission statement. It introduces the dictionary plan for all managerial aspects.

3.2 Mission statement

Formulating the mission statement of a project or an enterprise comprises identifying its aim and task in such a way that the nature of the work clearly emerges. The aim of this project is to compile dictionaries, within lexicographic units in Gabonese languages in order to serve indigenous languages and the language users. The line of function of that lexicographic unit will be to compile, as soon as possible, pedagogical dictionaries for the mother tongue education in primary schools and for adult literacy programmes. This line of function will be supplemented by the following functions:

- (i) To endeavour to bring about appropriate lexicographic products for the different languages of Gabon;
- (ii) To promote national languages through mother tongue education;
- (iii) To empower national languages speakers through literacy campaigns via their respective mother tongues;
- (iv) To raise the general level of lexicography in Gabon by cooperation with external dictionary projects, e.g. the *Woordeboek van die Afrikaanse Taal* (Dictionary of the Afrikaans Language, South Africa).
- (v) To actively address the imbalances by training and generating work; and
- (vi) To endeavour to foster quality and a unity of character, and to cater for the needs of the user in all its activities and products.

3.3 Areas of strategic focus and language data collection

Two essential questions need to be answered: are the right things being done? Are the right things being done right? The answers enable the lexicographic team to consider carefully those areas of the project in which it needs to succeed if it is to remain true to its aims. Such an analysis yields the following areas of performance: language data collection, editorial processing, printing and layout of the dictionary, marketing and research. The nature and the extent of the lexicographic project determine whether all these areas will be applicable. However, all the performances areas of the lexicographer are incorporated into the planning of the project. Further analysis of each of these performance areas is necessary in order to have a precise indication of their content.

This article will not discuss all these areas but it gives a view on the language data collection. Language data collection involves the excerpting of actual speech material from written sources and spoken sources. The selection, sorting, control integration, safekeeping and computerisation of the material are all part of this performance area. A language data collection policy will be needed to ensure that the collection of data proceeds purposefully. This policy ensure that a corpus is constructed for the project. The construction of a database is a costly and time-consuming task. For this reason, it is necessary to

determine in advance, what is to be collected and how it should be collected. The possibility of utilising an existing corpus is an alternative to an own corpus. For instance, in the case of Fang, a French corpus could be used because of the lack of terminological items for certain fields. However, no dictionary project should be without an authoritative corpus.

One possibility for obtaining or compiling an authoritative corpus for Fang would be the amalgamation of existing small corpora, i.e. those assembled by Afane-Otsaga (2004), Nyangone Assam (2006), Ella (2007) and Ekwa Ebanega (2007). This amalgamation will have certain advantages such as

- (i) producing a bigger corpus for the Fang language,
- (ii) the possibility to produce various dictionaries of different types from one single corpus, and
- (iii) the possibility of regular improvement and additions into the corpus.

The compilation of the Fang amalgamated corpus will learn from the experiences of the YCY corpus (Ndinga-Koumba-Binza and Mabika Mbokou 2010), a common wordlist including at least 100 000 most frequently used similar words of three related languages, i.e. Yipunu, Civili and Yilumbu. The corpus was compiled from various existing dictionaries in the three languages as well as from a number of existing different corpora, i.e. Mabika Mbokou (2006), Mihindou (2006), Tomba Moussavou (2007) and Soami (2010) for Yipunu; Blanchon (1990), Mabika Mbokou (1999), Ndinga-Koumba-Binza (2000 and 2008), Loembe (2005) and Tele-Pemba (2009) for Civili; and Emejulu and Pambo (1990), Mavoungou (2000 and 2002), Mboumba (2009), Saphou-Bivigat (2010) and Mavoungou and Plumel (2010) for Yilumbu.

Amalgamated corpora of this kind can be made for various Gabonese languages on the basis of the groupings of Gabonese languages by Kwenzi Mikala (1998) into language-units. According to Ndinga-Koumba-Binza (2017), each Kwenzi Mikala's language-unit can be divided into sub-groups based on lexical similarities and mutual intelligibility.

Each lexicographic team would have also to consider Gabonese languages actually still unwritten languages. Therefore, the collection of data will follow the collection criteria for unwritten languages as indicated by Singh (1982: 88). He states that for unwritten languages the data collected according to the field method with the help of informants. The criteria for the selection of informants, their age, sex, cultural and psychological qualities like intelligence, memory, alertness, patience, honesty, dependability, cheerfulness, etc. must be considered.

The fieldworker should approach and deal with the informant in a way that will elicit as much data as required without either causing annoyance to the informant or antagonism in him/her. This will ensure faithful and proper elicitation of the proper language data. That is why the speech community must be involved and kept enthusiastic about a dictionary in their language. For instance, in Gabon, the Fang speech community is commonly known to be

very enthusiastic about their language. The number of informants to be employed depends on the scope and the type of dictionary.

If all the regional varieties of the language are to be included, informants should be selected from all regions, in order to ensure optimum data, it is advisable to select more than one informant for every variety. This would also help on checking and rechecking the data. In order to elicit lexical units of as many varied types as possible, it would be advisable to select from all the following groups: men and women; people of all ages; people belonging to different economic and social groups.

Many typical items common among women may not be elicited from male informants and vice versa. Many unwritten languages are fast coming under the influence of some other languages regarding modernisation and technology. As a result, new lexical items are introduced in the stock of the languages. For instance, in Fang, there is an equivalence for the French lexical item *avion* (aeroplane) and there are many other lexical items, which the older generation were familiar with. The younger generation has adopted the borrowed lexical item, which leads to a general loss of older lexical items in the language. Only the older generation knows many of the lexical units, which presently have a low usage frequency.

Therefore, the older informants would be useful for providing lexical units of the type mentioned above. The younger generation informants would be equally useful for providing new words introduced into the language. The representation from different social groups will ensure the inclusion of words from those groups. For the preparation of a dictionary of an unwritten language, the lexicographic team or the lexicographer should have knowledge of the life and culture of the speech community. For eliciting words in an unwritten language, word lists of especially a modern language like French or English might be utilised (Zgusta 1971). Nevertheless, in making use of word list there are limitations to be considered.

The list may contain a good number of lexical items that are quite unknown, or even irrelevant, to native language situations. The general response to enquiries about such items is either that the informant gives generic words for specific objects or tries to coin some equivalents. Some of the newly coined lexical items are very artificial. Their artificiality can be tested by getting them checked with other speakers of the language. In some cases, the native speaker confesses total ignorance of a lexical item. There is a total unawareness of the local environment and objects in such lists.

How can a lexicographer ensure maximum elicitation for a dictionary of such a language Singh (1982: 91) suggests that a list of basic words belonging to different semantic domain and grammatical classes may be tentatively prepared for elicitation of data for unwritten languages the semantic domains include for instance: nature, earth, sky, natural events, geographical and astronomical items, directions, winds, weather, season, etc. mankind, sex, family, relationship, body parts, bodily functions and conditions, diseases and cure.

Clothing and personal adornments; food and drink, methods of preparation; dwelling, part of the house and furniture, etc. cooking utensils, tools, weapons, etc. flora and fauna (including part of animal anatomy, diseases, cures, etc.); occupations and professions, equipment, rituals and customs connected with them; road and transport; sense of perception; emotions, temperamental, moral and aesthetic (including insults, curses etc.); government, war, laws; religion; education; games and amusement, entertainment, music, dance, dram; metals; numerals and systems of enumeration; measurement of time, space volume, weight, quantity; functions words including classifiers; fairs, festival, customs, beliefs, etc. verbs: physical activity, instrument verbs, verbs of fighting, music verb, motion verbs, occupational verbs, culinary verbs, cosmetic verbs, communicative verbs, stationary verbs, sensory verbs, emotion verbs, other verbs.

This list is not exhaustive. It might be treated as a sort of referee point and related words on the semantic domain might be elicited based on this list. For example, while collection word about agriculture, words about the different agricultural products, the sowing and harvesting time, rituals and ceremonies connected with them, names of the different parts at different times of growth of these products and the verbs connected with different actions connected with them might be elicited. The data collected by using word list might not be adequate, especially from a semantic view. According to Samarin (1967: 208): "*the chief failure of the field dictionary is that it indicates not so much the meaning of word, but the fact that they exist. They do not define, they document*". The data for a dictionary collected from the word list should be supplemented by the data from different types of discourses, some of which are listed as follow (Samarin 1967: 2018):

- (i) **Narration:** eyewitness account, reminiscences and instructions on how to perform certain tasks or how to get to certain destinations.
- (ii) **Conversation:** arguments, dialogues about 'where you have been'.
- (iii) **Songs:** lullabies, dirges, dance songs. Folk tales: legends, how things come to be, amusing stories, proverbs and riddles.
- (iv) **Names:** personal, topographic, village, pseudo-onomatopoeic calls of animals or birds.

The collection of data from unwritten languages still has many problems. One of the most important problems is the segmentation and identification of a word from the phonetic continuum of texts, in written languages, which have a tradition of grammar; there are certain devices and fixed criteria to identify a word. According to Singh (1982: 94), "*a word in written languages is generally identified as a meaningful unit, a cluster of sounds or letters written between spaces, or with potential pauses. In unwritten languages, there is no such device. The lexicographer has to analyse the data, to make a grammatical analysis of the language, and then to fix the word and the lexicographic unit*". The collection of material from oral and written sources will serve the building of a corpus. The corpus for the intended

dictionary has to take into account the needs of the target group, learners in school. The learners need text reception and text production skills in the language being learned. A more comprehensive representation of data and additional guidance is necessary.

3.4 Description of the situation

It is important for proper planning to negotiate the situation in which a project finds itself. The internal and external environment determine this situation. The internal environment of a project comprises everything that is within the control of the project. All other matters that can influence the activities of the project but fall outside its control are regarded as part of the external environment, for example, the politico-economic circumstances prevailing in the country. A sound grasp of the external environment makes those involved more sensitive to the milieu in which they operate.

In the internal environment of this project, the collection and selection of language data, the editorial processing and layout depend on the lexicographic team. When the planning is thorough, the task becomes much easier and less time consuming. The finance and support services depend on the external environment.

The nature and extent of the reference and research work that the lexicographic team undertakes must be considered during the planning phase. A lexicographer's efficiency is compromised in the absence of reference material and library facilities. If these cannot be utilised elsewhere, they will have to be made available. The purchase and maintenance of office equipment, and computers and software requires large capital expenditures, which must be considered thoroughly during this planning phase.

In fact, all supplies needed for a lexicographic project must be calculated and taken into account so that it can be incorporated into the budget.

3.5 Analysis of the environment

The success of a project or enterprise depends largely on whether those involved are capable of identifying and acting upon the weak as well as the strong points in the internal environment. The weak points of a project increase its vulnerability in a changing world. If the project is aware of them, all the mechanisms can be put into operation to soften their effect as far as possible or even to eliminate them. The strong points of a project in the internal environment present the project with the opportunity to take proactive steps to ensure success.

The weak point of this project will be the lack of trained people to carry out the fieldwork for the language data collection. It is important to involve linguists and others interested in language matters in the project. Of course, the

language speech community must be kept enthusiastic about their language and the project. Sound metalexigraphic training of the team must be planned and encouraged. This will ensure that the Fang speaking community will get involved in Fang lexicography, the Punu speaking community in the Punu lexicography, the Nzebi speaking community in the Nzebi lexicography and dictionary production, etc.

The strong point is that the government supports the project entirely. It will create jobs and promote the local languages. It will introduce the culture of dictionary use among the young and the older generation.

3.6 Environment tendencies

Threats or constraining factors in the external environment must be identified, analysed and overcome, because it is in this way that the risks to which a project is exposed are reduced. A sound grasp and utilisation of the opportunities or facilitations factors in the external environment can assist in neutralising the constraining factors.

In a mother tongue education dictionary project, the only great threat could be a possible change in the politico-economic situation of Gabon because it is a government project and they will supply all financial and material support needed. There is also the threat that decisions to be taken by linguists and lexicographical experts regarding language issues become bureaucratic and political. An awareness of the endangered situation of local languages will prevent such turn.

3.7 Scenario

The aim of a scenario in a mother tongue education dictionary project would to outline a potential future in which the project must function with possible linguistic changes taken into account. This can give an indication of the challenges awaiting the project, but it also give an indication of the impact that the decision currently taken will have in future. A scenario can help the decision makers to evaluate the effect their decision critically.

In five years' perhaps, the language policy would not have changed because French will remain the sole official language. There will however be other changes because the government has recognised the necessity of mother tongue education. Acting on this awareness, it has initiated the training of teachers at university level for mother tongue education. This teaching section will be functional in five years' time. This means that it invests in the preparation of pedagogical books and dictionaries through the Ministry of Education. The training of translators, interpreters and broadcasters will also take place to anticipate the various needs of the Ministry of Communication and parliament. The direct and evident manifestation will be the compilation of language dic-

tionaries alongside other pedagogical material and textbooks. The establishment of lexicographic units is therefore a matter of urgency in the future. The development of lexicography in Gabon depends on the implementation by government of these programmes.

3.8 Assumptions regarding the future strategic position

Planning is always directed towards the future and, since the future is not known, planning involves working with assumptions. These assumptions are supported by the achievements of the project, by its current situation and level of development, as well as by the vision of those involved.

It can be envisaged that in the next five years for a mother tongue education dictionary project starting at the current period, the project would have trained lexicographers to start working. The collection of language data and computerisation should be advanced. There should be experienced linguists, language practitioners as well as any other necessary language experts.

3.9 Objectives

In order to carry out its functions efficiently and effectively, to strive for excellence in living up to its stated mission, and to remain on tract and not to lose its vision, an enterprise or project must work according to explicitly formulated aims and objectives.

The long-term objective of a mother tongue education dictionary project is to provide Gabonese languages with dictionaries to revitalise languages that are now endangered.

The medium-term objective will be to build a corpus of the main languages that will serve for dictionaries. The strategic aim is to start the establishment of a lexicographic unit.

3.10 Plan of action

Planning is seen as a team effort and is of no value if it cannot be brought to fruition. Efficient planning and management of goals keep the time and money expenditures of a lexicographic project within bounds. An estimation of the time and capital needed for a lexicographic project demands extensive preparation. The planning of a lexicographic project is practically impossible without extensive lexicographic experience. This issue is more and more being solved with the constant training of Gabonese lexicographers at university level. Nyangone Assam et al. (2016: 180) reports that "*training is the one aspect where the strategic planning of Gabonese lexicography has been quite successful*". Gabon has been for the past twenty years in the process of raising a generation of qualified and experienced lexicographers.

When lexicographic planning is carried out, the most important things to bear in mind are the specific lexicographic needs of the given language and its dictionary users. These needs will lead the decision makers to adapt the necessary organisational approach. In the planning of any kind of dictionary, the emphasis should always be on the needs of the intended users. This can be illustrated with the specific case of Yilumbu for which the need was twofold, i.e. first the language has never had a dictionary previously, and second the language is under great influence attacks from the neighbouring languages which are Civili in the town of Mayumba, Yipunu in the town of Moulengui-Binza and in Dousségoussou region, and Varama and Nkomi in the towns of Gamba and Sette-Cama (Mavoungou 2002). The metalexicographical planning conducted by Mavoungou (2002 and 2012) resulted successively in a general bilingual translation dictionary (Mavoungou and Plumel 2010) and in a translation dictionary of idiomatic phrases (Mavoungou and Ndinga-Koumba-Binza, submitted).

Furthermore, during the planning of a dictionary of a not yet stabilised language, the lexicographer must make up his/her mind about which policy to follow. It is during this phase that the lexicographer has to negotiate some important issues, e.g. the functions, structure and contents of the dictionary. He/she decides which variety of the language will be used in the dictionary. While doing so, the users' needs should guide his/her choices on the lexical items to be included or not. Considering this case, the first target group of users will be schoolchildren, which implies a need for the core vocabulary. The context of communication and the vocabulary in daily use will be adequate for learners.

For the Fang language for instance, Afane-Otsaga (2004, 2010a and 2010b) has been insisting that for the stabilisation and/or standardisation of Fang, the variety of the language spoken in Oyem (capital city of the Woleu-Ntem province² in the north of Gabon)

4. The dictionary conceptualisation plan

The dictionary conceptualisation plan takes into account all the decisions made about the compilation process. It can be considered as the practical side of the lexicographic process. According to Wiegand (1998: 151), the dictionary conceptualisation plan can be divided into five subdivisions: the general preparation phase, the material acquisition phase, the material preparation phase, the material-processing phase and the publishing phase.

Gouws (2001: 65) points out that the organisational plan can be regarded as a model formal the primary lexicographic process for that dictionary. It concerns all the managerial aspects dealt with previously in the present study. He suggests that the dictionary conceptualisation plan represents a model for the secondary lexicographic process of a specific dictionary. The genuine purpose of the future bilingualized school dictionary is to enable the learners to achieve

successful retrieval of information in the dictionary consultation procedure.

It aims to give them a practical instrument to help in the acquisition of their mother tongue via French as the source language. The needs and the reference skills of the target user group determines the genuine purpose of the dictionary, which is therefore the basis for the dictionary conceptualisation plan. The conceptualisation plan for a bilingualized school dictionary is presented according to the following aspects identified by Wiegand (1998).

4.1 The general preparation phase

The general presentation phase of the dictionary conceptualisation plan lays the foundation for the structure, contents and presentation of the final product. Gouws (2001: 68) identifies three issues to tackle in the general preparation phase. The compilation of a lexicographic instruction book or a lexicographic style guide is the first assignment of the team members. The second assignment to achieve is the microstructural programme of the dictionary. A third issue to be dealt with in the general preparation phase is the identification, establishment, nature, extent and description of a dictionary basis.

4.1.1 The instruction book for or style guide

The style guide includes instructions for the making of a dictionary. According to Bergenholtz (1990), the lexicographic style guide should contain a comprehensive description of a system applied in the dictionary. The editorial process will be much more efficient with the setting up of a style guide, which is a set of generalisations with accompanying illustrations from the envisaged dictionary. It is a set of instructions showing the order of components, typological presentation and input instructions for the database. The style guide is needed to ensure confidence and consistency on the part of the lexicographer, i.e. to ensure a uniform character and prevent deviation from the rules.

It is a learnable system for new lexicographers and contains detailed guidelines on how to treat each type of data category, how to treat difficult or prominent semantic and grammatical phenomena (homophony, polysemy, etc.) and other important data, for instance: spelling convention; data on pronunciation; parts of speech; cross-references, list of labels, list of editorial abbreviations; definition or translation equivalent, example material, homonymy, polysemy and synonymy; treatment of sensitive lexical items; treatment of cultural data.

General editorial principles have to be established. A rough style guide of data categories abstracted from samples of dictionary articles should be used for consistency by compiling computerised system. It is important to refine sample entries and the style guide to reflect changes during the compilation of the dictionary. A frequent update of the style guide is required. The computer-

ised system design helps to root out inconsistencies in the style guide, especially in the treatment of senses and sub-senses. It makes it easier to apply the style guide consistently, especially with regard to the ordering of data and the formatting of dictionaries articles.

Bergenholtz (1990) includes more of the typical issues to be dealt with an instruction book, for example the lemmatisation process (with reference to, for instance, the influence of an initial capital letter, diacritics, the order of the word and stem forms), the use of typographical and non-typographical markers in the articles, the making of different senses of a lemma, the use of abbreviations in the metalanguage of the dictionary, and the positioning and marking of new search zones in the article.

A style guide has to be developed. It shall cover all aspects of lexicography relevant to the dictionary to be compiled. It should preferably be an alphabetically arranged reference guide for making the manuscript, starting, for instance, with the abbreviations an editor may use and ending with the treatment of verbs. It should discuss the different kinds of definitions and when to use them, synonymy and semantic opposition, polysemy and homophony, the ordering of senses, and the treatment of gender and derogatory items. The style guide covers in detail in all the types of data that need to be presented in the envisaged dictionary. It has to be indicated in a consistent pattern so that it facilitates the work of the compilers. It should give clear guidance on labelling and the presentation of information. Typography and punctuation should be discussed in detail, for instance that all lemmas should be in bold print and grammatical data like parts of speech and plurals should be in italics.

The development of a style guide is essential as it serves as a training manual for new editors with little or no lexicographical training; it ensures that every editor makes the manuscript in the same way and that the dictionary eventually has a uniform character. It also prevents gradual deviation from the treatment of raw material originally decided on.

4.1.2 The microstructural programme

According to Gouws (2001: 68), the microstructural programme is another important issue in the general preparation phase that needs to receive attention. He states that at a very early stage of the lexicographic process, the team members should know the microstructural programme thoroughly. The microstructural programme involves the different data categories to be included in the treatment of the lemmata and the typical article slots allocated to these categories. It assists the team members to determine the quantitative extent of the dictionary. It enables team members to ensure a functional space budget because a dictionary always has space limitation and an early indication of the microstructural programmes provides good space management. Gouws (2001: 68) points out that a further value of the early identification of the microstructural programme is that it assists the lexicographers during the early phases of the

compilation process to focus on those data categories that will be included in the dictionary.

With regard to Fang, Nyangone Assam (2006: 85) indicates that in response to the genuine purpose and the functions determined by the needs of the users, the French-Fang bilingualized school dictionary would present an integrated microstructure. Gouws (2001: 87) defines integrated microstructures as one presenting a relation of direct addressing between a paraphrase of meaning/translation equivalent and its co-text entry/entries. It means that each paraphrase of meaning/translation equivalent is immediately followed, within the same sub-comment on semantics, by the co-text entry illustrating the typical usage of lexical item in question.

An integrated microstructure ideally suits this bilingualized dictionary because it makes the appropriate interpretation of the contents of the sub-comment on semantics for the user very easy. Especially in the treatment of a lexical item with many polysemous senses, the direct relation between co-text entry and paraphrase of meaning/translation equivalent ensures an optimal retrieval of information. It decreases text condensation because no other paraphrases of meaning/translation equivalences come between a given paraphrase of meaning/translation equivalent and its co-text entry. This is illustrated by the example below adapted from Nyangone Assam (2006: 86):

Abum [ábùm] nom. Pluriel *mebum* (Var: *amum*, *memum*)

1. Ventre, abdomen. *Makon abum, ma wok abum*, j'ai mal au ventre.
2. Mère. *Bine abum avor*, nous sommes issus de la même mère.
3. Grossesse. *E minga nyi ane abum*, Cette femme est enceinte.

The dictionary article above is an example of the treatment of a lemma in the integrated microstructure of the bilingualized school dictionary. The lemma **abum** (stomach) has polysemous senses: **1.** (stomach) the translation equivalents in French are given followed by the example sentences in Fang then in French to illustrate the usage of lemma in this specific sense. It is the same for the last two senses. There is a direct relationship between the translation equivalents and the co-text entry.

In the above article, the data categories that are included in the microstructure are the following:

- (i) **Comment on form:** an indication of the pronunciation is given for each lemma so that the learners will be helped with the **pronunciation** of each lemma. Pronunciation is given in the phonetic transcription and the tones for Fang lemmata will be included to give learners a solid basis for pronunciation. One of the main functions of the bilingualized school dictionary is to enable learners to read and write in their mother tongue Fang. The **class of the lemma, part of speech**, is indicated for the learners and they will know whether the lemma is a noun, a verb, an adverb, a pronoun, etc. the morphological data on the lemma, whether it is the **plural or**

diminutive forms, is also indicated when necessary to give learners a thorough perspective of the lemma treated. Fang is a language with different variants and the treatment makes provision for an **indication of the variants** when they occur.

- (ii) **Comment on semantics:** a translation equivalent and a definition are given followed by the example sentences in the language pair French and Fang. Idioms will be presented and treated as sub-lemmata treatment, but very few clusters will appear in the microstructure. Only those that are part of the learners' core language and have a high frequency of usage as indicated by the corpus will be given treatment as lemma. Collocations will not be treated as units, but each component, the base and the collocator of the combination, will be treated under its alphabet stretch. The collocations will be treated as illustrated examples in both the articles representing the base and the collocutor.

4.1.3 The dictionary basis

According to Wiegand (1998: 139), a dictionary basis can be described as the total of the source language data for the specific lexicographic process. This includes all the possible sources, which accommodate such material, as well as informants and mother tongue speakers of the language who can assist the editorial staff in the building up of a material collection. The dictionary basis of the planned school dictionary will be compiled from three types of sources. Wiegand (1998: 139) shows that the dictionary basis of a general monolingual or bilingual dictionary can be compiled from three sources. The primary sources of the dictionary basis will be all written material that reflects typical communication situations. The secondary sources are all available dictionaries in the specific language. The tertiary sources involve all other linguistic material that can be used, for example linguistic monographs, papers and grammars.

The dictionary basis of the school dictionary should be compiled for French and Gabonese native languages. As far as French is concerned, all three sources are available and abundant. The primary sources, written material that reflects typical communication are available since it is the language of wider communication and of instruction in Gabon. These include schoolbooks, storybooks, textbooks and all other material used in the curriculum at this level. Gabonese novels, poetry books and any creative writings in French can also serve as primary sources.

The secondary sources, available dictionaries in French, are abundant and will be easy to collect, but the collection of material must be guided by the purpose and functions of the envisaged bilingualized school dictionary. Gabonese French lexicography has also been in a developing process with a few Gabonese French dictionaries that have been published (Nyangone Assam et al. 2016). These dictionaries can also contribute as secondary sources.

The tertiary sources are the other linguistic material, particularly grammars, available in French for the primary education level. It is important to emphasise that cooperation within the team that included teachers and educational experts at the primary level is crucial.

Compiling the dictionary basis for a Gabonese native language is not as unproblematic as it is for French. All Gabonese native languages are unwritten and less-resourced languages with limited amount of written sources. Gouws (2001: 68) indicates that, although the primary sources will usually be texts, the dictionary basis of a dictionary compiled for a language with a strong oral tradition can also use speech recordings of the orature as primary sources.

4.2 Editorial processing

Editorial processing is one of the line functions of all dictionary projects, and includes the systematisation of linguistic material and the production of manuscripts and dictionary texts. The question as to what material should be included and what should be excluded is fundamental to the lexicographic team. This question will be answered in the light of the kind of dictionary that is compiled as well as its scope. The inclusion policy needs to cater for the macro- as well as microstructure of the dictionary. It must be determined which elements become lemmas and what lexicographic data should be supplied for each of these elements. Gouws (1997: 42) points out the importance of efficient well-planned editorial processing:

The macrostructure represents the lexicon and this selection may not be on a random basis. It is important that the compilation of any dictionary be dominated by well-defined principles regarding the collection and selection of lexicon items. The lemmata have to be drawn from a representative corpus of that specific language and a vital part of the planning of any dictionary is the development of a corpus. The successful retrieval of information in any dictionary often depends on an unimpeded access to the needed lemma sign.

Depending on the size of the editorial staff, one or more editors should revise the work of the others. Amongst the others, he/she must see it that the editorial manual is followed correctly to ensure that the dictionary has a uniform character. He/she has to look at the data and decide whether it has been interpreted correctly. Firstly, he/she must judge whether the inclusion of the lemma complies with the policy. He/she must check the pronunciation indicated, the grammatical information supplied, and so forth. He/she must ascertain whether the information in the definition is correct, and whether it has been formulated clearly and unambiguously. He/she has to check all the cross-references to ensure that the dictionary is reliable and closely-knit unit.

Fortunately, editing is only about checking up and correcting mistakes. It is also of invaluable help to younger colleagues who learn a tremendous amount from the changes suggested. In the beginning, the changes should always be discussed with the younger colleagues because this promotes the learning process.

4.3 A dictionary-making system and customisation of computer software

The corpus should be the point of departure from which the lexicographic team will abstract items to include and exclude from the pedagogical dictionaries that are needed for a specific dictionary project. However, it is vital for the project to have an authoritative database, which would also be utilised for dictionary projects.

Lexicographers' knowledge of dictionary-making systems could ensure that they gain as much as possible from systems developed from them. In order to make their contributions count in the design of dictionary-making systems, according to Harteveld and Louw (2004), they should consider the following:

1. Speed: The system should speed up the dictionary-making process as much as possible.
2. User-friendliness: The editor must find the use of the system and the editing of the dictionary entries on the system simple and sufficient.
3. Flexibility/adaptability: It should be easy to modify the system to suit individual or changing needs
4. Language support: All characters required by the lexicographer should be available.
5. Access to language data: Language data resources should be easily, if not simultaneously, accessible on the workstation used for editing.
6. Interface: Queries of manuscript data as well as other language resources should be possible by means of a simple interface.
7. Group work/Networking/Internet: It should be possible to be able to communicate, to exchange and to share material quickly and over long distances.
8. Compatibility with other/external data: The system should use file and data formats, which make exchange, sharing, and incorporation with other data simple and accurate.
9. Multiple output formats (electronic, printed): The system must allow for the output of data in both hard copy (printed) and electronic format (CD-ROM, online, Internet).
10. Cost: The system should be economical to run and maintain.

Harteveld and Louw (2004) advise that the steps indicated above should be a prerequisite for any adaptation and use of software in the editorial process. It is important to consider these because the cost and time of customisation will depend on the type and complexity of the dictionary being compiled. The program that will be provided initially may be a generic platform and will need to

be customised for each specific dictionary. New dictionaries may be created at a later stage and further customisation will then be required. The contract with the creators of the software must not only make provision for continued support of the software, but also for further customisation at minimal or no cost to the compilers. There must be training and support for the chosen tools, as well as a fixed period of familiarisation with the tools.

Computers have affected dictionaries in two major areas: production and research. On the production side, dictionaries can now be stored in databases, so that a change can be made very quickly and changes that affect the entire dictionary can be done in one-step. New editions of dictionaries can be produced without having to retypeset the entire work. Computers have already become important as a method for publishing dictionaries in electronic format instead of paper, and this will likely be increasingly important in the future.

For research, lexicographers can have instant access to corpora of the language and can thus look for example sentences of rare works, compare grammatical information, gather example sentences, and in general rely on real life examples of the language instead of having to fall back on personal impressions. They can also communicate rapidly with consultants across the world by using e-mail. It has become difficult for any lexicographic project to avoid the use of computers and all the tools that come along with it.

5. Conclusion

Dictionary making is a long, complex and time-consuming activity. As the work involved is vast, it is necessary that detailed planning be done before it begins. The compilation of every dictionary has to be preceded by the formulation of a lexicographic plan adhering to the aims of the typological criteria of that specific dictionary, and aimed at the specific needs and reference skills of a well-defined target user. This plan must not be done in a haphazard way, but has to be rooted in a general theory of lexicography.

According to Wiegand (1984: 14-15), one of the components of a general theory of lexicography is the theory of organisation. This includes all the activities leading to the drawing up of a plan that has to precede that compilation of every dictionary. There are some crucial issues to be considered for the lexicographic planning. For example, the planning cannot start without a decision about the type of dictionary because the macrostructure and the microstructure dictionary and a bilingual dictionary are different, and even more so in the case of school dictionary. The lexicographer should decide about the language of the dictionary, and the social and stylistic variations of the language. In this regard, all the structural components of a dictionary, including the macrostructure, microstructure, mediostructure and access structure come into play.

These decisions should be made before starting the actual work on the dictionary and should be strictly adhered to. All decisions must be printed and recorded to ensure continuity when new staff join the project. Instructions must

be completed and detailed. A blueprint for the project, which may contain descriptions and instructions for the collection of material, which sources mentioned, the compilation of the word list, and the articles of the intended dictionary, is advisable. Besides these details, the project should reflect the scope of the dictionary, its purpose and the target user group. The preparation of such a blueprint will not only help as a guidebook for the compilers in which the data distribution is explained, but can also be used to prepare the front and back matter texts.

In view of revamping mother tongue education dictionary perspectives with the strategic planning of Gabonese lexicography, the present article has provided a tentative conceptual framework for the production of such specific dictionaries in Gabonese native languages. While it can be recognized that Gabonese lexicography has embarked itself in a fast-growing developing process, current dictionary projects in progress and recently published dictionaries in Gabonese languages do not show any purpose for mother tongue education. This article has pleaded for mother tongue education dictionary projects in Gabon.

Notes

1. Mr Guy Nzouba Ndama was Speaker of Gabonese National Assembly for 20 years, and is currently one of the main leaders of opposition in Gabon.
2. Woleu-Ntem province is the only province in Gabon that is about 90% monolingual. The regional language of the province is Fang, the mother tongue of the vast majority of the population of this. There are also in very limited numbers speakers of Hausa and Baka, which are also indigenous to the province, but they also speak Fang on a daily basis. Woleu-Ntem is also the only Gabonese province where Hausa and Baka can originally be found. Although a few varieties of Fang can be found in four other provinces of Gabon (Estuaire, Moyen-Ogooué, Ogooué-Ivindo and Ogooué-Maritime), the Fang spoken in Woleu-Ntem shows different dialectal variants. Every major town in Woleu-Ntem, i.e. Oyem, Bitam, Minvoul, Mitzic and Medouneu has its own variety of Fang.

References

Dictionaries and Encyclopaedias

- Biton, A.** 1907. *Dictionnaire français-ndumu et ndumu-français, précédé d'éléments de grammaire*. Nantes: Imprimerie Bourgeois.
- Biton, A. and J.J. Adam.** 1969. *Dictionnaire Ndumu-Mbede-Français et Français-Ndumu-Mbede. Petite Flore de la Région de Franceville (Gabon). Grammaire Ndumu-Mbede*. Archevêché de Libreville. Bar-le-Duc: Imprimerie St Paul.
- ILALOK (Institut des Langues Locales au Kouilou).** 2008. *Dictionnaire Vili-Français/Mpisukulu Bi Kum' Bi Tshi Vili Ku Tshi Mputu*. Paris: L'Harmattan.
- Mavoungou, P.A. and B. Plumel.** 2010. *Dictionnaire yilumbu-français*. Libreville: Éditions Raponda-Walker.

- Mavoungou, P.A. and H.S. Ndinga-Koumba-Binza.** Submitted. *Dictionnaire des Expressions Idiomaticques Lumbu*. Louvain/Paris: Éditions Peeters.
- Mickala-Manfoumbi, R.** 2004. *Lixique Pove–Français. Français–Pove*. Libreville: Éditions Raponda-Walker.
- Mickala-Manfoumbi, R.** 2012. *Dictionnaire des Noms Personnels Pove*. Paris: Éditions Alpha-Omega.
- Raponda-Walker, A.** 1934. *Dictionnaire Mpongwè–Français, suivi d'Éléments de Grammaire*. Metz: Imprimerie de la Libre Lorraine.
- Van der Veen, L.J. and S. Bodinga-bwa-Bodinga.** 2002. *Gedandedi Sa Geviya. Dictionnaire Geviya–Français*. Leuven/Paris/Sterling: Peeters.

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.** 2010a. Alphabet et orthographe: critères, qualités, conditions et vulgarisation dans le cas du Gabon. Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et Standardisation des Langues Gabonaises*: 69-89. Stellenbosch: SUN PRESS.
- Afane Otsaga, T.** 2010b. Propositions pour l'orthographe du fang. Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et Standardisation des Langues Gabonaises*: 188-191. Stellenbosch: SUN PRESS.
- Alberts, M.** 1997. Lexicographic Needs Assessment. *PanSALB Lexicographic Consultative Meeting, Johannesburg, 31 October 1997*: 7-12.
- Bergenholtz, H.** 1990. Lexikographische Instruktionen für ein zweisprachiges Wörterbuch. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43: 19-37.
- Blanchon, J.A.** 1984. Présentation du yi-lumbu dans ses rapports avec le yi-punu et le ci-vili à travers un conte traditionnel. *Pholia* 1: 7-35.
- Blanchon, J.A.** 1990. Civili. *Revue Gabonaise des Sciences de l'Homme* 2: 141-142.
- Ekwa Ebanéga, G-M.** 2007. *A Microstructural Programme for Dictionaries in Fang*. Unpublished D. Litt. Dissertation. Stellenbosch: Stellenbosch University.
- Ella, E.M.** 2002. Un dictionnaire scolaire de langue de spécialité. Emejulu, J.D. (Ed.). 2002. *Éléments de Lexicographie Gabonaise. Tome 2*: 305-325. New York: Jimacs-Hillman.
- 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.** 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.
- Emejulu, J.D.** 2003. Challenges and Promises of a Comprehensive Lexicography in the Developing World: The Case of Gabon. Botha, W.F. (Ed.). 2003. *'n Man wat beur. Huldingsbundel vir Dirk van Schalkwyk*: 195-212. Stellenbosch: Bureau of the WAT.
- Emejulu, J.D. and F. Pambo Loueya.** 1990. Yilumbu. *Revue Gabonaise des Sciences de l'Homme* 2: 197-201.
- Gouws, R.H.** 1997. Lexicographic Planning. *PanSALB Lexicographic Consultative Meeting, Johannesburg, 31 October 1997*: 16-20.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Ed.). 2001. *Éléments de Lexicographie Gabonaise. Tome 1*: 58-94. New York: Jimacs-Hillman.

- Harteveld, T. and P.A. Louw.** 2004. *Course on the Editing Tools in Dictionary Making*. Stellenbosch: Bureau of the WAT.
- 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.
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography*. New York: The Scribner Press.
- Loembe, G.** 2008. *Parlons Vili: Langue et Culture de Loango*. Paris: L'Harmattan.
- Mabika Mbokou, L.** 1999. *Les Phénomènes Analogiques en Civili: Étude Phonologique et Morphologique*. Unpublished M.A. Thesis. Libreville: Omar Bongo University.
- Mabika Mbokou, L.** 2001. Le rôle du dictionnaire dans le système éducatif. Emejulu, J.D. (Ed.). 2001. *Éléments de Lexicographie Gabonaise. Tome 1*: 206-222. New York: Jimacs-Hillman.
- Mabika Mbokou, L.** 2006. *A Model for the Macro- and Microstructure of a Yipunu-French School Dictionary*. Unpublished D.Litt. Dissertation. Stellenbosch: Stellenbosch University.
- Mavoungou, P.A.** 2000. *A Frequency List of the Yilumbu Language*. Unpublished Study Conducted at the Bureau of the Woordeboek van die Afrikaanse Taal (WAT), Stellenbosch.
- Mavoungou, P.A.** 2002. *Metalexical 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.** 2012. Towards a Yilumbu Dictionary of Idiomatic Phrases. Ndinga-Koumba-Binza, H.S. and S.E. Bosch (Eds.). 2012. *Language Science and Language Technology in Africa: A Festschrift for Justus C. Roux*: 177-198. Stellenbosch: SUN PRESS.
- Mboumba, L.H.** 2009. *Quelques Problèmes Métalexicographiques en Yilumbu: Le Cas des Parties du Discours*. M.A. Thesis. Libreville: University of Omar Bongo.
- Mickala-Manfoumbi, R.** 2013. *Essai de Grammaire Pove*. Paris: Éditions Alpha-Oméga.
- Mihindou, G.-R.** 2006. *A Theoretical Model for a Yipunu-English-French Explanatory Dictionary of Medical Terms*. Unpublished D.Litt. Dissertation. Stellenbosch: Stellenbosch University.
- Ndinga-Koumba-Binza, H.S.** 2000. *Phonologie du Civili de Mayumba: Langue Bantu du Gabon (H12a)*. Unpublished M.A. Thesis. Libreville: University of Omar Bongo.
- Ndinga-Koumba-Binza, H.S.** 2005. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ndinga-Koumba-Binza, H.S.** 2008. *Phonetic and Phonological Aspects of the Civili Vowel Duration: An Experimental Approach*. Unpublished D.Litt. Dissertation. Stellenbosch: University of Stellenbosch.
- Ndinga-Koumba-Binza, H.S.** 2017. Language and Dialects in Gabon: An Analysis of Language-units towards Language Inventory. Prah, K.K. and L.M. Miti (Eds.). 2017. *Deconstructing of the African Tower of Babel. Between the Harmonization and Fragmentation of African Language Orthographies*. Cape Town: The Centre for Advanced Studies in the African Society.
- Ndinga-Koumba-Binza, H.S. and L. Mabika Mbokou.** 2010. *Combined Corpus Project of Yipunu, Civili and Yilumbu*. Paper presented at the Interim Conference of the African Language Association of Southern Africa (ALASA) held at the University of Botswana, Gaborone, Botswana, 21-23 July 2010.
- Ndinga-Koumba-Binza, H.S., B. Nyangone Assam and V. Ompoussa.** 2017. *Perspectives for Lexicography Units in Multilingual Gabon*. Paper presented at the 22nd International Conference of the African Association for Lexicography (AFRILEX) held at Rhodes University, Grahamstown, South Africa, 26-29 June 2017.

- Nyangone Assam, B.** 2001. La lexicographie dans l'apprentissage des langues gabonaises. Emejulu, J.D. (Ed.). 2001. *Éléments de Lexicographie Gabonaise. Tome 1*: 187-205. New York: Jimacs-Hillman.
- Nyangone Assam, B.** 2002. Le dictionnaire pédagogique scolaire: quelques aspects du traitement lexicographique. Emejulu, J.D. (Ed.). 2002. *Éléments de Lexicographie Gabonaise. Tome 2*: 292-304. New York: Jimacs-Hillman.
- 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., H.S. Ndinga-Koumba-Binza and V. Ompoussa.** 2016. What French for Gabonese French Lexicography? *Lexikos* 26: 162-192.
- Nzang-Bie, Y.** 2001. Vers une éducation multilingue au Gabon: première approche. *Revue Gabonaise des Sciences du Langage / Gabonese Journal of Language Sciences* 2: 17-29.
- République Gabonaise.** 2010. *Etats Généraux de l'Éducation, de la Recherche et de l'Adéquation Formation-Emploi. Les Actes Adoptés. Libreville, les 17 et 18 mai 2010*. Libreville: Ministère de l'Éducation Nationale, de l'Enseignement Supérieur, de la Recherche Scientifique et de l'Innovation.
- Samarin, W.J.** 1967. *Field Linguistics: A Guide to Linguistic Field Work*. New York: Holt, Rinehart and Winston.
- 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.
- Singh, R.A.** 1982. *An Introduction to Lexicography*. Manasagangotri, Mysore: Central Institute of Indian Languages.
- Soami, L.S.** 2010. *Towards the Development and Application of Representative Lexicographic Corpora for the Gabonese Languages*. Ph.D. Thesis. Stellenbosch: Stellenbosch University.
- Tele-Pemba, A.M.** 2009. *Éléments pour une Approche Comparée des Emprunts Lexicaux du civil du Gabon, du Congo-Brazzaville et du Cabinda: Proposition d'un Modèle de dictionnaire*. M.A. Thesis. Libreville: University of Omar Bongo.
- Tomba Moussavou, F.** 2007. *Metalexigraphic Criteria for a Monolingual Descriptive Dictionary Presenting the Standard Variety of Yipunu*. Ph.D. Thesis. Stellenbosch: Stellenbosch University.
- Van Schalkwyk, D.J.** 1999. Planning and Management — The Most Neglected Activities in Lexicography. *Lexikos* 9: 198-207.
- 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 Verlag.
- Wiegand, H.E.** 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.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia/The Hague/Paris: Mouton.

Direct User Guidance in e-Dictionaries for Text Production and Text Reception — The Verbal Relative in Sepedi as a Case Study*

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

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

Ulrich Heid, *Department of Information Science and Natural Language
Processing, Hildesheim University, Hildesheim, Germany and Department of
African Languages, University of Pretoria (heid@uni-hildesheim.de)*

and

Daniel J. Prinsloo, *Entelect, Melrose Arch, Johannesburg, South Africa
and Department of African Languages, University of Pretoria,
Pretoria, South Africa (dprinsloo@gmail.com)*

Abstract: This article introduces a prototype of a writing (and learning) assistant for verbal relative clauses of the African language Sepedi, accessible from within a dictionary or from a word processor. It is an example of how a user support tool for complicated grammatical structures in a scarcely resourced language can be compiled. We describe a dynamic light-weight tool aimed at combining user-knowledge with text production support, i.e., user-involved interactive text production of the complicated verbal relative in Sepedi. In this article, the focus is on access in a dictionary use situation. Although the tool is intended as a writing assistant to support users in text production; it also satisfies text reception and cognitive needs, but its focus is on solving text production issues related with the interaction between lexical items and complex grammatical structures in the African (Bantu) languages and for learning by users and/or training users in this interaction.

Keywords: WRITING TOOLS, USER-GUIDANCE, USER SUPPORT, TEXT PRODUCTION, E-DICTIONARIES, AFRICAN LANGUAGES, SEPEDI, COMPLEX GRAMMATICAL STRUC-

* This article represents follow-up work on an initial design study for user support in complex grammatical structures presented at *Euralex 2014* (Prinsloo et al. 2014). The website *SepediHelper.co.za* was introduced at *eLex 2015* (Prinsloo et al. 2015).

TURES, RELATIVE CONSTRUCTION

Opsomming: Direkte gebruiksleiding in e-woordeboeke vir teksproduksie en teksresepsie — die werkwoordrelatief in Sepedi as gevallestudie. Hierdie artikel stel 'n prototipe van 'n skryf- en (leer)hulpmiddel bekend vir werkwoordrelatiewe konstruksies in die Afrikataal Sepedi, wat vanuit 'n woordeboek of 'n woordverwerker toeganklik is. Dit dien as voorbeeld van hoe 'n gebruikershulpmiddel vir ingewikkelde grammatikale strukture in 'n hulpbronbeperkte taal saamgestel kan word. Ons beskryf 'n dinamiese liggewig hulpmiddel wat gemik is op die kombinerings van gebruikerskennis met teksproduksie-ondersteuning, dit wil sê, gebruikersbetrokkenheid by interaktiewe teksproduksie van die ingewikkelde werkwoordrelatiewe in Sepedi. In hierdie artikel is die fokus op toegang tydens 'n woordeboekgebruiksituasie. Hoewel die werktuig bedoel is as 'n skryfhulpmiddel om gebruikers in die produksie van teks te ondersteun, voldoen dit ook aan teksresepsie- en kognitiewe behoeftes. Die fokus is egter op die oplossing van teksproduksiekwessies wat verband hou met die interaksie tussen leksikale items en komplekse grammatikale strukture in die Afrikatale asook op die aanleer van taal deur gebruikers en/of die opleiding van gebruikers in hierdie interaksie.

Sleutelwoorde: SKRYFHULPMIDDELS, GEBRUIKERSLEIDING, GEBRUIKERSONDERSTEUNING, TEKSPRODUKSIE, E-WOORDEBOEKE, AFRIKATALE, SEPEDI, KOMPLEKSE GRAMMATIKALE STRUKTURE, RELATIEFKONSTRUKSIE

1. Introduction

Over the last ten years, several writing aid tools have been developed (see below for an overview). Their purpose is to support users who need to produce texts (e.g. in a language that is not their L1). This support can be obtained in a dictionary use situation or from a word processor. The focus in this article is on guidance in text production when using an e-dictionary for text production with verbal relatives in Sepedi. Such user support can either be obtained by checking words, sentences or paragraphs produced by the user, or by guiding him/her to adequate solutions. Most such tools focus on lexical choice (e.g. in collocations). For languages with complicated morphosyntactic structures, such tools should cover not only lexical choice, but also the interaction between lexicon and grammar. The South African African (Bantu) languages are a typical example of such languages, and we will use Sepedi as a case in point in the present article.

We will present the prototype of *Sepedi-helper*, a tool that can assist dictionary users in the construction of Sepedi verbal relative constructions, which are a typical example of the complexity that arises from the interaction between lexical choice and the grammatical system of the language. The prototype is presented in a stand-alone version, but the objective is to include it into an interactive e-dictionary, e.g. an English–Sepedi translation dictionary.

In the remainder of this introduction, we recall the main lines of the state of the art in writing aids; in section 2, we present the concept of direct user

guidance which underlies the *Sepedi-helper*. Sections 3 and 4 are devoted to the morphosyntactic properties of the relative construction and show the complexity involved in the interaction between lexical choice and the building-up of correct grammatical constructions. In sections 5 and 6 we show the principles underlying the writing support for Sepedi relatives, as well as the properties of the actual implementation of the tool. A "guided" tour from the dictionary user perspective follows in section 7 and we conclude in sections 8 and 9 with remarks on first experiences with dictionary users, as well as plans for future work. While we exemplify the principles of direct guidance for dictionary users on the Sepedi relative construction, we are convinced that more constructions from the African (Bantu) languages, as well as more generally any kind of interaction between lexical choice and grammatical (or morphosyntactic) constraints of a given language, could be dealt with along the same lines.

Writing tools have a great potential for user support in an e-dictionary, especially for text production but also for text reception of complicated grammatical structures in any language. Such tools should be designed to take the dictionary user's expertise into account in terms of the level or strategy for guidance provided.

Regarding text production, this article illustrates the working of a Builder for assisting users to write relatives in Sepedi (see Section 3). In a similar vein, the tool should be able to translate a Sepedi relative phrase into English. The nature of the support should typically also link to a user's level of knowledge of the grammatical system of L2 and should therefore take different user types, based on their knowledge of the L2 and their information needs into consideration (cf. Tarp (2008)), which can be summarised as follows:

- A user with a very limited knowledge of the language or a casual user, e.g., may prefer a machine translation option in the dictionary, with links to the grammar rules which may be consulted on demand. Cf. Bosch and Faaß (2014) as an example of direct user guidance to the correct answer in the compilation of possessive constructions in Zulu plus rule-based machine translation technology. Possessive constructions in Zulu can be regarded as complicated, requiring substantial knowledge of the nominal class system, possessive concords, exceptions to the formation rules, etc. which many inexperienced users may not have at the time of consulting the tool.
- On the other hand, a user who has a fair knowledge of the language may require a different type of support, e.g. through decision trees, i.e., a series of basic choices made by the user. Examples have been discussed for copulatives, kinship terminology, colour terms, etc. Cf. Bothma et al. (2013), Prinsloo and Bosch (2012), Prinsloo et al. (2011), Taljard and Prinsloo (2013).
- In certain cases, a user might benefit more from a bird's eye view through well-structured guidance paths such as tables and diagrams on e.g. kinship relations, grammatical moods and meanings, etc. Cf. Prinsloo et al. (2012).

Such technologies, integrated into the dictionary, may enable the user to find the correct information at an adequate level of detail and complexity required to solve his/her information need, thereby individualising the data presented to the user in terms of his/her information need. Cf. Bothma (2011), Fuertes-Olivera and Tarp (2014), Tarp (2008, 2011, 2012), Verlinde (2011).

The purpose of such tools is to guide users to the information they are looking for, i.e., without having to first study complicated grammatical structures in order to find the required information. We use the term *user support (technologies)* as an umbrella term for all such technologies. To date, we have described only the three technologies listed above. Additional such technologies and designs exist, e.g., *Interactive Language Toolbox* (<https://ilt.kuleuven.be/inlato/>), *Writing assistants and automatic lexical error correction: word combinatories* (Wanner et al. 2013), *A collocation writing assistant for learners of Spanish* (Alonso Ramos et al. 2014), user driven task and problem-oriented multifunctional leximats (Verlinde et al. 2010), online data-driven lexicographic instruments on foreign language learning (Buyse and Verlinde 2013), the work of Bertels and Verlinde on lexicography and corpus analysis (Bertels and Verlinde 2011), etc. All such techniques can be embedded in an e-dictionary and are intended to give information on demand, i.e., the user has the option to consult the tool if the "standard" dictionary article does not provide sufficient data to solve the user's specific information need. It would also be possible to embed such tools in a word processor, or for the user to consult such tools as stand-alone tools; this, however, is not the focus of this article and will not be addressed further. The focus is therefore on a tool embedded in an e-dictionary; access in this case is from the e-dictionary. The user therefore consults the dictionary, in the current case about the translation of the English word "who" into Sepedi (i.e., a translation/text production information need), and upon finding that (s)he needs more help than is available in the "standard" dictionary article, accesses the *Sepedihelper* on demand.

2. Direct user guidance as a support technique

Direct user guidance as an additional technique in the e-dictionary to provide user support for complex grammatical structures is not a solution for all user support. We regard it as a complementary technology that may be used in conjunction with other user support technologies for specific grammatical constructions, and the same way as these it should only be available to the user on demand, depending on the user's level of language knowledge, the nature of the information need and the user's choice of support tool. In Prinsloo et al. (2014), we presented a design study to show that user support through direct user guidance can provide solutions in the case of complex concordial relationships between nouns and pronouns. In terms of the Function Theory of Lexicography (Tarp (2008), Bothma and Tarp (2012), Fuertes-Olivera and Tarp (2014)), the design provides for text production, text reception and cognitive information

needs. In this article, we report on further work that has been done in this regard, viz. the development of a small-scale prototype to demonstrate the feasibility of such a tool. We describe it from the perspective of the end-user, i.e., how (s)he could go about solving his/her information need by using the prototype tool. We also briefly describe the technologies we used to develop the prototype tool and report on some observations from user-studies.

As will be clear from the discussion above and from Prinsloo et al. (2011, 2012), such techniques are made available "on demand", i.e., users are not forced to use them if they feel that their information needs have been solved by the "standard" dictionary article. In every case, the use of such a technique is therefore a conscious choice of the user to find more information or information that is easier to use, digest or apply than the information available in the dictionary, the outer text of the dictionary or other reference tools such as grammar books that the user may have available.

The importance of the user perspective as the main thrust in the compilation of modern dictionaries has been emphasized in numerous publications, e.g., Gouws and Prinsloo (2005), Tarp (2008, 2011, 2012), Fuertes-Olivera and Tarp (2014). The concept of *user-support* appropriately puts the user in focus. Compare Tarp's (2012: 253) idea of individualization when he refers to "quicker, more accurate and personalized satisfaction of the corresponding user needs". Our approach to user support furthermore does not necessarily put the user into a specific category (e.g., as a learner of the language): it is not profile-based and does not assume that the user will be interested to study a complete grammatical paradigm before being able to produce (or understand) texts. We therefore also cater for the casual, on-the-fly user, who is not interested or in a position to devote time to the in-depth learning of a foreign language, but relies on access to appropriate information from the e-dictionary and additional tools on demand.

3. Grammatical distinctions as a problem for African language lexicographers

Constructing phrases and sentences in African languages is a complicated process resulting from the classification of nouns into different noun classes. Text production, be it through dictionary consultation or creative writing, requires a substantial amount of grammatical knowledge. Traditionally the user had to rely on paper dictionaries and grammar books. Most print dictionaries for African languages are not helpful for text production and complicated grammatical issues are dealt with in many pages of fine print in grammar books which the user has to study as a prerequisite to text production.

3.1 The notion of grammatical distinctions

A given grammatical property may be expressed in many different forms. For

example, there are different equivalents for a pronoun such as *he*, determined by the grammatical class of the noun. Nouns in African (Bantu) languages are subdivided into different noun classes, as illustrated in Table 1. These classes have their own sets of subject concords and object concords, as well as different sets of pronouns such as demonstrative, possessive, emphatic and quantitative. This means that e.g. in Sepedi, an English personal pronoun such as *he* can be expressed by up to ten different subject concords, a form like *him* by ten object concords and more than 20 pronominal forms. Consider Table 1 which distinguishes 15 different noun classes, each having their own 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-			nna	
1st Person plural	rena 'we'	re	re			rena	
2nd Person sing.	wena 'you' (singular)	o	go			wena	
2nd Person plural	lena 'you' (plural)	le	le			lena	
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	gona	gohle
Class 16	fase 'below'			fa			

Class 17	godimo 'above'	go	go		ga	gona	gohle
Class 18	morago 'behind'			mo			

Table 1: The noun class system of Sepedi with a few sets of concords and pronouns

In Table 1 the Sepedi equivalent of the demonstrative 'this' varies depending on the class of the noun, e.g.,

- (1) class 1: *monna yo* 'this man', but
class 10: *dikgomo tše* 'these cattle' and
class 14: *bogobe bjo* 'this porridge'

Likewise, the Sepedi equivalent of the possessive 'of' differs for each class, e.g.,

- (2) class 1: *mosadi wa monna* 'wife of the man', but
class 2: *basadi ba monna* 'wives of the man' and
class 7: *selepe sa Madika* 'Madika's axe'

Concords and pronouns representing subjects and objects also vary according to the nominal class, e.g.:

- (3) a. O e bone 'He saw it'
o (e.g. *monna* class 1) e (e.g. *tau* class 9) bone
he (the man) it (the lion) saw
- b. o (e.g. *monna* class 1) bone yona (e.g. *tau* class 9)
he (the man) saw it (the (specific) lion)
- c. Ba rekile tšohle 'They bought everything'
ba (e.g. *banna* class 2) rekile tšohle (e.g. *dikgomo* class 10)
they (the men) bought all (the cattle)

In 3a *o* is a subject concord and *e* is an object concord. In 3b *yona* is an emphatic pronoun and in 3c *tšohle* is a quantitative pronoun.

3.2 Grammatical distinctions in the sentence context

If Table 1 is interpreted from a translation-based viewpoint (e.g. EN → Sepedi), the grammatical distinctions paradigm is mono-dimensional in the sense that it is always given for a single source language item which diverges into a single set of equivalents. More than one instance of grammatical distinction can, however, co-occur in a single construction or phrase; thus multiple choice points from the grammatical paradigms in Table 1 may (co-)occur in one sentence, and below we will illustrate cases with one, two and three occurrences. In (4)

the user has to determine the correct subject concord from the paradigm *o/a/le/se/e* to complete the sentence.

Example 1: *he/they*, as the subject of a sentence (subject concords):

- (4) a. **O/a/le/se/e** thušitše mosadi
 He helped the woman
- b. **ba/e/a/di** ja bogobe
 They eat porridge

In (5) the same situation prevails for the selection of the appropriate quantitative pronoun (used as an object of the verb).

Example 2: how to express *all* (quantitative pronouns):

- (5) Go bolaya **bohle/yohle/ohle/tšohle**
 To kill **all**

In (4) and (5) respectively the user has to deal with a single paradigm to complete the sentence. More complicated are situations where (s)he has to negotiate two, as in (6) or even more such grammatical paradigms in a single sentence, cf. (7).

In (6) the user has to find the correct subject concord and the applicable object concords from the *two* paradigms *o/a/le/se/e* and *ba/e/a/di* to complete the sentence: *he* as a subject and *them* as an object:

- (6) **O/a/le/se/e** tlo **ba/e/a/di** thuša.
 He will them help

The construction involves varying subjects and objects. The subject and object are in most cases not belonging to the same class as in (6) if *o* (class 1) is to represent the subject *he* and *ba* (class 2) the object *them*.

In (7) the correct demonstrative, subject concord and object concord need to be selected from the *three* paradigms *yo/wo/le/se/ye* (demonstratives), *o/a/le/se/e* (subject concords) and *ba/e/a/di* (object concords) to complete the sentence; and again involves varying subjects and objects in terms of the correct demonstrative, subject concord and object concord:

- (7) **Yo/wo/le/se/ye** **a/wo/le/se/e** **ba/e/a/di** thušitšego
 He he them helped
 He who helped *them*.

(7) is an example of a relative construction which can be regarded as one of the complicated structures for text production, especially for inexperienced users. It will first be described in more detail and then followed by an introduction of

the first prototype relative builder for Sepedi. The standard structure of the relative is noun + demonstrative + subject concord + verb stem with the relative suffix *-go*, as described below.

4. The relative construction in Sepedi

The relative is described in detail in traditional Sepedi grammars such as Van Wyk et al. (1992), Lombard et al. (1985), Ziervogel (1969) and Poulos and Louwrens (1994). They agree in principle that the relative modifies a noun or pronoun and that two main types are distinguished, i.e., direct and indirect relatives. Both direct and indirect relatives typically consist of nouns, demonstratives, subject concords, object concords, verb stems, relative suffixes and pronouns. For the user to produce a who-sentence, knowledge of at least 10 pages in fine print in Poulos and Louwrens (1994) is required, and the option to use Google Translate or Microsoft/Bing Translator for a translation does not exist.

(8) Direct relative

- a. Monna yo a sepelago. 'The man who is walking.'
- b. Monna yo a rekago puku. 'The man who buys/is buying the book.'
- c. Monna yo a rekelago bana puku. 'The man who buys a book for the children.'

In (8a) the relative consists of a noun of class 1 (*monna*), a demonstrative of class 1 (*yo*), a subject concord *a*, an intransitive verb stem *-sepela* and a suffix *-go* indicating relative mood on the verb form. In (8b) the verb stem *-reka* is transitive and is followed by a direct object *puku*. In (8c) the verb is double transitive, indicated by the suffix *-el*, and followed by an indirect object *bana* and a direct object *puku*. Objects can be pronominalized by means of object concords or pronouns.

So, e.g. the objects in (8c) can be pronominalized as in (9).

(9) Direct relative with a pronominalized object

- a. Monna yo a **ba** rekelago puku. 'The man who buys **them** a book.'
- b. Monna yo a rekelago **bona** puku. 'The man who buys **them** a book.'
- c. Monna yo a **e** rekelago bana. 'The man who buys **it** for the children.'
- d. Monna yo a rekelago bana **yona**. 'The man who buys **it** for the children.'

In (9a) and (9b) *bana* is pronominalized by its object concord *ba* and emphatic pronoun *bona* respectively. Likewise, in (9c) and (9d) *puku* is pronominalized by its object concord *e* and emphatic pronoun *yona* respectively. All of these constructions can occur in the present, future and past tense, in the positive or negative. Consider, e.g., (8a) in the three tenses present, future and past in the positive and negative in Table 2:

A	Direct Relative intransitive		
1	Present tense positive	Monna yo a sepelago	The man who is walking
2	Present tense negative	Monna yo a sa sepelego	The man who is not walking
3	Future tense positive	Monna yo a tlogo sepela	The man who will walk
4	Future tense negative	Monna yo a ka se sepelego	The man who will not walk
5	Past tense positive	Monna yo a sepetšego	The man who walked
6	Past tense negative	Monna yo a sa sepelago	The man who did not walk

Table 2: Present, past and future tense, positive and negative

All of the constructions in (8), (9) and Table 2 also apply for the indirect relative which differs from the direct relative in the use of an additional nominal before the verb as in (10):

(10) Object concord as pronominalized object

Monna yo mosadi a mo rekelago puku. 'The man for whom the woman buys a book.'

In this case the demonstrative belongs to *monna* but the subject concord to *mosadi*. A detailed discussion of the indirect relative is given in the traditional grammar books cited above.

Several hundreds if not thousands of possible relative constructions could be formed for relatives through the combination and permutation of the following possibilities:

- direct/indirect relative
- present tense/future tense/past tense
- positive/negative
- intransitive/transitive/double transitive
- object concord/object pronoun
- 18 noun classes, etc.

The 18 most typical types for the direct relative are the following:

1. Direct relative intransitive positive
2. Direct relative intransitive negative
3. Direct relative transitive positive with object noun
4. Direct relative transitive negative with object noun
5. Direct relative transitive positive with object noun pronominalized with a concord

6. Direct relative transitive negative with object noun pronominalized with a concord
7. Direct relative transitive positive with object noun pronominalized with a pronoun
8. Direct relative transitive negative with object noun pronominalized with a pronoun
9. Direct relative double transitive positive with indirect and direct object nouns
10. Direct relative double transitive negative with indirect and direct object nouns
11. Direct relative double transitive positive with indirect object pronominalized with a concord and direct object
12. Direct relative double transitive negative with indirect object pronominalized with a concord and direct object
13. Direct relative double transitive positive with indirect object pronominalized with a pronoun and direct object
14. Direct relative double transitive negative with indirect object pronominalized with a pronoun and direct object
15. Direct relative double transitive positive with indirect object and direct object pronominalized with a concord
16. Direct relative double transitive negative with indirect object and direct object pronominalized with a concord
17. Direct relative double transitive positive with indirect object and direct object pronominalized with a pronoun
18. Direct relative double transitive negative with indirect object and direct object pronominalized with a pronoun

Consider, e.g., the amount of knowledge presupposed from the dictionary user if (s)he wants to produce a fairly simple English single transitive sentence such as the *man who buys her a book* in Sepedi. (S)he has to know

- (a) the Sepedi word for *man*, i.e., *monna*,
- (b) to which of the possible 15 noun classes it belongs, i.e., class 1, in order to
- (c) select the correct demonstrative from 15 possibilities, i.e., class 1 *yo*,
- (d) the subject concord for class 1,
- (e) that an irregular relative concord is used for this noun class, i.e., *a* and not *o*,
- (f) the Sepedi word for *buy*, i.e., *reka*,
- (g) the Sepedi word for *book*, i.e., *puku*,

- (h) the Sepedi word to which *her* refers, e.g. *mosadi* 'woman', *mosetsana* 'girl', etc.,
- (i) to which of the noun classes the object belongs in order to select the correct object concord from 15 possibilities, i.e., *mo*,
- (j) what the relative suffix is, i.e., *go*
- (k) that the object concord is used pre-verbally.

in order to construct *monna yo a mo rekelago puku*.

5. Direct guidance for relative constructions

In principle, in respect of the relative, guidance can be given by means of three possible types of access depending on the user's need in terms of text production and his/her knowledge of the language.

The on-the-fly user will benefit most from assistance resembling machine translation for both text production and text reception purposes. The typical situation could be where the user simply wants to know how to say an English sentence such as *the man who bought the car* in Sepedi or how to translate the equivalent Sepedi sentence *monna yo a rekilego mmotoro* into English. As remarked by Prinsloo et al. (2014: 820), they simply need an on-the-spot solution and might not even be interested in learning Sepedi or English.

The focus of this article is on the user with limited knowledge of Sepedi who needs help — to a greater, or lesser extent — to create relative sentences in Sepedi. The user might even be someone like a second or third language speaker of Sepedi who is quite proficient in the language but requires confirmation as to the correctness of the sentences produced, e.g. in the case of relative constructions with irregular nouns and verbs. The relative builder, accessible via the article for "who" in the e-dictionary, attempts to cater for these different proficiency levels in a natural way by offering the user the opportunity to take shortcuts, e.g. if (s)he knows the Sepedi words or a longer route to the Sepedi words through dictionary lookup. No attempt was made to cater for formal user proficiency levels, e.g. users requested to indicate their level of expertise (e.g. Bothma 2011 and De Schryver 2003).

In building the relative construction, the system performs the following steps for the sentence *the children who love the food*:

- (i) *children*: the tool provides the correct equivalent from the e-dictionary, i.e., *bana* tagged for part of speech as N02 (noun of class 2, cf. Table 1);
- (ii) *who*: keeping the agreement constraint from the sentence formation rule (noun + demonstrative + subject concord + verb + relative suffix (-*go*), cf. (7) and (8)), the tool extracts the demonstrative for class 2 from the closed-class list of demonstratives, i.e., *ba*;

- (iii) (subject concord): The insertion of the SC is coded in the rule for relatives: it requires, in addition to the demonstrative in (ii), the subject concord for the noun in (i). As in (ii), the tool proposes *ba*, i.e., the subject concord for class 2.
- (iv) *food*: the tool provides the correct equivalent from the dictionary, i.e., *dijo* tagged for part of speech as N08 (noun of class 8, cf. Table 1);
- (v) *love*: as for (i), the system selects the correct Sepedi equivalent: *rata*, plus adding the relative suffix. The adding of *-go* is built into the relative construction rule.

Result: Bana ba ba ratago dijo. *The children who love the food.*

The processes (i) to (v) are the same for automated text production support. The user enters the entire English phrase, e.g. *the man who bought the car* and the system applies (i) to (v) to construct the Sepedi sentence *monna yo a rekilego mmotoro*. It is also possible to type the full Sepedi sentence, in which case the process is reversed to produce the English translation *the man who bought the car*.

In these specific cases, no user knowledge (neither lexical nor grammatical knowledge) is required, and the process is fully automated. However, the tool can also be used interactively, which requires the user to make specific choices in the construction of the relative, as discussed in section 7 below.

6. A software implementation: underlying technology and essential components of the relative builder

Essential grammatical components are a machine readable English/Sepedi dictionary with part of speech markup. The syntax and components of the relative construction are hard-coded. The functioning prototype was developed using AngularJS and Bootstrap on the front-end. The back-end was developed in PHP that uses a SQLite database. The current prototype is hosted at www.sepedihelper.co.za. The application is written using best practices to clearly differentiate between logic/content and interface/display, as well as between input and output, to allow maximum flexibility. These characteristics allow for easier improvement, maintenance and extension of the application. Due to the differentiation between front-end and back-end, both can be replaced with other technologies, if necessary. Integrating this application with Microsoft Word or with the Open-Source alternatives like Libre-Office and Open-Office will require a rewrite of the software for each. Integration with e-dictionaries, on the other hand, should be less complex. The writing tool would need minimal improvements and refactoring to allow it to be used as a component inside e-dictionaries.

For the prototype, a limited subset of the data of an existing e-dictionary was used, copying only the relevant data fields of the e-dictionary to a new e-dictionary database. We therefore foresee that a new e-dictionary database will not be required for a full implementation of this tool — the tool will simply access the database of the dictionary. In a full version, a standard bilingual Sepedi/English e-dictionary will be used. The e-dictionary database of this standard dictionary will, however, have to be modified to make provision for the fields that are required by the proposed tool.

The main form of input data required are tagged wordlists. Nominal word lists require the noun itself, a translation equivalent paradigm as well as a noun class indication, e.g. "*badiredi* (employees, workers), N02" is a full database entry represented in the database in three fields. Verbs require more information, e.g. the verb itself, its translation equivalents, tense, transitivity and lastly, if past tense, the verb entry indicating the present tense form of the verb. The latter is required when converting to and from certain rules and from present to past tense. So, e.g., the database entry for the present tense verb, *reka*, is as follows: "*reka* (buy, buying, buys), present, transitive". The past tense entry for *reka* i.e., *bought* would be indicated in the database as "*rekile* (bought), past, transitive, *reka*". The *reka* at the end is given to enable the transformation rule applicable to verbs.

Lastly, it is worth mentioning that great emphasis was placed on performance. The SQLite database duplicates the data in a special full-text-search (FTS) table that is extremely fast to query. Users start typing and the entire wordlist is searched for a partial match. This would take unacceptably long to complete with normal database "LIKE" operators. A negative constraint for using SQLite and FTS tables is that database performance degrades with writing operations (changes, inserts and deletions) due to locking tables during updates. Wordlists are, however, not constantly modified, so updating the wordlists during maintenance periods is an acceptable trade-off and constraint for a cost-free and performant database. Wordlists that remain static as part of the grammar rules of Sepedi (demonstratives, subject concords, etc.) are hardcoded in PHP. Such lists require manual updating but allow performance gains that are well worth the trade-off. The development choices described above should allow the application to scale easily to accommodate much larger wordlists or a full e-dictionary (modified to contain the required database fields).

7. Using the relative builder

A prototype of the relative sentence builder is available at www.sepedihelper.co.za and is briefly described in the following section.

The user will be able to access the tool directly from the web address or in future from an e-dictionary or word processing software, as illustrated in Figure 1:

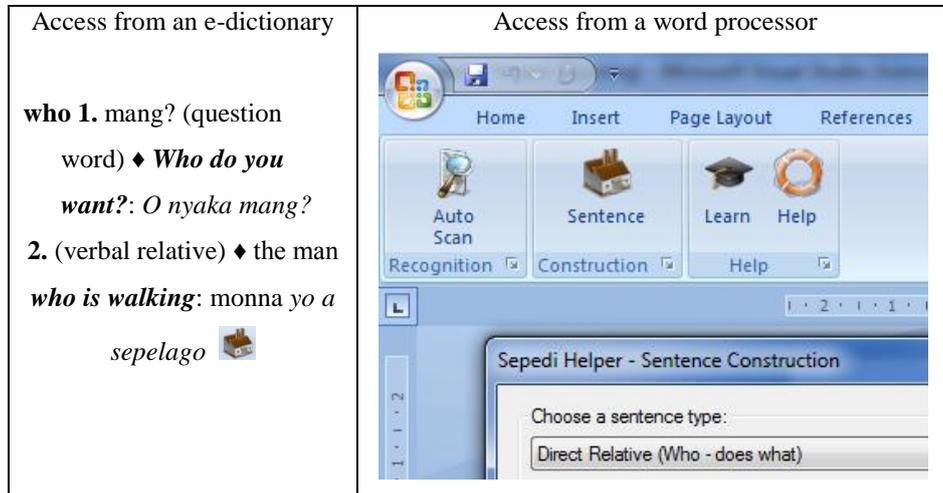


Figure 1: Accessing the relative construction tool

In the dictionary, an additional icon (picture of a factory) allows the user to launch the tool, which opens the tool in a pop-up window or in the user's word processor. When the user is already working in the word processor, an icon on the toolbar allows him/her to launch the tool. The Relative Builder can therefore equally well be used as a dictionary component or as a writing assistant. The dictionary user will be offered guidance from within the dictionary article of all English and Sepedi lemmas which are relevant to the relative construction as in Figure 1, left column, e.g. *who, what, which* and all of the Sepedi demonstratives *yo, ba, wo*, etc. as well as the relative suffix *-go*. The user who requires assistance to build a relative construction from within a word processor can click on the factory icon in the taskbar as in Figure 1, right column.

The tool currently offers assistance for all 18 types of relatives listed in section 3.2 above.

The user's existing knowledge at any given point is taken into account by offering them choices e.g. to enter Sepedi words directly or to go via English. In the build-up process given below, the user would like to express in Sepedi the sentence *the children who love/like it*, where *it* refers to "food". The user departs from the dictionary article for *who* by clicking on the factory icon in the article for *who* in the e-dictionary (or consults the helper by clicking on the factory icon if the helper is accessed from a word processor). In both cases (s)he is presented with the *Sepedihelper* screen in Figure 2.

Sepedi:
Prototype sentence builder [? Instructions](#)

Sentence Type: **Relative** [?](#)

WHO/WHAT (IS DOING SOMETHING)

Sentence:
[Choose Subject Noun] [Choose Verb]

Tense: **Present** [?](#) Type: **Positive** [?](#)

Input List:

- Tense - Present
- Type - Positive
- Subject Noun - [Not Chosen]
- Verb - [Not Chosen]
- Direct Object Noun - [Not Chosen]
- Indirect Object Noun - [Not Chosen]

Figure 2: The Sepedi relative builder main screen

The user is informed that cognitive information at different levels can be obtained by clicking on the question mark "?" icons. The inexperienced or first-time user can obtain cognitive information on direct relatives and direct relatives with an object concord. Typical examples are given and a few suggestions for building relative constructions are also presented, cf. Figure 3.

Relative mood:

Relative expresses: **who/what is doing something**

Present

Most common structure: Noun + demonstrative + subject concord + verb + go + object

Example: **Positive:** Monna yo a rekago puku.
 'The man who buys a book'

Negative: Monna yo a sa rekego puku. 'The man who does not buy a book'

Negation strategy: sa ... verb ends in -e

Can occur with direct and indirect object:

Monna yo a rekelago bana puku
 indirect object direct object
 'The man who buys for the children a book.'

Nouns can be replaced by **pronouns**:

Example (**bana**): Monna yo a rekelago bona puku

Object nouns can also be replaced by object concords:

Example: Monna yo a ba rekelago puku

Future

Most common structure: Noun + demonstrative + subject concord + fut. + go + verb + object

Example: **Positive:** Monna yo a tlogo reka puku.
 'The man who will buy a book'

Negative: Monna yo a ka se rekego puku.
 'The man who will not buy a book'

Negation strategy: ka se ... verb ends in -e

The combination ke = I + ka se becomes **nka se**
Nka se reke puku 'I shall not buy the book'

Past

Most common structure: Noun + demonstrative + subject concord + verb + go + object

Example: **Positive:** Monna yo a rekilego puku.
 'The man who bought a book'

Negative: Monna yo a sa rekago puku.
 'The man who did not buy a book'

Negation strategy: sa + verb back to the present tense

Figure 3: Help screen explaining the basics of relative sentence construction

The gradual build-up and eventual completed sentence will be displayed in a horizontal line under the heading **Sentence**, as in Figure 9. Initially this line reflects all the required and optional elements of the relative construction, e.g. [Choose Subject Noun] and [Choose Verb] to be replaced step-by-step with real words and concords in the build-up process.

For step 1 the instruction is to choose a noun. If the user knows the Sepedi word (s)he can type it in, as in Figure 4.

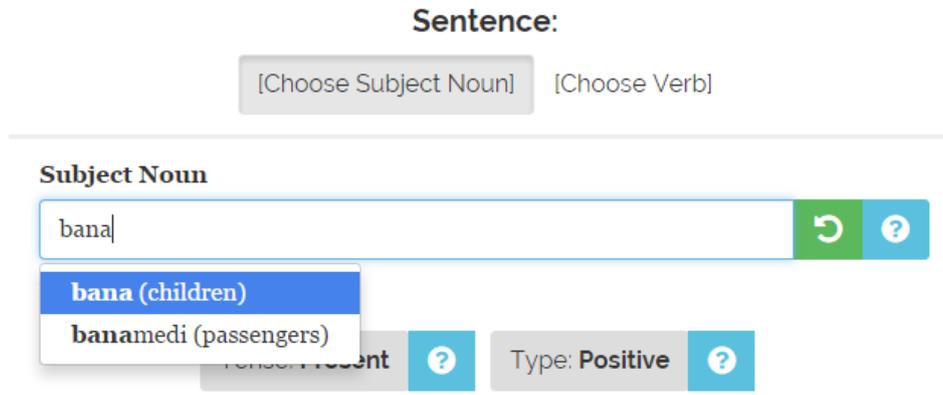


Figure 4: Inserting the Sepedi subject noun in Step 1 of the relative builder

The system presents the word with its translation equivalent(s) (currently from a limited database). In a full implementation a more comprehensive set of entries with direct translations will be provided. Words as well as navigation links that would open up viewing it as a dictionary entry for more comprehensive help with a word will be offered. The translation equivalents help the user to ascertain that (s)he is dealing with the right Sepedi noun. If the user does not know the Sepedi word, (s)he simply types the English word to find the Sepedi translation and selects *bana*, the required Sepedi item, as in Figure 5:

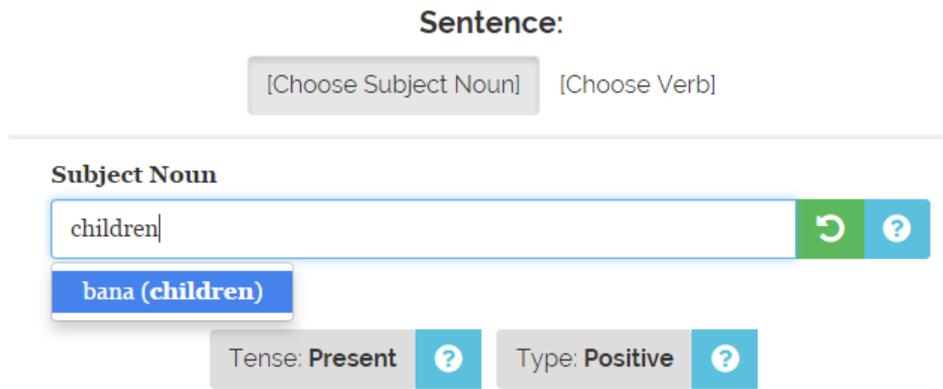


Figure 5: Inserting the Sepedi subject noun by typing the English noun in Step 1 of the relative builder

The system automatically performs a lookup for the part of speech of *bana*, i.e., a noun from Class 2, and generates the demonstrative *ba* and the subject concord *ba* in the next two fields, i.e., the generated demonstrative and the generated subject concord as well as displaying the current stage in the build-up process in the Generated Complete Sentence line, i.e., *bana ba ba* prompting the user to enter the remaining required component, i.e., the verb, Figure 6.

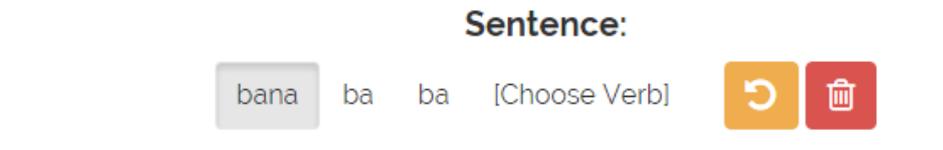


Figure 6: Generating the demonstrative and subject concord in the Sepedi relative builder

The user repeats the same process as in Step 1 to enter the verb *rata* in Step 2. The system automatically adds the required relative suffix *-go* to the verb and prompts the user to add a direct object, Figure 7.



Figure 7: The Sepedi relative phrase for *the children who love*

The user enters the object noun *dijo* "food", the builder completes the relative and asks the user if (s)he would rather prefer to pronominalize the object noun, i.e., the children who love *it*. At this point the system has already determined the part of speech of *dijo*, i.e., a noun in class 8. A choice is offered between insertion of the object concord *di* or substitution of *dijo* by its pronoun *tšona*.



Figure 8: The Sepedi relative phrase for *the children who love the food*

If the user chooses the object concord, the system inserts it in the correct syntactic order which is pre-verbal.



Figure 9: The Sepedi relative phrase for *the children who love it*

Consider the same situation as in Figures 8 and 9 for a noun from a different nominal class, as in Figures 10 and 11.



Figure 10: The Sepedi relative phrase for *the children who love the employee*

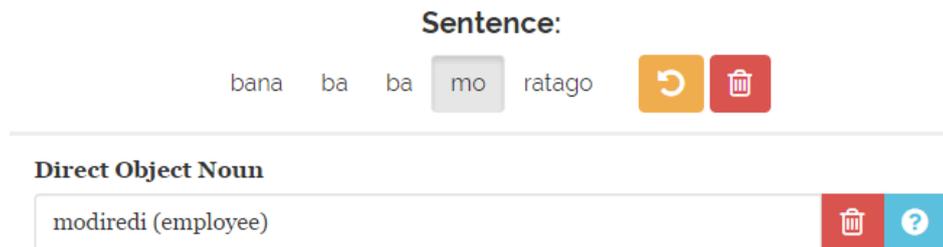


Figure 11: The Sepedi relative phrase for *the children who love him*

In Figure 10, the noun is *modiredi* (employee) which belongs to class 1 and the object concord of class 1 is *mo*. Thus the system correctly generated the object concords for *dijo* and *modiredi* as *di* versus *mo* respectively.

8. User-studies

Reflecting on user-studies is not the aim of this article — user feedback will be dealt with in more detail in a forthcoming publication. Recently, two user studies totaling 109 users were conducted on the SepediHelper and the relative construction in particular. These studies, of which some were performed in class by students on their cellphones, indicate that most users fail to produce "who"-phrases correctly without external help and benefitted much from the guidance by the writing assistant. In these studies the user had to attempt compilation of the phrase first and then to use the writing assistant. The value of the assistant was clear. The most typical problems in respect of producing a "who"-sentence are summarized as follows:

- Wrong mood
- Wrong sentence position for the subject of the sentence
- Both demonstrative and/or subject concord left out
- Wrong negation morpheme or negation morpheme in wrong syntactic position
- Did not change the concord of class 1 to *a*
- Omission of the relative suffix
- Incorrect spelling / word division
- Got it right — no guidance but confirmation from the tool

Users generally regarded the tool as user-friendly and easy to use.

9. Conclusion and future work

User support through direct guidance (and other support mechanisms) for complex grammatical structures allows the user to navigate via the shortest route to the information (s)he is looking for in an e-dictionary without having to work through long and often complicated grammar-type representations of complex grammatical structures. Such guidance is always available on demand, i.e., the user is not forced to work through any such support mechanisms if (s)he finds that the "standard" data in the e-dictionary are sufficient to solve his/her information need in a given situation. However, if more information is needed or if the standard presentation of the information (be this in the e-dictionary, in outer texts or in reference tools) is too difficult or complex to be easily understood, the user would have an alternative mechanism (or alternative mechanisms), accessible on demand from within the e-dictionary (or from within a word processor or as a stand-alone application) to obtain the relevant information.

The proposed direct guidance functions also successfully combat information overload and fulfil the needs of not only the learner of the language but also of the casual on-the-fly-user of the dictionary; its flexibility is intended to provide a step towards individualization. Different access points are available to the user depending on his/her pre-existing knowledge. It is therefore not a profile-based tool.

We envisage that such mechanisms be implemented as "plug-in modules" in entries of specific lemmas of an e-dictionary, i.e., an additional link/button is shown to the user on screen which (s)he can follow on demand. Such tools could therefore be used as writing tools integrated in a word processor, again activated by the user on demand, if (s)he requires to check the correct formulation of a complex grammatical construction, i.e., checking whether his/her own original construction is grammatically correct. Such additional functionality could be part of our future work — this is different from constructing a sentence in a word processor (or dictionary-linked tool), and rather similar to spelling and grammar checkers that currently occur in popular word processing software.

Future work includes the full-scale implementation of user support for complex structures proposed in this paper as a module within e-dictionaries. Identifying and categorising additional support techniques and developing prototypes and the full-scale implementation of such additional support techniques are also envisaged, as well as identifying further complex grammatical structures for which additional user support techniques may need to be developed. We will also investigate the possibility of the reuse of all such modules for user support in word processors and writing tools, as well as for language instruction and computer-assisted language learning (CALL).

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, BMBF, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (Grant specific unique reference numbers 85763 and 95925). The Grantholders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the author, and that the NRF accepts no liability whatsoever in this regard.

References

- Alonso Ramos, M., M. Garcia Salido and O. Vincze.** 2014. Towards a Collocation Writing Assistant for Learners of Spanish. Faaß, G. and J. Ruppenhofer (Eds.). 2014. *Workshop Proceedings of the 12th Edition of the KONVENS Conference, Hildesheim, Germany, October 8–10, 2014*: 77-88. Hildesheim: Universitätsverlag Hildesheim.
- Bertels, A. and S. Verlinde.** 2011. La lexicographie et l'analyse de corpus: nouvelles perspectives. *Meta: Journal des Traducteurs* 56(2): 247-265.
- Bosch, S.E. and G. Faaß.** 2014. Towards an Integrated e-Dictionary Application — The Case of an English to Zulu Dictionary of Possessives. Abel, Andrea, Chiara Vettori and Natascia Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*: 739-747. Bolzano/Bozen: EURAC Research.
- Bothma, T.J.D.** 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 71-102.
- Bothma, T.J.D., U. Heid and D.J. Prinsloo.** 2013. Implementing Decision Support for Text Production in e-Dictionaries. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper*. Institute of the Estonian Language and Trojina, Institute for Applied Slovene Studies, 17–19 October 2013, Tallinn, Estonia.
- Bothma, T.J.D. and S. Tarp.** 2012. Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.
- Buyse, K. and S. Verlinde.** 2013. Possible Effects of Free On Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of Linguee and the Interactive Language Toolbox. *Procedia: Social and Behavioral Sciences* 95: 507-512.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.).** 2011. *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Lexicographica. Series Maior 146. Berlin/Boston: De Gruyter Mouton.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

- Lombard, D.P., E.B. van Wyk and P.C. Mokgokong.** 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: Van Schaik.
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.
- Prinsloo, D.J. and S.E. Bosch.** 2012. Kinship Terminology in English–Zulu/Northern Sotho Dictionaries — A Challenge for the Bantu Lexicographer. Fjeld, Ruth Vatvedt and Julie Matilde Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012*: 296-303. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Prinsloo, D.J., T.J.D. Bothma and U. Heid.** 2014. User Support in e-Dictionaries for Complex Grammatical Structures in the Bantu Languages. Abel, Andrea, Chiara Vettori and Natascia Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*: 819-827. Bolzano/Bozen: EURAC Research.
- Prinsloo, D.J., U. Heid, T.J.D. Bothma and G. Faaß.** 2011. Interactive, Dynamic Electronic Dictionaries for Text Production. Kosem, I. and K. Kosem. 2011. *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011, Bled, 10–12 November 2011*: 215-220. Ljubljana: Trojina, Institute for Applied Slovene Studies. http://www.trojina.si/elex2011/elex2011_proceedings.pdf.
- Prinsloo, D.J., U. Heid, T.J.D. Bothma and G. Faaß.** 2012. Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22: 290-320.
- Prinsloo, D.J., Daniel Prinsloo and J.V. Prinsloo.** 2015. A Writing Tool for Sepedi. *E-Lex 2015*. Herstonceux Castle, United Kingdom, 11–13 August 2015.
- Taljard, E. and D.J. Prinsloo.** 2013. *Lexicographic Treatment of So-called Cattle Colour Terms in Northern Sotho*. Paper delivered at AFRILEX 2013 — 18th Annual International Conference of the African Association for Lexicography, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa, 2–5 July 2013.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. *Lexicographica Series Maior* 134. Tübingen: Niemeyer.
- Tarp, S.** 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 54-70.
- Tarp, S.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28: 253-267.
- Van Wyk, E.B., P.S. Groenewald, D.J. Prinsloo, J.H.M. Kock and E. Taljard.** 1992. *Northern Sotho for First-Years*. Pretoria: Van Schaik.
- Verlinde, S.** 2011. Modelling Interactive Reading, Translation and Writing Assistants. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 275-286.
- Verlinde, S., P. Leroyer, and J. Binon.** 2010. Search and You Will Find. From Stand-alone Lexicographic Tools to User Driven Task and Problem-oriented Multifunctional Leximats. *International Journal of Lexicography* 23(1): 1-17.
- Wanner, L., S. Verlinde and M. Alonso Ramos.** 2013. Writing Assistants and Automatic Lexical Error Correction: Word Combinatorics. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia*: 472-487. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.
- Ziervogel, D.** 1969. *Handboek van Noord-Sotho*. Pretoria: Van Schaik.

Afrikataalleksikografie: Gister, vandag en môre

D.J. Prinsloo, *Departement Afrikatale, Universiteit van Pretoria, Pretoria,
Suid-Afrika (danie.prinsloo@up.ac.za)*

en

Elsabé Taljard, *Departement Afrikatale, Universiteit van Pretoria, Pretoria,
Suid-Afrika (elsabe.taljard@up.ac.za)*

Opsomming: Hierdie artikel het ten doel om 'n perspektief te bied op Afrikataalleksikografie se reaksie op ontwikkelings in die internasionale leksikografiepraktyk met betrekking tot nuwe tendense en veranderinge wat hoofsaaklik deur die elektroniese era teweeg gebring word. Die ontwikkeling van die Afrikataalleksikografie sal met dié van hoofsaaklik Engelse papier- en elektroniese woordeboeke vergelyk word. Drie tydperke van leksikografiese ontwikkeling word onderskei. Daar word geargumenteer dat die algehele prentjie vir die ontwikkeling van Afrikataalleksikografie minder gunstig is as dié van Engels en tale soos Duits en Frans. Die fokus is op die beskikbaarheid van woordeboeke, leksikografiese gehalte, die gebruik van nuwe beskikbare rekenaar-tegnologieë en die gebruik van inligtingsbronne van die inligtingsera. Daar word tot die gevolgtrekking gekom dat die leksikografiepraktyk in die Afrikatale in die meeste gevalle nie tred hou met internasionale ontwikkelings in die leksikografie nie en dat die aantal uitdagings vir dié tale toeneem het. Ten slotte word 'n aantal stukragte geïdentifiseer wat ten goede kan meewerk om die kwaliteit van Afrikataalwoordeboeke te verbeter.

Sleutelwoorde: AFRIKATAALLEKSIKOGRAFIE, AFRIKATAALWOORDEBOEKE, AAN-
LYN WOORDEBOEKE, GEBRUIKERSPERSPEKTIEF, INLIGTINGSERA, INTROSPEKSIE, KORPUS-
GEBASEERDE LEKSIKOGRAFIE, LEMMASELEKSIE, WOORDEBOEKKULTUUR

Abstract: African Language Lexicography: Yesterday, Today and Tomorrow.

This article aims to give a perspective on African language lexicography's response to developments in international lexicographic practice in relation to new trends and changes mainly brought about by the electronic era. The development of African language lexicography will be compared to that of mainly English paper and electronic dictionaries. Three periods of lexicographic development are distinguished. It will be argued that the overall picture for the development of African language lexicography compares less favourably to that of English and languages such as German and French. The focus will be on the availability of dictionaries, lexicographic quality, use of new computer technologies and the use of information sources made available by the information era. It will be concluded that in most instances African language lexicography is not keeping up with international developments in lexicography and that the number of challenges for these languages has increased. Finally a number of positive factors will be identified which could contribute to the improvement of the quality of dictionaries for African languages.

Keywords: AFRICAN LANGUAGE DICTIONARIES, AFRICAN LANGUAGE LEXICOG-

RAPHY, CORPUS-BASED LEXICOGRAPHY, DICTIONARY CULTURE, INFORMATION ERA, INTROSPECTION, LEMMA SELECTION, ONLINE DICTIONARIES, USER PERSPECTIVE

1. Inleiding

Die doel van die artikel is om vas te stel tot watter mate leksikografie in die Afrikatale vanaf die pre-elektroniese korpusera in die leksikografie tot op hede tred gehou het met die jongste internasionale tendense in die ontwikkeling van die leksikografie. Tweedens word die huidige en toekomstige uitdagings wat die Afrikataalleksikografie in die gesig staar, geïdentifiseer. In die derde plek word stukragte wat Afrikataalleksikografie bevorder, maar ook negatiewe aspekte wat leksikografiese ontwikkeling strem, aangestip. Ten slotte word 'n aantal aanbevelings gemaak wat kan meehelp tot die verbetering van Afrika-taalwoordeboeke en dat Afrikataalleksikografie nie nog verder uit pas raak met internasionale tendense nie.

Die ondersoek is geïnspireer deur die baanbrekerswerk van Reinhard Hartmann getiteld *Lexicography in Africa*, (Hartmann 1990: 90) wat die Afrikatale in sy wydste interpretasie as "tale wat in Afrika gepraat word" as vertrekpunt neem. Hierdie artikel fokus in beginsel op wat internasionaal as Bantoetale bekend staan, 'n deelversameling van die breër term *Afrikatale* wat in werklikheid in vier taalfamilies verdeel kan word (Heine en Nurse 2000: 1). Aangesien die term Bantoetale in Suid-Afrika gestigmatiseer geraak het, word na hierdie tale wat in Suid-Afrika gepraat word ook as *Afrikatale* verwys.

Vir die doel van hierdie artikel is die vroeë negentigerjare as skeidslyn tussen "gister" en "vandag" in die Afrikataalleksikografie gekies, en wel om die volgende redes:

- Hierdie tydperk word allerweë as die begin van die elektroniese korpusrewolusie beskou — Hanks (2012: 398) praat van "the corpus revolution in lexicography";
- dit verteenwoordig die beginpunt van die elektroniese leksikografie waar digitale media as platform vir woordeboekdata gebruik word (Rundell 2012: 72); en
- 1990 is die jaartal waarop 'n insiggewende verslag oor leksikografie in Afrika verskyn het (Hartmann 1990).

As maatstaf vir beoordeling van die mees basiese leksikografiese vereistes wat in hierdie artikel as kriteria voorgehou word kan woordeboeke vir Afrikatale met enige woordeboek vir byvoorbeeld Engels, Frans, of Duits vergelyk word. Die sogenaamde "Big Five" woordeboeke, te wete *Cambridge Advanced Learner's Dictionary* (CALD), *Collins COBUILD Advanced Dictionary of English* (COBUILD), *Longman Dictionary of Contemporary English* (LDOCE), *Macmillan English Dictionary for Advanced Learners* (MED) en *Oxford Advanced Learner's Dictionary of Current English* (OALD) word internasionaal erken as woordeboeke van beson-

der hoë leksikografiese gehalte (De Schryver en Prinsloo 2000: 292) en word daarom as basis vir vergelyking in hierdie studie gebruik. Daar sou geargumenteer kon word dat 'n vergelyking van Afrikataalwoordeboeke met woordeboeke wat uit 'n ryke leksikografiese tradisie stam enigsins ongebalanseerd is. Ideaal gesproke moet ooreenstemmende woordeboektypes met mekaar vergelyk word — aanleerderswoordeboeke met aanleerderswoordeboeke, omvattende verklarende woordeboeke met omvattende verklarende woordeboeke, ensovoorts. Indien die gebrek aan Afrikataalwoordeboeke egter in ag geneem word, is dit duidelik dat so 'n ideale vergelyking gewoon nie moontlik is nie. Die geskatte aantal Afrikatale is tans 2 035 (Heine en Nurse 2000: 1), en volgens Awak (1990: 13) is daar slegs ongeveer 1 300 leksikons wat naastenby as Afrikataalwoordeboeke beskou kan word. Wanneer hierdie getal vergelyk word met die 6 700 woordeboeke wat volgens Awak (op cit.) reeds 30 jaar gelede vir Engels bestaan het, is dit duidelik dat so 'n een-tot-een vergelyking nie uitvoerbaar is nie. In die lig hiervan, is die visier van hierdie artikel juis grof gestel om 'n globale oorsig oor die stand van die Afrikataalleksikografie te gee.

2. Basiese kriteria vir die beoordeling van leksikografiese kwaliteit

In die lig van die voorafgaande paragraaf behoort dit reeds duidelik te wees dat formulering van die kriteria waarvolgens leksikografiese kwaliteit beoordeel word, problematies kan wees. Bestaande woordeboeke het verskillende teikengebruikers, verskillende funksies, hulle verskil ten opsigte van omvang en is daarom op beide makro- en mikrovlak uiteenlopend van aard. Vir die doel van hierdie artikel moes die navorsers gevolglik noodwendig terugval op die mees basiese doel van 'n woordeboek as kriterium vir die evaluering van Afrikataalwoordeboeke as goeie of swak woordeboeke.

Macmillan Dictionaries (*From Corpus to Dictionary*) gee die basiese vereiste vir 'n woordeboek kort en kragtig:

A dictionary is a description of the vocabulary of a language. It explains what words mean, and shows how they work together to form sentences.

Hoewel treffend, is hierdie oorkoepelende vereiste relatief vaag en is verdere verfyning nodig. Wat die mikrostruktuur betref, word goeie woordeboeke gekenmerk deur die aanbod van 'n verskeidenheid inligtingstipes:

As containers of knowledge, dictionaries should accommodate a variety of information types. (Gouws 1990: 52)

Gouws (1989: 113) benadruk veral die belangrikheid van betekenisinligting in 'n woordeboekartikel:

Vir die deursneewoordeboekgebruiker is betekenis die inligtingstipe wat die algemeenste in woordeboeke nageslaan word. As 'n mens na die struktuur van 'n woordeboekartikel kyk, is dit ook duidelik dat betekenisbeskrywing nie net die

omvangrykste komponent van die artikel is nie, maar dat dit ook as die sentrale deel van 'n woordeboekartikel beskou moet word.

Gouws en Prinsloo (2005: 144) sê dat 'n woord so gedefinieer moet word dat die gebruiker al die antwoorde verkry op die vraag wat hom of haar genoep het om die woord in die woordeboek na te slaan, en verwys na Laufer (1992: 71) wat presies uitspel wat dit behels om 'n woord "te ken":

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

Laufer (op cit.) bevestig dus die noodsaak van 'n verskeidenheid inligtingstipes binne 'n gegewe woordeboekartikel.

Gouws en Prinsloo (2005: 39) dui verder aan dat 'n woordeboek wat maksimaal in die gebruiker se behoefte voorsien ook 'n eienskap van 'n goeie woordeboek is. Die mate waarin 'n woordeboek dus in die behoefte van sy teiken-gebruiker voorsien, is dus 'n belangrike kriterium vir leksikografiese kwaliteit:

Good dictionaries do not only display a linguistically sound treatment of a specific selection of lexical items. Good dictionaries are products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are, cf. Haas (1962), Barnhart (1962), Otto (1989) and Prinsloo and Gouws (1996). ... The user-perspective, so prevalent in modern-day metalexigraphy, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality.

Die voorafgaande verwysings handel oor die basiese vereistes op mikrostruktuurvlak, dit wil sê inligtingstipes en die kwaliteit van bewerking, maar 'n tweede belangrike aspek vir die woordeboekgebruiker lê op die makrostruktuurvlak, naamlik of die woord (lemma) waaroor inligting verlang word wel in die woordeboek opgeneem is.

A good dictionary is one in which you can find the information you are looking for — preferably in the very first place you look. (Haas 1962: 48)

... it is the function of a popular dictionary to answer the questions that the user of the dictionary asks, and dictionaries on the commercial market will be successful in proportion to the extent to which they answer these questions of the buyer. (Barnhart 1962: 161)

De Schryver en Prinsloo (2001) gee 'n uitvoerige bespreking van die tekortkominge in Afrikataalwoordeboeke op makrostrukturele vlak as gevolg van

ontoereikende lemma-aanbod. Afrikataalwoordeboeke haal ook nie die paal op mikrostrukturele vlak nie as gevolg van ontoereikende bewerking van die lemmas soos hieronder uitvoeriger aangetoon sal word.

Gegewe die algemene gebrek aan woordeboeke vir die Afrikatale soos hierbo uiteengesit, en die feit dat geen gesofistikeerde analise en vergelykingstrategie tussen tipes woordeboeke moontlik is nie, word daar in die hieropvolgende bespreking volstaan met die evaluering van woordeboeke op grond van die basiese makro- en mikrostrukturele eienskappe oor die drie tydperke wat in hierdie artikel voorgelê word. Ons gebruik dus die term *leksikografiese kwaliteit* vir woordeboeke volgens ons subjektiewe oordeel oor die bewerking van lemmas, gebruikersleiding deur produktiewe en reseptiewe gebruik, lemma-aanbod, lemmaseleksie en die korrektheid van die inligting.

Ter illustrasie kan die woordeboekinskrywing vir *cheap* in OALD en MED in Figure 1 en 2 as 'n voorbeeld van goeie leksikografiese gehalte beskou word. Die "rykheid en kompleksiteit" as eienskappe van goeie woordeboeke waarna Atkins (1998: 3) verwys, kan met die eerste oogopslag raakgesien word in die aanbod van verskeie betekenisonderskeidings, verskillende tipes gebruiksvoorbeelde, idiomatiese uitdrukkings, afgeleide vorme, uitspraakleiding, woordsoort-aanduiding, nommering en selfs 'n voorafopsomming van betekenisonderskeiding, ensovoorts.

cheap /tʃi:p/ *adj* (-er, -est) **1(a)** low in price; costing (relatively) little money: *cheap tickets/fares* ◦ *the cheap seats in a theatre* ◦ *Cauliflowers are very cheap at the moment.* **(b)** worth more than the cost; offering good value: *These wool carpets are cheap at £10.99 a metre.* ◦ *\$3 is very cheap for a hard-cover book.* ◦ *immigrant workers, used as a source of cheap labour* (ie labour that is paid very little, esp unfairly). **2** charging (CHARGE² 1) low prices: *a cheap hairdresser/restaurant.* **3** of poor quality; inferior: *cheap furniture/jewellery/shoes* ◦ *a cheap and nasty bottle of wine.* **4** (of people, words or actions) unpleasant and mean²(2); not clever or worthy of respect: *a cheap gibe/joke/remark/retort* ◦ *That was a cheap trick to play on her.* ◦ *He's just a cheap crook.* ◦ *His treatment of her made her feel cheap.* **5** (US *infml derog*) extremely careful with one's money; mean²(1). **IDM** **cheap at the 'price;** (US) **cheap for the 'price** so well worth having that the price, however high it is, does not seem too much: *The holiday will be very expensive but if it helps to make you fit and healthy again it will be cheap at the price.* **on the 'cheap** (*infml*) without paying or asking the usual, or a fair, price: *buy/sell/get sth on the cheap.* **► cheap** *adv* (*infml*) for a low price: *get sth cheap* ◦ *sell sth off cheap.* **IDM** **go 'cheap** (*infml*) to be offered for sale at a low price: *The local shop has some radios going cheap.* **not come 'cheap** to cost a lot of money: *High quality electronic equipment doesn't come cheap.* **cheaply** *adv* **1** for a low price: *buy/sell/get sth cheaply.* **2** in a cheap(1a,3) manner: *The room was cheaply furnished.* **IDM** **get off lightly/cheaply** ⇨ LIGHTLY.

Figuur 1: Die artikel vir *cheap* in OALD

cheap¹ /tʃi:p/ adj ★★★

1 not expensive	4 without value
2 not of good quality	5 against spending money
3 unkind/unfair	+ PHRASES

1 not expensive: *People should have access to cheap, fresh food.* ♦ *The local buses are cheap and reliable.* ♦ *a cheap and plentiful source of energy* ♦ *The machine is relatively cheap to build but expensive to maintain.* — opposite DEAR, EXPENSIVE **1a.** used about something that is lower in price than usual or than you expected: *Silver is very cheap in Mexico.* **1b. cheap and cheerful** Br E not expensive and of reasonable quality: *It's not the best place you'll ever eat at, but it's cheap and cheerful.*

2 not expensive and not of good quality: *People were drinking cheap wine from plastic cups.* ♦ *He was always buying her cheap jewellery.* ♦ **a cheap imitation** *Always ask for our products by name – beware of cheap imitations!*

2a. cheap and nasty Br E unattractive, not expensive and of very bad quality: *a market stall selling cheap and nasty watches*

3 a cheap action or remark is unfair or unkind and does not deserve respect: *I'm not interested in scoring cheap points in this debate.* ♦ **a cheap joke/jibe** *The Minister should be dealing with the problem, not making cheap jibes.* ♦ **a cheap trick** *He said he realized it had been a cheap trick and he was sorry.*

4 not considered important or valuable: *It happened during the war when life was cheap.*

5 mainly Am E not willing to spend money: *She's so cheap she wouldn't even buy her own mother a birthday card.*

cheap at half the price Br E informal used for emphasizing that something is good value and not considered expensive

cheap at the price Br E used for emphasizing that something is worth more than it costs: *He'd spent £60,000 but said his victory was cheap at the price.*

on the cheap informal **1** if you buy something on the cheap you do not pay the usual price: *It's a great opportunity to pick up a car on the cheap.* **2** if you do something on the cheap you spend less money and get a product or service that is not of the best quality — **cheaply** adv, **cheapness** noun [U]

→ TALK²

cheap² /tʃi:p/ adv at a low price: *I can't believe I managed to get it so cheap.* ♦ **dirt cheap** informal (=at a very low price) *I got this bike dirt cheap at a car boot sale.* **a. going cheap** informal being sold for a lower price than is usual: *They've got a load of furniture going cheap.* **b. not come cheap** to cost or charge a lot: *He doesn't come cheap, but he gets the job done on time.*

Figuur 2: Die artikel vir *cheap* in MED

3. Die leksikografiese nalatenskap van die pre-1990-era

Die leksikografiese hede kan beswaarlik van die leksikografiese verlede losgemaak word. Baie van die leksikografiese probleme en tekortkominge waarmee huidige Afrikataalleksikograwe worstel, is die nalatenskap van leksikografiese tradisies en praktyke uit die pre-1990-era. Hierdie aangeleenthede word vervolgens onder die loep geneem.

3.1 Woordeboeksamestelling in die prekorpusera¹

Op enkele uitsonderings na is die tydperk voor die 1990's gekenmerk deur die samestelling van papierwoordeboeke. Woordboeksamestelling het op drie strategieë berus: introspeksie, observasie en uitbreiding van lemmalyste van bestaande woordeboeke. Introspeksie maak staat op die leksikograaf se kennis, maar soos tereg op Macmillan Dictionaries se webblad (*From Corpus to Dictionary*) uitgewys word, was (en is) sodanige inligting nie voldoende nie. Gevolglik het leksikograwe ook gebruik gemaak van die handmatige bestudering van potensiële bronne van leksikografiese inligting. In die laaste instansie was (en is) die uitbreiding en herbewerking van die lemmalyste van bestaande woordeboeke aanvaarde leksikografiese praktyk. 'n Goeie voorbeeld van laasgenoemde is die *Pukuntšu Woordeboek* (1989) van Kriel en Van Wyk, waarin dit duidelik in die voorwoord gestel word dat dié woordeboek 'n 'omgewerkte' uitgawe van T.J. Kriel se *Pukuntšu* (1983) is.

Hierdie werkswyse ten spyte, kan wat Engels betref sonder twyfel gesê word dat vroeëre weergawes van die sogenaamde "Big Five" reeds van goeie leksikografiese gehalte was. Dit is te verstane omdat die samestellers van hierdie woordeboeke die voordeel van onder andere 'n lang en goed ontwikkelde leksikografiese tradisie gehad het — Samuel Johnson het immers reeds in 1755 sy *A Dictionary of the English Language* gepubliseer.

Vir die Afrikatale was die situasie veel minder gunstig. Daar was byvoorbeeld 'n tekort aan omvattende, eentalige verklarende woordeboeke. Wanneer die nege amptelike Afrikatale van Suid-Afrika beskou word, blyk dit dat slegs een van hierdie tale voor 1990 oor 'n eentalige, verklarende woordeboek beskik het — dit is die *Thanodi ya Setswana ya Dikole* (Tswana skoolwoordeboek) deur M. Kgasa, gepubliseer in 1976. Dit is egter nie 'n omvattende verklarende woordeboek nie, maar soos die titel ook aandui, is dit gemik op skoolleerders. Eentalige woordeboeke wat voor 1990 selfs vir Afrikatale wat as dominante tale beskou word, saamgestel is, is skaars en eerder die uitsondering as die reël.² Vir Swahili, een van die Afrikatale met die grootste aantal sprekers (98.3 miljoen), is daar in 1935 die eentalige *Kamusi ya Kiswahili, yaani kitabu cha maneno ya Kiswahili* deur F. Johnson gepubliseer. Geen inligting met betrekking tot die bestaan van sodanige woordeboeke kon vir tale soos Kinyarwanda (12.1 miljoen sprekers), Lingala (10 miljoen sprekers) en Rundi (9 miljoen sprekers) gevind word nie. Die eerste eentalige verklarende woordeboeke in Shona (meer as 8 miljoen sprekers) het eers in 1996 (*Duramazwi ReChiShona*) en 2001 (*Duramazwi Guru ReChiShona*) verskyn. Bestaande woordeboeke was meestal twee- of drietalige woordeboeke met beperkte bewerking soos die blote aangee van vertaal-ekwivalente, en sommige woordeboeke was dikwels nie veel meer as woorde-lyste nie. Sulke woordeboeke reflekteer swak leksikografiese kwaliteit omdat daar nie byvoorbeeld enige betekenisinligting gegee word nie. Die hoofstukke van verskeie outeurs in Hartmann (1990) soos Gouws (1990: 55), Mann (1990: 7) en Awak (1990: 2) bevestig dat die leksikografie van die Afrikatale in dié tyd-

perk in 'n ontwikkelingsfase was. Gouws (1990: 55) stel dit soos volg:

[L]exicographical activities on the various indigenous African languages [... have] resulted in a wide range of dictionaries. Unfortunately, the majority of these dictionaries are the products of limited efforts not reflecting a high standard of lexicographic achievement. ... [and] display a lack of comprehensive monolingual lexicographical description. Important aspects of linguistic information like semantics, syntax and pragmatics are neglected because the scope of translation dictionaries only allows a restricted treatment of these categories.

Gouws (1990: 53) wys twee tipes ongelyke dekking van informasie in Afrika-taalwoordeboeke en selfs ongelyke aanbod in dieselfde woordeboek uit.

On the one hand, it can be shown that dictionaries for the different language groups do not exhibit a similar degree of lexicographical sophistication and variety. On the other hand, the dictionaries of one specific language group do not treat all information categories alike, and within one dictionary an unbalanced presentation of information categories can often be identified.

Busane (1990: 27) stel dit soos volg:

It is thus clear that the bulk of dictionaries of major national or regional languages referred to above still leave the African user in the lurch, because of the nature of their presentation and the arrangement of the entries.

3.2 Eurosentriese karakter

'n Verdere kenmerk van die vroeë Afrikataalleksikografie wat dikwels as kritiek geopper word, is dat dit Eurosentries was; deur Europeërs vir hulle eie gebruik saamgestel en nie déúr moedertaalsprekers vir moedertaalsprekers nie.

The history of lexicography in Africa began as a result of European activities: exploration, evangelization and colonialization. The early lexicons, whether compiled by explorers, missionaries or colonial administrators, were 'Euro-centered', produced in Europe for Europeans rather than for African users. ... Even with the emergence of modern linguistics, lexicographic works have been primarily intended for scholarly interest and not for the needs of ordinary Africans. (Awak 1990: 17)

Dit is ironies dat Europese woordeboeke as van hoogstaande gehalte beskou word maar dat Afrikataalwoordeboeke se Eurosentriese karakter as een van die redes vir swak Afrikataalwoordeboeke voorgehou word. Die antwoord is eenvoudig — woordeboeke vir tale soos Engels, Frans en Duits is deur leksikograwe saamgestel terwyl Afrikataalwoordeboeke deur sendelinge wie se fokus bloot kommunikasie op die sendingveld met Afrikataalsprekers was, saamgestel is. Hulle was dus volgens Awak (1990: 17) nie *per se* daarop ingestel om goeie woordeboeke te maak nie, veral nie vir Afrikataalsprekers as teiken-gebruikers nie.

3.3 Hantering van grammatikale kompleksiteit en problematiese benadering tot lemmatisering

Reeds in die preteoretiese era het die pioniers geworstel met die kompleksiteit van die grammatikale sisteme van die Afrikatale. Verskeie benaderings tot lemmatisering, 'n uiters problematiese aspek in Afrikatale, is gevolg, waaronder die paradigmabenedering (Ziervogel en Mokgokong in *Groot Noord-Sotho woordeboek* (GNSW)) en 'n reëlgerigte benadering (Kriel en Van Wyk in *Pukuntšu woordeboek, Noord-Sotho-Afrikaans*) (PUKU).

Die stam- versus woordtradisies vir lemmatisering was sekerlik die grootste bron van frustrasie vir die gebruiker van Afrikataalwoordeboeke. Die stambenedering vereis kennis van die grammatika ten einde 'n komplekse woord te kan opsoek. Dit is veral problematies in tale met 'n konjunktiewe skryfwyse soos Zulu waar die gebruiker gekonfronteer word met lang ortografiese eenhede, bestaande uit 'n woordstam en verskeie affikse. Gebruikers moet oor 'n redelike mate van grammatikakennis beskik alvorens hulle in staat is om die stam te kan identifiseer ten einde dit in 'n woordeboek op te soek. So is dit byvoorbeeld problematies vir veral die aanleerders van Zulu om die stam *-sebenza* in *ngokusebenzisa* ("deur te gebruik/aktiveer") te identifiseer ten einde dit in 'n Zuluwoordeboek te kan opsoek. Hierteenoor staan tale soos Noord-Sotho en Tswana met 'n disjunktiewe skryfwyse waar woordlemmatisering die aangewese weg is, maar ten spyte hiervan het sommige leksikograwe die stamtradisie gevolg, waarskynlik omdat dit vanuit 'n linguïstiese oogpunt meer korrek sou wees. Volgens hierdie tradisie word stamme as lemmas opgeneem. So sal die Noord-Sothowoord *mosadi* ("vrou") onder "S" as *-sadi* gelemmatiseer word in plaas van onder "M" as *mosadi*. Van Wyk (1995: 89) lewer besondere fel kritiek op stamlemmatisering:

Would the user of an English dictionary, for example, take kindly to the lemmatization of *perception*, *conception*, *reception* and *deception* as *-ception* just because they are lexically related? Or to *pre-empt* and *empty* being lemmatized as *-empt* (as ZM would do)?

Leksikograwe soos Ziervogel en Mokgokong wyk selfs van die standaard alfabetiese volgorde af ten einde meer linguïsties korrek te wees — lemmas wat byvoorbeeld met *tl-*, *ts-*, *tš-* en *tsh-* begin, word nie onder die letter *T* gelemmatiseer nie, maar vorm aparte alfabetiese kategorieë. Hierdie praktyk werk verdere gebruikersonvriendelikheid in die hand en lok nog meer kritiek uit. Die hedendaagse klem op die gebruikersperspektief waar die gebruiker en nie die linguis/leksikograaf nie, die kwaliteit van 'n woordeboek bepaal, reflekteer nog slegter op hierdie woordeboeke.

3.4 Gebrek aan 'n woordeboekkultuur

Afrikataalleksikografie in die pre-1990-era is verder gekortwiek deur die

gebrek aan 'n sterk woordeboekkultuur (Atkins 1998: 3) — 'n situasie wat nog weinig verander het soos bevestig deur 'n gebruikerstudie deur Taljard et al. (2011: 103). Min woordeboeke van werklik goeie leksikografiese gehalte is beskikbaar wat gevolglik nie 'n stewige woordeboekkultuur in die hand werk nie. Atkins (1998: 3) beoordeel die situasie soos volg:

[T]he speakers of African languages have not in their formative years had access to dictionaries of the richness and complexity of those currently available for European languages. They have not had the chance to internalize the structure and objectives of a good dictionary, monolingual, bilingual or trilingual.

Met verwysing na Afrikataalwoordeboeke, lê Gouws (1990: 55) ook 'n duidelike verband tussen die kwaliteit van woordeboeke en woordeboekkultuur:

With a few exceptions, these dictionaries offer only restricted translation equivalents, aimed at decoding, with no or little attention given to the encoding function of a pedagogical dictionary. ... there is less information to be exploited by the user. This results in a vicious circle, with the dictionary user not realizing what he can expect to find in his dictionary or how to interpret the given entries because the lexicographer does not include all the possible information categories or treat them on an equal basis.

Om dit baie eenvoudig te stel: dit kan beswaarlik van sprekers van 'n Afrika-taal verwag word om 'n gesofistikeerde woordeboekkultuur te hê as daar selfs nie eers 'n enkele woordeboek vir dié betrokke taal bestaan nie.

In 'n taalgemeenskap waarin daar 'n goed gevestigde woordeboekkultuur is, is die woordeboekgebruiker vertrouwd met die verskillende woordeboeksoorte, "die tipes data wat in woordeboeke aangetref word, die waarde en beperkinge van verskillende woordeboeksoorte en met naslaanvaardighede om woordeboeke suksesvol te kan gebruik" (Gouws 2016: 112). In so 'n gemeenskap is daar dus 'n wisselwerking tussen leksikografie en die woordeboekgebruiker — hoe meer gesofistikeerd die woordeboekgebruiker, hoe meer gesofistikeerd sal die woordeboeke wees ten einde die gebruiker se behoeftes ten beste te bevredig (Gouws 2016: 112). Woordeboekgebruik in die Afrikatale — wat maar een aspek van 'n woordeboekkultuur verteenwoordig — steek sleg af by woordeboekgebruik vir Engels, Frans, Duits, ens. (Taljard et al. 2011: 103 en Atkins 1998: 3). Hierdie feit word ook duidelik geïllustreer deur Nkomo (2012: 21, 45 en 47) wat telkens verwys na die afwesigheid van 'n volledig ontwikkelde woordeboekkultuur in Zimbabwe. Wanneer daar 'n sterk woordeboekkultuur in 'n gemeenskap is, sal daar kundige woordeboekgebruikers wees. Die kwaliteit van woordeboeke is dus uiters belangrik, maar so ook die kundigheid van die woordeboekgebruiker:

There are two direct routes to more effective dictionary use: the first is to radically improve the dictionary: the second is to radically improve the users. (Atkins en Varantola 1998: 83)

Teen 1990 vertoon Afrikataalleksikografie reeds 'n groot agterstand in vergelyking met woordeboeke vir Engels, Duits en Frans, maar tog met tekens van ontluikende potensiaal vir die samestelling van goeie woordeboeke deur moedertaalsprekers self. Hartmann (1990: v) verwys na "lexicography in Africa, with all its exciting possibilities and external constraints".

4. Nuwe uitdagings

Benewens uitdagings wat die gevolg van bepaalde historiese leksikografiese tradisies en praktyke is, word die Afrikataalleksikografie voortdurend deur nuwe uitdagings in die gesig gestaar. Die samestelling en gebruik van woordeboeke vir Afrikatale word naamlik gekortwiek deur wat Hartmann (1990: v) eksterne beperkings ("external constraints") noem. Die Afrikatale word algemeen as hulpbronbeperkte tale beskryf (Prinsloo 2012: 121, Jones et al. 2005: 143,156 en 157). Tale kan as hulpbronbeperk beskou word op grond van die volgende:

- gebrek aan fondse vir die samestelling en aankoop van woordeboeke,
- gebrek aan teks- en spraakkorpusse,
- afwesigheid van 'n gestandaardiseerde ortografie,
- beperkte aantal sprekers,
- taal en/of sprekers wat aan politieke onderdrukking onderhewig is,
- taaldata wat nie bewaar word nie, en
- taalliggame wat verantwoordelik is vir standaardisering en normering van die taal ontbreek of is wanfunksioneel (Wolvaardt 2017: 9, Prinsloo 2009: 162).

Genoemde beperkinge is almal tot 'n mindere of meerdere mate relevant vir die Afrikataalleksikografie. In hierdie artikel word gefokus op finansiële beperkings en die implikasies daarvan vir leksikografiese produkte. Hier is drie aspekte ter sprake: eerstens die finansiële beperkings waaronder uitgewers funksioneer, tweedens, met verwysing na die Suid-Afrikaanse situasie, die onderbefondsing van die Nasionale Woordeboekeenhede, en laastens die bekostigbaarheid van woordeboeke vir die potensiele woordeboekgebruikers.

Gebrek aan fondse vir die samestelling van woordeboeke vir Afrikatale is 'n groot probleem. Uitgewers beperk noodgedwonge die omvang en prys van papierwoordeboeke (Prinsloo 2009: 162). Dit beteken dat die leksikograaf beperk word tot opname van relatief min lemmas met 'n meer uitvoerige bewerking, byvoorbeeld *Oxford Bilingual School Dictionary, Northern Sotho and English* (ONSD) (ongeveer 10 000 lemmas), of relatief meer lemmas met slegs enkele vertaalekwivalente, soos *Popular Northern Sotho Dictionary* (POP) (ongeveer 28 000 lemmas). Prinsloo (2009: 162) verwys hierna as 'n driehoekige wis-

selwerking tussen prys, aantal lemmas en uitvoerigheid van bewerking:

Bidirectional dictionaries bridging English with a Bantu language in South Africa are currently caught up in a triangulation of number of lemmas versus exhaustiveness of treatment versus price. This simply means that 500–600 pages are the default limit within which the compiler can operate as prescribed by the publishers.

ONSD	POP
<p>cheap * <i>adjective</i> (cheaper, cheapest) ■ [PC +] theko ya fase • I can't buy an expensive watch, so I will buy this cheap one. <i>Nka se reke sešupanako sa go bitša mašelang a mantši, ke tla reka se sa theko ya fase.</i></p> <p>cheat <i>verb</i> (cheats, cheating, cheated) ■ fora ◊ cheat someone ■ tšhela phori mahlong</p> <p>cheating <i>noun</i> (no plural) ■ bohvirihwiri • The men were starting to suspect that it was he who was doing the cheating. <i>Banna ba be ba thomile go mo naganela gore ke yena a dirilego bohvirihwiri bjoo.</i></p> <p>check ** <i>verb</i> (checks, checking, checked) ■ hlola ◊ check carefully ■ lekodišiša ◊ check (for/on) ■ hlolola</p> <p>cheek <i>noun</i> (no plural) ■ kgang; moreba (<i>impudence; nerve</i>)</p> <p>cheekiness <i>noun</i> (no plural) ■ makoko</p> <p>cheeky <i>adjective</i> (cheekier, cheekiest) ■ [PC +] kgang; [PC +] makoko • A cheeky person is annoying. <i>Motho wa kgang o a tena.</i></p> <p>cheer <i>verb</i> (cheers, cheering, cheered) 1 ■ thakgala • A few members of the audience cheered when he said so. <i>Batheetši ba sego bakae ba thakgetše ge a realo.</i> 2 ■ thakgalela • The crowd cheered the players. <i>Lešaba le ile la thakgalela baraloki.</i> ◊ cheer up ■ thabiša • We tried to cheer her up, because she was feeling sad. <i>Re lekile go mo thabiša ka ge a be a nyamile.</i></p> <p>cheerful <i>adjective</i> (more cheerful, most cheerful) ■ [DEM + SC +] thakgetšego; [DEM + SC +] thabilego • He is always friendly and cheerful. <i>O phela e le yo botho e bile e le yo a thakgetšego ka mehla.</i></p> <p>cheese <i>noun</i> (pl. cheeses) ■ tšhese • I put some cheese on my toast. <i>Ke beile tšhese borothong bja ka bja go bešwa.</i></p>	<p>cheap tšhipiša, tšhipilē, fokōtša thēkō</p> <p>cheat n., mothētši, sehvirihwiri; v., fora, hlalefētša, hwirihwitša</p> <p>check v., tšhēka; n., sekōtšhō, tšhēkō</p> <p>cheek lerama, thama; - bone rapōthama</p> <p>cheekiness kgang, mereba</p> <p>cheer n., thaba, ditlatse; v., thabiša, kgothatša</p> <p>cheerful thabilēng, thabilēgo</p> <p>cheese kase, tšhese</p>

Figuur 3: Aantal lemmas versus omvang van die bewerking in ONSD en POP: *cheap* tot *cheese*

Uit Figuur 3 blyk dat die bewerking van lemmas in POP beperk is tot die blote aanbod van een of meer vertaalekwivalente, terwyl ONSD heelwat meer inlig-

tingstipes aan die gebruiker bied soos gebruiksfrekwensie, woordsoortkategorie, naamwoordklasnommer, enkelvoud of meervoud en hulp met enkelvoudsvormen en meervoudsvorme. Bloot op sigwaarde geneem, wys hierdie eenvoudige vergelyking ook dat ongeveer agt artikels in POP in dieselfde woordeboekspasie aangebied word as drie in ONSD. Dit is dus duidelik dat die finansiële omstandighede van uitgewers as opdraggewers vir die saamstel van woordeboeke noodwendig 'n direkte invloed op die kwaliteit van die eindproduk het.

Woordeboeke word nie slegs in opdrag van uitgewers saamgestel nie — in enkele Afrikalande bestaan formele strukture wat hul onder andere met leksikografie bemoei. Vergelyk byvoorbeeld die *Institute of Kiswahili Research* (TUKI) (Universiteit van Dar-es-Salaam, Tanzanië), die Allexprojek (later: *African Languages Research Institute* of ALRI) in Zimbabwe, en die nasionale leksikografie-eenhede in Suid-Afrika.

Wat laasgenoemde strukture betref, is dit alombekend dat hierdie eenhede deur gebrekkige befondsing van die Pan Suid-Afrikaanse Taalraad (PanSAT) ernstig aan bande gelê word:

How have the flag bearers for South Africa's bold approach to restoring the nation's indigenous languages, become the neglected poor relations of the deeply flawed institution that is the Pan South African Language Board (PanSALB)? How has the national lexicography project, pioneered in the early years of South Africa's democratic transition by some of the country's greatest language activists and academics, been permitted to degenerate into the scattered efforts of a diminishing band of lexicographers? Forced for the last decade into perpetual begging for adequate funding, the National Lexicography Units (NLUs) hover on the verge of extinction. The critical question is, 'does anyone care?'. Dishearteningly, indications from government seem to imply that the response is, 'Not really.' (Wolvaardt 2017: 9)

Gebrekkige befondsing het noodwendig 'n negatiewe uitwerking op die effektiewe funksionering van hierdie eenhede. Die begroting wat aan die eenhede toegesê word, maak byvoorbeeld nie voorsiening vir die uitgee van (papier)woordeboeke nie, en kommersiële uitgewers is nie altyd gretig om die woordeboeke wat deur die eenhede saamgestel word, uit te gee nie — waarskynlik as gevolg van die relatief klein omset van woordeboeke. Dit het tot gevolg dat potensiële woordeboekgebruikers nie toegang tot woordeboeke het nie, wat die ontwikkeling van 'n woordeboekkultuur onder Afrikataalsprekers verder kortwiek. Laastens, hoewel toegang tot woordeboeke nie noodwendig die aankoop en besit daarvan impliseer nie, sou dit tog ideaal wees as 'n beduidende persentasie gebruikers wel woordeboeke aankoop. Volgens StatsSA se 2014/15-opname is die gemiddelde huishoudelike inkomste van swart Suid-Afrikaners R92 893 per jaar, dit is R7 741 per maand. Dit is onwaarskynlik dat die aankoop van 'n woordeboek of woordeboeke hoog op die prioriteitslys van hierdie huishoudings sal wees.

5. Die era van rekenaartegnologie

Die vroeë negentigs is gekenmerk deur die ontluiking van rekenaartegnologie en die gevolglike inligtingsontploffing van die internet. Hierdie ontwikkelings het 'n reuse-impak op die leksikografie — dit lei naamlik die era van die korpusrewolusie en van elektroniese en aanlyn woordeboeke in.

Rekenaartegnologie bring ongekende prosesseringspoed en feitlik onbeperkte stoorruimte vir woordeboekdata op CD-ROM, DVD en hardeskywe mee. Skielik is die eeueoue ruimteprobleem iets van die verlede — die rekenaar verstom met nuwe tegnologie, kapasiteit en spoed waarmee inligting gestoor en opgeroep kan word:

The arrival of the modern computer set in motion a series of lexicographers' dreams without equal in the history of dictionary making. Achieving the wildest of those electronic dictionary vistas has the potential to result in reference works beyond all recognition. (De Schryver 2003: 143)

De Schryver (2003) voorspel verder die samestelling van "adaptive and intelligent dictionaries" wat in staat sal wees om hulle gebruikers "te verstaan" en aan 'n gebruiker inligting te verskaf op sy spesifieke kundigheidsvlak.

5.1 Die korpusera vir leksikografie

Vir die leksikografie is die belangrikste die aanbreek van die korpusera met elektroniese korpora wat deur die rekenaar moontlik gemaak word. In 1987 bring *Collins Cobuild* inderdaad 'n algeheel nuwe dimensie met die verskyning van 'n korpusgebaseerde woordeboek wat woordeboeksamestelling hoofsaaklik op grond van introspeksie eensklaps in die skadu stel. COBUILD neem die gebruiksfrekwensie van woorde in ag en identifiseer vyf frekwensiekategorieë wat in die woordeboek skematies aangedui word deur middel van gevulde en ongepulde diamante. Die samestellers bestudeer ook trefwoorde in konteks, onder meer met die oog op betekenisonderskeidings, kollokasies en gebruiksvoorbeelde. Introspeksie en korpusdata vorm 'n voortreflike kombinasie. Soos Wójtowicz (2016: 12) tereg opmerk: "The implementation of the most evolved instruments in the preparation of a lemma-sign list should be accompanied by the use of basic common sense".

Die korpusera het egter nie die verwagte omwenteling in Afrikataalwoordeboeke teweeg gebring nie. Lemmaseleksie weerspieël steeds opvallende leemtes wat betref woorde wat in 'n bepaalde woordeboek opgeneem behoort te word en woorde wat liefsv uitgelaat moet word ten gunste van ander woorde wat onder andere op grond van gebruiksfrekwensie na alle waarskynlikheid deur die teikengebruiker opgesoek sal word.

Vergelyk die woordeboekartikels vir die resente woordeboeke: *The Official Setswana-English Dictionary* (OSED), *The Official Tshivenda-English Dictionary* (OTED) en *The Official English-Isindebele-Afrikaans Dictionary* (OENAD) in figure 4, 5 en 6 onderskeidelik:

<p>babble ¹ (<i>noun</i>) balabala; ² (<i>verb</i>) balabala babe (<i>noun</i>) lesea babies (<i>noun</i>) masea baboon (<i>noun</i>) tshwene baby (<i>noun</i>) lesea baby disease characterised by the bumping head of a child which is believed to be very dangerous if not treated earlier (<i>noun</i>) tlhogwana bachelor (<i>noun</i>) kgope bachelors (<i>noun</i>) dikgope back (<i>noun</i>) mokokotlo; mokotla; mokwatla backbite (<i>verb</i>) tshoma backs (<i>noun</i>) dikhularo; mekwatla bacon (<i>noun</i>) sepeke bacteria (<i>noun</i>) baketeria; baketheria bad (<i>verb</i>) senyega; sulafala bad beer (<i>noun</i>) phiri bad look (<i>noun</i>) medilolo bad luck (<i>noun</i>) madimabe; sefifi bad omen (<i>noun</i>) kung</p>	<p>bar (<i>noun</i>) bara bare fruits (<i>verb</i>) ungwa baritone (<i>noun</i>) kodu bark (<i>noun</i>) lekwmamati bark of trees (<i>noun</i>) makwati barn (<i>noun</i>) polokelo barn owl (<i>noun</i>) kgaribišane barns (<i>noun</i>) dipolokelo barrel (<i>noun</i>) faki barren woman (<i>noun</i>) moopa barricades (<i>noun</i>) mephakarego barrier (<i>noun</i>) mopako; thibedi barrister (<i>noun</i>) mmueledi barter (<i>verb</i>) papatso based on (<i>verb</i>) itshetega based upon ¹ (<i>noun</i>) theilweng; ² (<i>verb</i>) theilwe bashed (<i>verb</i>) ketekwa basic noun (<i>noun</i>) inakemo basin (<i>noun</i>) mogopo; sekotlele basis (<i>noun</i>) motheo bask (<i>verb</i>) ikomosa; ora</p>
--	--

Figuur 4: Die eerste gedeelte van die alfabetiese strek B in OSED

<p>avocado <i>n</i> hafukhada, gadaphele avoid <i>v</i> tutshela, phipha, ilisa avow <i>v</i> tenda, dibula, diamba, ana awaken <i>v</i> vusa, karusa, fhaṭusa awakening <i>n</i> phaṭhuso, phaṭuso award¹ <i>n</i> tshifhiwa, mpho award² <i>v</i> ŋea, fha, hwedza awe <i>n</i> ganuko, nyofho, khanganeo awful <i>adj</i> ofhisaho, vhiṭhaho, shushaho awl <i>n</i> unzhi, nyunzhi axe <i>n</i> mbaḍo * mbaḍo i a rema an axe cuts axis <i>n</i> asi</p> <p style="text-align: center;">..... B </p> <p>baboon <i>n</i> pṭene baby <i>n</i> lushie, lutshetshe, ŋwana baby-grand <i>n</i> piamalufhafha, ṭhukhu babyhood <i>n</i> vhuṭuku, vuhana bachelor <i>n</i> khombe * <i>munna uyu ndi khombe</i> this man is a bachelor back <i>n</i> murahu; muṭana; tsenzhe back- saw <i>n</i> tsahaswali backbite <i>v</i> sevha, remba backbone <i>n</i> muṭoḍo background <i>n</i> valafulu back-stitch <i>n</i> muṭhavhela bacon <i>n</i> ŋama ya nguluvhe bacteriology <i>n</i> qivhazwisinisa bacterium <i>n</i> tshisinisa bad <i>adj</i> vhiṭhaho, u sa luga badly <i>adv</i> luvhi bag <i>n</i> mukhwama, saga bagpipe <i>n</i> tshiilavuvha bagworm <i>n</i> tshireḍakhuni</p>	<p>bamboo <i>n</i> musengele, luṭanga banana <i>n</i> muomva band¹ <i>n</i> bennde, tshigwada band² <i>v</i> vhoṭha, pomba bandage <i>n</i> banditshi bandit <i>n</i> livemu, tshigevhenga bangle <i>n</i> bengele banish <i>v</i> thatha * <i>vho mu thatha kha lino</i> he has been banished from this country banishment <i>n</i> muthatho bank <i>n</i> bannga bank-note <i>n</i> tshedele ya mabambiri bankruptcy <i>n</i> khundwalandu banner <i>n</i> fulaha banter <i>v</i> swaswa baobab <i>n</i> muvhuyu baptise <i>v</i> lovhedza, golovhedza baptism <i>n</i> ndovhedzo baptist <i>n</i> mulovhedzi bar <i>n</i> bara; dungo; mukonde (wa tshisibe) barbarism <i>n</i> vhuṭuka barbarity <i>n</i> vhuṭema barbarousness <i>n</i> vhonani barbarity barber <i>n</i> mugeri bare <i>adj</i> vha hu si na tshithu bargain <i>n</i> dzhiawolala barge <i>n</i> khumbi bark¹ <i>n</i> gwati bark² <i>v</i> huvha barley <i>n</i> goroi barn <i>n</i> goha barrack <i>n</i> nṇḍu barrel <i>n</i> fagi, lwobo; mulomo barren <i>adj</i> phanzhe; nyumba, muumba; bale</p>
---	---

Figuur 5: Die eerste gedeelte van die alfabetiese strek B in OTED

<p>babble¹ <i>n gebabbel</i>; ukukhuluma ngendlela engazwakaliko babble² <i>v babbel</i>; -khuluma ngendlela engazwakaliko Babel <i>n Babel</i>; abantu abanengi bakhuluma ngesikhathi sinye baboon <i>n bobbejaan</i>; ifene baby <i>n baba(tjie)</i>; umntwana bachelor <i>n vrygesel</i>; ipohlo back¹ <i>n rug</i>; umhlana back² <i>adv terug</i>; -seka backbone <i>n ruggraat</i>; ithambo lomgogodlha background <i>n agtergrond</i>; isendlalelo backing <i>n steun</i>; isekelo backwards <i>adv agteruit</i>; -emuva bacon <i>n (vark)spek</i>; isipeke senyama yefarigi bacteria <i>n bakterieë</i>; imilwana bacteriology <i>n bakteriologie</i>; isifundo sebhaktheriya</p>	<p>bacterium <i>n bakterie</i>; umulwana bad <i>adj sleg</i>; -mbi badge <i>n kenteken</i>; ibheji badly <i>adv erg</i>; kumbi bag <i>n sak</i>; umgodla baggage <i>n bagasie</i>; umthwalo bail¹ <i>n borg</i>; ibheyili bail² <i>v borg staan</i>; -bheyila bailiff <i>n balju</i>; isithunywa sekhotho bait <i>n aas</i>; ibheyithi bake <i>v bak</i>; -bhaga (amakhekhe) baker <i>n bakker</i>; umbhagi bakery <i>n bakkery</i>; indawo la kubhagwa khona uburotho balaclava <i>n balaklawamus</i>; ikhophayede balance¹ <i>n balans</i>; isilinganiso balance² <i>v balanseer</i>; -linganisa balanced <i>adj gebalanseerde</i>; -linganisako balcony <i>n balkon</i>; amathuri bald <i>adj kaalkop</i>; -nepandla balderdash <i>n nonsens</i>; ikulumo engazwakaliko</p>
--	---

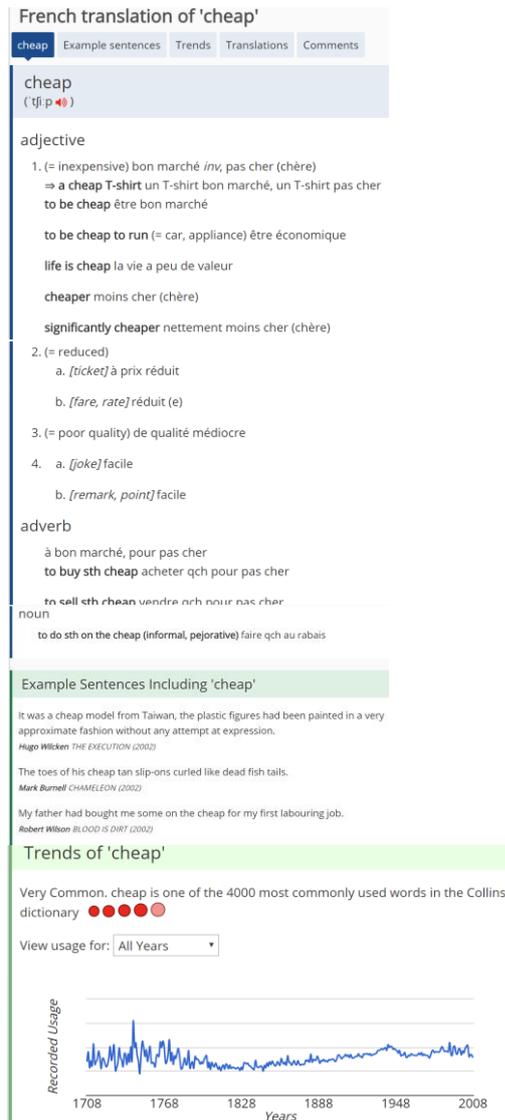
Figuur 6: Die eerste gedeelte van die alfabetiese strek B in OENAD

Dit is in die eerste plek nie duidelik waarom woorde soos 'baby disease characterised by the bumping head of a child which is believed to be very dangerous if not treated earlier' en 'bare fruits' (sic) (OSED), 'baby-grand' en 'barbarousness' (OTED) en 'balderdash' (OENAD) enigsins in die betrokke woordeboeke opgeneem is nie — dit is sonder twyfel lemmas wat deur Prinsloo en De Schryver (1999: 258) as "unlikely to be looked for by the target user at the expense of essential ones" beskou kan word, veral vir sulke relatief klein woordeboeke. Dit terwyl 'n groot aantal topfrekwente woorde wat met die hoogste stergradering (***) in MED en die twee hoogste kategorieë gevulde sirkelgraderings (**** en *****) in *Collins English Dictionary* (CED) as sodanig gemerk is, soos *base*, *beside*, *best*, *blue*, *born*, *brief*, *brilliant*, *bus* en *busy* nie in die lemmalys van OSED opgeneem is nie. Vergelyk ook De Schryver en Prinsloo (2000: 292) vir 'n uitvoerige bespreking van leemtes in lemma-aanbod in bestaande Afrikataalwoordeboeke wat nie met behulp van 'n korpus saamgestel is nie. Hierdie bewerkings bied verder soos wat Gouws (1990: 55) beklemtoon, slegs beperkte vertaalekwivalente wat hoogstens vir teksresepsie van hulp kan wees, maar nie veel bydra tot teksproduksiebehoefte van gebruikers nie. Die yl leksikografiese aanbod maak kennelik nie gebruik van die oorvloed van leksikografies relevante inligting wat deur 'n korpus gebied word nie.

Die vraag kan maar net weer in die woorde van Van Wyk (1995: 89) gevra word of die gebruikers van woordeboeke vir Engels, en argumentsonhalwe Duits en Frans tevrede sou wees met die leksikografiese kwaliteit as sulke woordeboeke die amptelike twee- of drietalige woordeboeke tot hulle beskikking was? Die antwoord is stellig "nee" — die omvang van die bewerkings in OENAD, OSED en OTED voldoen nie aan die minimum vereistes soos in para-

graaf 3 in navolging van Atkins (1998: 3), Gouws en Prinsloo (2005: 144), en ander uiteengesit is nie.

Tweetalige woordeboeke vir byvoorbeeld Engels en Frans soos in figure 7 en 8 staan in skrilte kontras met dié van die Afrikatale in figure 4, 5 en 6 wat betref die rykheid van die informasie-aanbod.



Figuur 7: Cheap in Collins English–French Dictionary (CEFD) <https://www.collinsdictionary.com/dictionary/english-french/cheap>

German translation of 'cheap'

cheap Example sentences Trends Translations Comments

cheap
(tʃi:p)

adjective

1. (also adverb) billig
behaviour, appearance ordinär

⇒ **to feel cheap** sich *dat* schäbig vorkommen
⇒ **it doesn't come cheap** es ist nicht billig
⇒ **it's cheap at the price** es ist spottbillig

2. (= poor quality) minderwertig
⇒ **cheap money** (economics) billiges Geld

noun

to buy sth on the cheap (informal) etw für einen Pappenstiel kaufen (inf)
⇒ **to make sth on the cheap** (informal) etw ganz billig produzieren

Example Sentences Including 'cheap'

It was a cheap model from Taiwan, the plastic figures had been painted in a very approximate fashion without any attempt at expression. **i**
Hugo Wilcken THE EXECUTION (2002)

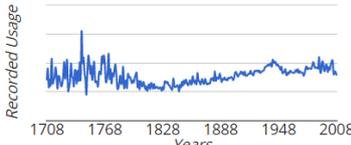
The toes of his cheap tan slip-ons curled like dead fish tails.
Mark Burnell CHAMELEON (2002)

My father had bought me some on the cheap for my first labouring job.
Robert Wilson BLOOD IS DIRT (2002)

Trends of 'cheap'

Very Common. cheap is one of the 4000 most commonly used words in the Collins dictionary ●●●●●

View usage for: All Years



Figuur 8: *Cheap* in *Collins English–German Dictionary* (CEGD)
<https://www.collinsdictionary.com/dictionary/english-german/cheap>

Ter wille van goeie perspektief moet egter vermeld word dat in die samestelling van Afrikataalwoordeboeke die korpusrewolusie nie totaal ongesiens verbygegaan het nie. In 2007 verskyn die *Oxford Bilingual School Dictionary: Northern Sotho and English* (ONSD). Die samestelling van die lemmalys aan die Noord-Sothokant is gebaseer op frekwensietellings uit die korpus, gebruiksvoorbeelde word uit die korpus onttrek en korpusdata word ook aangewend om betekenisonderskeiding te fasiliteer. Die woordeboek bevat ook 'n korpusgebaseerde grammatika as deel van die makrostruktuur. Hierdie woordeboek word in 2010 en 2014 deur twee stalmaats, die *Oxford Bilingual School Dictionary: Zulu and English* (OZD) en die *Oxford Bilingual School Dictionary: IsiXhosa and English* (OXD) onderskeidelik gevolg waarin dieselfde samestellingstrategieë aangewend is. In 2012 verskyn die *Tlhalosi ya Medi ya Setswana*, deur Thapelo J. Otlogetswe, 'n korpusgebaseerde, eentalige, verklarende woordeboek in Tswana. Benewens die gebruik van 'n korpus vir die samestelling van 'n lemmalys, hulp met betekenisonderskeiding en onttrekking van gebruiksvoorbeelde, word korpusdata hier ook gebruik om woordgroepe ("word clusters") waarin bepaalde Tswana-terme optree te identifiseer. Die eerste korpusgebaseerde woordeboek vir Swahili, *Swahili–Suomi–Swahili sanakirja* deur Abdulla et al., word in 2002 gepubliseer. Dit is 'n tweetalige Swahili–Fins–Swahili woordeboek en die *Helsinki Corpus of Swahili* (HCS), in daardie stadium bestaande uit ongeveer 12 miljoen woorde (<http://urn.fi/urn:nbn:fi:lb-2014032624>) is gebruik vir die samestelling daarvan.

Vir die Afrikatale en ander hulpbronneperekte tale word die samestelling van korpusgebaseerde woordeboeke waarin die volle potensiaal van die korpus benut word, egter steeds aan bande gelê deur 'n gebrek aan hulpbronne, spesifiek tekskorpora van gepaste omvang en kwaliteit; dit is gebalanseerde en verteenwoordigende korpora. Aan die ander kant word daar ook geen sigbare pogings van die samestellers van Afrikataalwoordeboeke aangewend om sodanige korpora saam te stel nie.

5.2 Aanlyn woordeboeke

Wat aanlyn woordeboeke betref, was daar aanvanklik groot opgewondenheid. Meijs (1990) het selfs voorspel dat papierwoordeboeke voor die einde van die vorige eeu tot 'n einde sou kom. Dit het natuurlik nie gebeur nie. Die algemene neiging was aanvanklik om die beste van twee wêrelde te verkry deur, soos in MED, 'n CD-ROM-weergawe van die papierwoordeboek saam met die gedrukte weergawe te voorsien. In ander gevalle is papierwoordeboeke feitlik net so op rekenaar geplaas — die tipiese "dictionary behind glass" — of met enkele soekfunksies verryk. Teleurstelling word in die literatuur teen sulke woordeboeke uitgespreek:

The dictionary of the present is at heart little different from the dictionary of the past. Will the dictionary of the future simply blip its little electronic way off into the sunset dazzling its readers with the speed which it dishes up the same old facts on a technicolor screen? It is up to us to take up the real challenge of the computer age, by asking not how the computer can help us to produce old-style dictionaries better, but how it can help us to create something new. (Atkins 1996: 515-6)

Ten spyte van 'n aanvanklik trae wegspring, het elektroniese woordeboeke wat van eenvoudige elektroniese funksies gebruik maak gaandeweg ontwikkel tot aanlyn woordeboeke wat meer unieke elektroniese eienskappe vertoon soos klikbare audio-uitspraakleiding, opwipkieslyste ("drop down menus"), opwip-inligting wat verkry word wanneer die muis oor die skerm beweeg word ("mouse-over" funksies), koppeling met addisionele inligting soos portale en verskillende tipes leksikografiese hulpmiddels (Heid et al. 2012: 274, 276).

Though 'electronic lexicography' — the use of digital media for delivering dictionary data — dates back at least as far as 1990, the pace of change has picked up dramatically in the last five years, after a leisurely start. (Rundell 2012: 72)

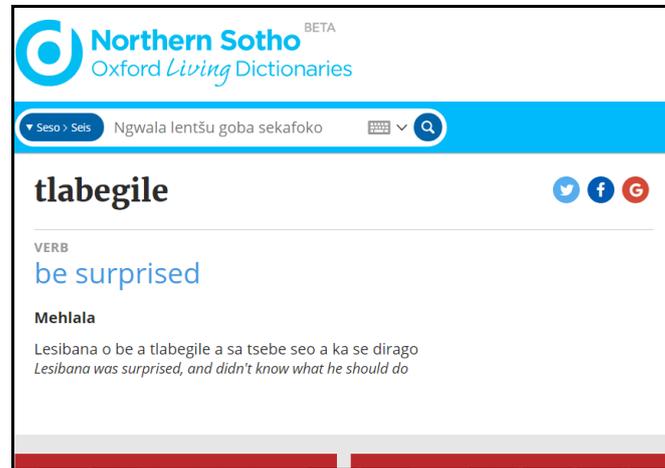
Reeds in 2003 berig De Schryver (2003: 6) dat 182 aanlyn woordeboeke vir die Afrikatale in 117 verskillende tale beskikbaar is. Hy merk egter terselfdertyd op dat die Suid-Afrikaanse Afrikatale onderverteenvoerdig is, dat aanlyn woordeboeke vir die Afrikatale oor die algemeen klein is en nie altyd van ewe goeie gehalte is nie. 'n Moontlike rede hiervoor is dat die oorgang van papierwoordeboeke na elektroniese/aanlyn woordeboeke gekom het in 'n stadium toe hierdie tale mank gegaan het aan goeie papierwoordeboeke. Potensiële woordeboeksamestellers het dus nie die voordeel gehad van goed deurdragte leksikografiese praktyk, waarop hulle sou kon voortbou nie. Hoewel steeds onderverteenvoerdig, is daar aanduidings dat belangstelling in die daarstel van aanlyn woordeboeke vir die Suid-Afrikaanse Afrikatale wel aan die toeneem is. Enkele sodanige inisiatiewe word kortliks bespreek.

In 2004 word die eerste Suid-Afrikaanse aanlyn verklarende Noord-Sotho-woordeboek vrygestel, die *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete* (TLH). Wat hierdie aanlyn woordeboek besonders maak, is dat dit deur Suid-Afrikanners, vir Suid-Afrikanners saamgestel is. TLex (voorheen TshwaneLex, <http://tshwanedje.com/tshwanelex/>), die sagteware wat gebruik is vir die samestelling van hierdie woordeboek, is byvoorbeeld uit die staanspoor ontwikkel met die oog op die Afrikataalleksikografie. 'n Soortgelyke inisiatief is die samestelling van die *Bukantswe: An online Sesotho dictionary* (<http://bukantswe.sesotho.org/>) wat Olivier (2016: 146) as 'n basiese Engels-Sesotho-aanlyn woordeboek beskryf. Dit bevat tans 10 075 lemmas en laat die gebruiker toe om beide Suid-Sotho en Engelse soekwoorde in te tik.

'n Punt van kritiek wat teen albei voorafgenoemde aanlyn woordeboeke geopper kan word, is dat nie een van hulle maksimaal gebruik maak van een van die belangrikste voordele van aanlyn woordeboeke nie, dit is naamlik die feit dat opdatering op 'n gereelde basis gedoen kan word, 'n tekortkoming wat deur Olivier (2016: 148) vir *Bukantswe* erken word. Dit is veral jammer dat hierdie voordeel nie in geval van die Noord-Sothowoordeboek benut word nie, aangesien voorsiening gemaak word vir simultane/gelyktydige terugvoer ('simultaneous feedback'), 'n proses waarvolgens 'n woordeboek aanlyn beskikbaar gemaak word wanneer die samestelling daarvan nog aan die gang is. Deur onder andere die analise van loglêers en direkte terugvoer van gebruikers via die woordeboekoppervlak kan daar noodsaaklike inligting ten opsigte van die teikengebruiker se inligtingsbehoefte ingewin word. Hierdie inligting kan dan aangewend word om seker te maak dat in hierdie behoeftes voorsien word. Sedert die aanvanklike vrystelling van aanlyn Noord-Sothowoordeboeke, is geen opdatering nog gedoen nie. Desnieteenstaande verteenwoordig veral die *Pukuntšuthaloši* 'n belangrike leksikografiese hulpbron waarvan die volle potensiaal nog ontgin moet word.

In teenstelling met voorafgenoemde aanlyn woordeboekprojekte wat plaaslik van aard is, is daar twee inisiatiewe wat buite Suid-Afrika gesetel is, maar met een of meer Suid-Afrikaanse Afrikatale as fokus, naamlik *Oxford Global Languages* (OGL) (<https://www.oxforddictionaries.com/ogl>) en die *Bilingual Zulu-English Dictionary* (<https://isizulu.net/>). Die doel van die OGL-projek is om nou saam te werk met plaaslike taalgemeenskappe om sogenaamde lewende woordeboeke te bou en uit te brei. Drie Suid-Afrikaanse tale vorm reeds deel van die projek, te wete Noord-Sotho, Zulu en Tswana. Aangesien Oxford University Press die uitgewers van die *Oxford Bilingual School Dictionary: Northern Sotho and English* (ONSD) was, is die elektroniese databasis wat die papierwoordeboek onderlê ook vir die aanlyn weergawe gebruik. Wat egter vreemd opval, is dat die aanlyn woordeboek in vergelyking met die papierwoordeboek 'n verskraalde leksikografiese aanbod aan die gebruiker bied. Vergelyk die inskrywings vir *-tlabegile* 'verbaas wees' in die papierwoordeboek en die aanlyn weergawe in Figuur 9 onderskeidelik:

tlabegile /tlabêgilê/ verb + neuter-passive (eg) + perfect (ile)<TLABA►
be surprised •Lesibana o be a **tlabegile**
a sa tsebe seo a ka se dirago. *Lesibana was surprised, and didn't know what he should do.*



Figuur 9: *-tlabegile* in die ONSD en die *Northern Sotho Oxford Living Dictionary*

Uit die voorbeeld hierbo is dit duidelik dat die ONSD 'n veel vollediger leksikografiese bewerking van die lemma gee. Dit verstrek naamlik uitspraakleiding en verskaf morfologiese inligting oor die verskillende werkwoordsuffikse wat aan die stam waarvan die lemma afgelei is, gevoeg is. Hierdie inligting ontbreek in die aanlyn weergawe.

Prinsloo (2012: 134) verwys tereg na *isiZulu.net* as die mees gesofistikeerde aanlyn woordeboek vir die Suid-Afrikaanse Afrikatale. Dit verskaf uitgebreide gebruikersleiding, 'n kompakte weergawe van die belangrikste aspekte van die Zulugrammatika, uitspraakleiding en maak op sinvolle wyse van hiperskakels en kruisverwysings gebruik. Dit breek ook weg van die tradisie van stamlemmatisering wat kenmerkend van die meerderheid van die Ngunitale is — gebruikers kan 'n soekwoord intik sonder om eers die stam te identifiseer. Prinsloo (2012: 135) kom tot die volgende gevolgtrekking: "isiZulu.net does well in analysing complex words in terms of their morphemes and related derivations and gives semantic guidance where possible".

Een van die voordele van aanlyn woordeboeke is dat dit aan verskillende tipes hulpmiddels ("NLP extensions") gekoppel kan word. Hierdie aspek word in 6 hieronder aan die orde gestel.

6. Die toekoms van leksikografie

Dit is uiteraard nie moontlik om presies te voorspel hoe die toekoms van die leksikografie daar sal uitsien nie. Toekomsverwagtinge kan maar ten beste uitgespel word vanuit die *status quo* en huidige tendense in die leksikografie.

6.1 Uitfasering van papierwoordeboeke ten gunste van aanlyn woordeboeke

Na verwagting sal papierwoordeboeke geleidelik uitgefaseer word en toeneemend deur aanlyn woordeboeke vervang word (Meijs 1990: 69-70). Dat aanlyn woordeboeke papierwoorde in die toekoms kan vervang, is nie meer bloot 'n voorspelling nie — Macmillan neem in 2012 die voortou deur nie meer papierwoordeboeke uit te gee nie, maar slegs elektroniese woordeboeke met die dramatiese aankondiging: "[A]t Macmillan, we've taken the decision to phase out printed dictionaries and focus on our rich and expanding collection of digital resources" (Rundell 2012). Dit wat reeds in 1990 deur Meijs vir die einde van die eeu voorspel is, kom 'n dekade later tog as 'n verrassing, soos gesien kan word in opmerkings soos dié op Macmillian Dictionary se webblad *Stop the Presses — The End of the Printed Dictionary* en *STOP THE PRESS: Dictionary No Longer a Page-turner*.

Dié proses sal na verwagting stadiger plaasvind vir Afrikataalleksikografie as vir Engels. Uitgewers in Suid-Afrika is dit eens dat die papierwoordeboekemark tans steeds oorheersend is (Persoonlike mededeling: Phillip Louw, Bestuurder: Woordeboekinhoudontwikkelaar en Tegniese stelsels, Oxford University Press, Kaapstad). Rundell (2015: 303) se stelling hieronder som waarskynlik die Suid-Afrikaanse situasie goed op:

In many parts of the world paper dictionaries still have a healthy future ahead of them. Furthermore, certain types of dictionaries — such as those designed for schools, or special-subject dictionaries, or dictionaries of 'smaller' languages — may show a preference for print for some time to come.

6.2 Die toevallige woordeboekgebruiker

Dit kan ook verwag word dat die inligtingsbehoefte van die toevallige woordeboekgebruiker ("on the fly user") sal toeneem en dat die huidige fokus op die bevrediging van hulle behoeftes voortgesit sal word. Leksikograwe moet dus toenemend ingestel wees op gebruikers wat nie noodwendig aanleerders van die taal is en derhalwe uitvoerige teksresepsie-, teksproduksie- en kognitiewe inligting benodig nie, maar ook op dié gebruikers wat 'n onmiddellike oplossing benodig. Verdere gebruikerstudies om die presiese behoeftes van hierdie tipe gebruiker vas te stel, is nodig.

6.3 Die inligtingsera

Daar is ook min twyfel dat die invloed van die inligtingsera skerp sal toeneem en dat dit selfs 'n gevaar vir die voortbestaan van woordeboeke kan inhou deurdat gebruikers bloot 'n soekenjin sal gebruik eerder as 'n woordeboek, veral as woordeboekdata nie gratis beskikbaar is nie. Rundell (2012: 72) maak

die ietwat onheilspellende voorspelling: "Dictionaries are going the same way as encyclopedias". Hy verwag dat "the new and ongoing digital revolution will be more disruptive. Its effects are 'external', in that it impacts directly on dictionary users, and is in a sense driven by their changing behaviour".

6.4 Koppeling en ondersteuningstrukture

Talle nuwe hulpmiddels vir die woordeboekgebruiker het in die afgelope aantal jare die lig gesien. Sulke hulpmiddels is van groot waarde vir die gebruiker, veral in gevalle waar gebruikersbehoefte nie binne die bestek van woordeboekartikels of selfs die addisionele komponente soos voor- en agterwerk in die woordeboek bevredig kan word nie. Hieronder tel *Interactive Language Toolbox* (<https://ilt.kuleuven.be/inlato/>), wat addisionele inligting verskaf oor teks, woorde en woordkombinasies in Nederlands, Engels en Frans, 'n hulpmiddel vir die gebruik van kollokasies vir aanleerders van Spaans (Alonso Ramos et al. 2014), Engels–Zulu possessiefkonstruksies (Bosch en Faaß 2014), die *Sepedi-helper* vir die konstruksie van sinne in die verskillende modi in Noord-Sotho (Prinsloo et al. 2015), ens. Dit is alles hulpmiddels wat beskikbaar is vir die woordeboekgebruiker in gevalle waar die standaardwoordeboekartikel nie voldoende inligting verskaf nie. Die skryfhulpmiddels *Zulu e-Dict test version* (Bosch en Faaß 2014) en die *Sepedi-helper* (Prinsloo et al. 2015) is goeie voorbeelde van interaktiewe hulpmiddels vir teksproduksie waar die gebruiker insette moet lewer en sy eie kennis met die verwerkingskrag van die rekenaarsagteware moet kombineer.

E-woordeboeke wat met sulke hulpmiddels gekoppel is ten einde meer inligting aan die gebruiker te verskaf, kan toekomstige woordeboeke meer kompetierend met soekenjins maak. Dit is dus verblydend dat daar 'n toename is in die samestelling van sulke hulpmiddels, maar Heid et al. (2015) meen dat die ontwikkeling van sulke hulpmiddels op 'n ongekoördineerde en onsistematiese wyse plaasvind. Heid et al. 2015 poog derhalwe om hulpmiddels te orden en hou 'n sogenaamde taksonomie van hulpmiddels voor vir kommunikatiewe en kognitiewe ondersteuning van woordeboekgebruikers in veral teksproduksie en teksresepsie.

Die taksonomie stel onder meer 'n basiese onderskeid tussen kommunikatiewe en kognitiewe hulpmiddels voor en fokus op verskillende tipes hulpmiddels vir teksresepsie versus teksproduksie. Hierdie taksonomie sal leksikografiese help om leemtes te identifiseer in die ontwikkeling van hulpmiddels.

Een aspek is egter seker, naamlik dat rekenaartegnologie 'n groot invloed op die samestelling van toekomstige woordeboeke sal hê en ook dat inligtingsbronne soos die internet 'n belangrike rol sal speel.

There is much to be done, but the direction of travel is for dictionaries to exploit the electronic medium more fully. What is harder to predict is where this process will end (or even if it will end). There is no particular reason why dictionaries should survive in anything like their present form ... (Rundell 2012: 75)

Rundell (2012) verwag dat daar in die toekoms nie meer formeel onderskeid gemaak sal word tussen woordeboeke, tesourusse en ensiklopedieë nie. Skarebenutting ("Crowdsourcing") en gebruikersgegenereerde inhoud asook woordeboeke wat by die gebruiker aanpas ("adaptable" en "adaptive" modelle) en selfs woordeboeke wat hulleself opdateer word in die vooruitsig gestel.

It is equally plausible to imagine that its heterogeneous functions might be better performed by separate, more specialized resources: among others, automatic translation tools, text-remediation software, or the kinds of tool described by Prinsloo et al. (2011), which are designed to guide users' lexical and grammatical choices in text-production mode. (Rundell 2012: 78)

7. Toekoms van die Afrikataalleksikografie

Die vraag is dan tot watter mate die Afrikataalleksikografie by die verwagte ontwikkelings sal aanpas? Sal Gouws (2007: 315, 316) se waarneming dat daar 'n ontluikende inisiatief by moedertaalsprekers van Afrikatale bestaan om goeie woordeboeke saam te stel werklik momentum kry?

Die gevare vir Afrikataalleksikografie kan as't ware in een sin opgesom word: Die omvang van probleme, bedreigings, behoeftes en tekort aan hulpmiddels het toegeneem. Is Afrikataalleksikografie besig om die stryd te wen of te verloor? Hoe kan bestaande stukragte ontwikkel en beter aangewend word vir die verbetering van Afrikataalwoordeboeke? Onder die stukragte tel leksikografieprojekte soos TUKI, ALRI, die nuwe inisiatiewe deur die *South African Centre for Digital Language Resources* (SADilaR) om korpusse te bou en algemeen beskikbaar te maak, die pogings van die *African Association for Lexicography* (Afrilex) om leksikografie te bevorder, inisiatiewe van uitgewers en entrepreneurs vir die publikasie van woordeboeke, opleiding in leksikografie aan tersiêre inrigtings, onderriginisiatiewe vir woordeboekgebruik, ontwikkeling van hulpmiddels soos die *Sepedi-helper* en die *Zulu e-Dict*, en deelname van gebruikers aan die samestelling van woordeboeke ("user-generated content").

Indien daar 'n enkele aspek uitgesonder kan word wat met groot vrug deur Afrikataalleksikograwe ontgin kan word, is dit die omvattende navorsing wat tot dusver oor spesifiek die Afrikataalleksikografie gedoen is, en steeds gedoen word. Indien leksikograwe kennis neem van hierdie navorsingsresultate kan dit ongetwyfeld lei tot die saamstel van beter woordeboeke.

8. Samevatting

Afrikataalleksikografie ontwikkel nie in isolasie nie, maar word beïnvloed deur dieselfde ontwikkelingstendense en dryfvere wat internasionaal voorkom. In hierdie artikel is aangevoer dat Afrikataalleksikografie nie so flink reageer op hierdie tendense en nuwe moontlikhede wat die rekenaarera bied nie. Dit kan onder meer toegeskryf word aan die feit dat die rekenaarera vir Afrikataalwoordeboeke aangebreek het op 'n tydstip toe die leksikografiese standaard van papierwoordeboeke nog nie op internasionale peil was nie. Die toene-

mende eise wat daar vandag en in die toekoms aan die Afrikataalleksikografie gestel sal word, oortref die beskikbare stukragte. Die probleme en vereistes vir Afrikataalleksikografie het toegeneem soos die gebrek aan goeie papier- en elektroniese woordeboeke, 'n gebrekkige woordeboekkultuur, onvoldoende en ontoeganklike korpora, onvoldoende leksikografiese hulpmiddels, finansiële beperkings, ensovoorts.

Die somber gevolgtrekking waartoe in hierdie artikel gekom word, is dat die leksikografiese agterstand van die Afrikataalleksikografie besig is om te vergroot.

Hierdie veronderstelde proses van agteruitgang kan egter omgekeer word deur maksimale benutting van die genoemde stukragte, gedryf deur Afrikataalleksikograwe. Woordeboeksamestellers in Afrika moet 'n deeglike voorafstudie doen van papier- en elektroniese woordeboeke van goeie kwaliteit vir tale soos Engels, Frans en Duits. Hulle moet ook deeglik kennis neem van bestaande literatuur oor problematiese aspekte van Afrikataalleksikografie. Bestaande Afrikataalwoordeboeke van goeie gehalte moet as rolmodelle geneem word. Leksikografieprojekte en beheerliggame soos Die Pan Suid-Afrikaanse Taalraad (PanSAT) moet in die toekoms meer doelgerigte ondersteuning en leiding aan woordeboekprojekte en leksikografie-eenhede bied. Woordeboeksamstelling moet toenemend korpusgebaseerd wees en meer woordeboekhulpmiddels soos die *Sepdihelper* moet vir Afrikataalwoordeboeke saamgestel word.

9. Erkennung van finansiële ondersteuning

Hierdie navorsing is ondersteun deur die National Research Foundation of South Africa (spesifieke toekennings-verwysingsnommers (UID) 85763 en 103898). Die ontvangers verklaar ook dat menings, bevindinge, gevolgtrekkings en aanbevelings in hierdie NRF-ondersteunde navorsing dié is van die outeurs en dat die NRF geen verantwoordelikheid daarvoor aanvaar nie.

Eindnote

1. Met 'korpus' word hier elektroniese korpus bedoel.
2. Statistiek vir aantal sprekers van genoemde tale verkry van: *Ethnologue Languages of the World*. <https://www.ethnologue.com/>. [Besoek: 03-10-2017.]

Verwysings

Woordeboeke

- (CALD) **McIntosh, C. (Red.)**. 2013. *Cambridge Advanced Learner's Dictionary*. Vierde uitgawe. Cambridge: Cambridge University Press.
- (CED) *Collins English Dictionary*. <https://www.collinsdictionary.com/dictionary/english>.
- (CEFD) *Collins English-French Dictionary*. <https://www.collinsdictionary.com/dictionary/english-french/cheap>.

- (CEGD) *Collins English–German Dictionary*. <https://www.collinsdictionary.com/dictionary/english-german/cheap>.
- (COBUILD) **Carroll, K. (Red.)**. 2012. *Collins COBUILD Advanced Dictionary of English*. Sewende uitgawe. Glasgow: HarperCollins.
- (GNSW) **Ziervogel, D. en P.C. Mokgokong**. 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*. Pretoria: J.L. van Schaik.
- (LDOCE) **Mayor, M. (Red.)**. 2009. *Longman Dictionary of Contemporary English*. Vyfde uitgawe. Harlow, Essex: Pearson Education.
- (MED) **Rundell, M. (Red.)**. 2007. *Macmillan English Dictionary for Advanced Learners*. Tweede uitgawe. Oxford: Macmillan.
- (OALD) **Deuter, M., J. Bradbery en J. Turnbull (Reds.)**. 2015. *Oxford Advanced Learner's Dictionary of Current English*. Negende uitgawe. Oxford: Oxford University Press.
- (OENAD) **Mahlangu, K.S. (Red.)**. 2015. *English–Isindebele–Afrikaans: IZiko lesiHlathululi-mezwi sesiNdebele Trilingual Dictionary*. (Isindebele National Lexicography Unit.) S.p.: South African Heritage Publishers.
- (ONSD) **De Schryver, G.-M. (Red.)**. 2007. *Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English*. Kaapstad: OUP Southern Africa.
- (OSED) **Mareme, G.B. (Red.)**. 2015. *Thanodi ya Temepedi Setswana-English*. (Setswana National Lexicography Unit.) S.p.: South African Heritage Publishers.
- (OTED) **Tshikota, S.L. (Red.)**. 2015. *Thalusamaipfi ya nyambo-mbili ya Tshivenda–English. English–Tshivenda. Bilingual Dictionary. Mathivha-Milubi-Maḡadzhe Edition*. (Tshivenda National Lexicography Unit.) S.p.: South African Heritage Publishers.
- (OXD) **De Schryver, G.-M. en M. Reynolds (Reds.)**. 2014. *Oxford Bilingual School Dictionary: IsiXhosa and English*. (Eerste uitgawe.) Kaapstad: OUP Southern Africa.
- (OZD) **De Schryver, G.-M. (Red.)**. 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi*. Kaapstad: OUP Southern Africa.
- (POP) **Kriel, T.J., D.J. Prinsloo en B.P. Satheke**. 2007. *Popular Northern Sotho Dictionary, Northern Sotho–English, English–Northern Sotho*. Vierde uitgawe. Kaapstad: Pharos.
- (PUKU) **Kriel, T.J., E.B. van Wyk en S.A. Makopo**. 1989. *Pukuntšu woordeboek (Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho)*. Vierde, hersiene en uitgebreide uitgawe. Pretoria: J.L. van Schaik.
- (TLH) *Pukuntšuthaloši ya Sesotho sa Leboa ka Inthanete*. Aanlyn beskikbaar: <http://africanlanguages.com/psl/>.

Ander bronne

- Abdulla, A., R. Halme, L. Harjula en M. Pesari-Pajunen (Reds.)**. 2002. *Swahili–Suomi–Swahili sanakirja*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Alonso Ramos, M., M. Garcia Salido en O. Vincze**. 2014. Towards a Collocation Writing Assistant for Learners of Spanish. Faaß, G. en J. Ruppenhofer (Reds.). 2014. *Workshop Proceedings of the 12th Edition of the KONVENS Conference, Hildesheim, Germany, October 8–10, 2014: 77–88*. Hildesheim: Universitätsverlag Hildesheim.

- Atkins, B.T.S.** 1996. *Bilingual Dictionaries: Past, Present and Future*. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström en Catarina Rödger Pappmehl (Reds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gotenburg: Departement Sweeds, Universiteit van Gotenburg.
- Atkins, B.T.S.** 1998. *Some Discussion Points Arising from Afrilex-Salex '98*. Ongepubliseerde kursus-evalueringsdokument, Universiteit van Pretoria.
- Atkins, B.T.S. en K. Varantola.** 1998. Monitoring Dictionary Use. Atkins, B.T.S. (Red.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 21-82. Tübingen: Max Niemeyer.
- Awak, M.K.** 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.R.K. (Red.). 1990: 8-18.
- Barnhart, C.L.** 1962. Problems in Editing Commercial Monolingual Dictionaries. Householder, F.W. en S. Saporta (Reds.). 1962. *Problems in Lexicography*: 161-181. Bloomington: Universiteit van Indiana.
- Bosch, S.E. en G. Faaß.** 2014. Towards an Integrated e-Dictionary Application — the Case of an English to Zulu Dictionary of Possessives. Abel, Andrea, Chiara Vettori and Natascia Ralli (Reds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*: 739-747. Bolzano/Bozen: EURAC Research.
- Busane, M.** 1990. Lexicography in Central Africa: The User Perspective, with Special Reference to Zaïre. Hartmann, R.R.K. (Red.). 1990: 19-35.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M. en D.J. Prinsloo.** 2000. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 1: The Macrostructure. *South African Journal of African Languages* 20(4): 291-309.
- De Schryver, G.-M. en D.J. Prinsloo.** 2001. Corpus-based Activities versus Intuition-based Compilations by Lexicographers. The Sepedi Lemma-Sign List as a Case in Point. *Nordic Journal of African Studies* 10(3): 374-398.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Red.). 1990: 52-65.
- Gouws, R.H.** 2007. On the Development of Bilingual Dictionaries in South Africa: Aspects of Dictionary Culture and Government Policy. *International Journal of Lexicography* 20(3): 313-327.
- Gouws, R.H.** 2016. Op pad na 'n omvattende woordeboekkultuur in die digitale era. *Lexikos* 26: 103-123.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Haas, M.R.** 1962. What Belongs in a Bilingual Dictionary? Householder, F.W. en S. Saporta (Reds.). 1962. *Problems in Lexicography*: 45-50. Bloomington: Universiteit van Indiana.
- Hanks, P.** 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.
- Hartmann, R.R.K. (Red.).** 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24-26 March 1989*: Exeter Linguistic Studies 15. Exeter: University of Exeter Press.

- Heid, U., T.J.D. Bothma en D.J. Prinsloo.** 2015. *Hidden e-Dictionaries as Part of Language-related User Guidance Tools*. Plakkaataanbieding. eLex 2015, Herstmonceux Kasteel, VK, 11–13 Augustus 2015.
- Heid, U., D.J. Prinsloo en T.J.D. Bothma.** 2012. Dictionary and Corpus Data in a Common Portal: State of the Art and Requirements for the Future. *Lexicographica* 28: 269-291.
- Heine, B. en D. Nurse.** 2000. *African Languages: An Introduction*. Cambridge: Cambridge University Press.
- Interactive Language Toolbox*. Aanlyn beskikbaar: <https://ilt.kuleuven.be/inlato/>.
- Jones, J., S.E. Bosch, L. Pretorius en D.J. Prinsloo.** 2005. Development of Reusable Resources for Human Language Technologies (HLT) Applications: Practice and Experience. *SA Journal of African Languages* 25(2): 141-159.
- Lauffer, Batia.** 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, H. et al. (Reds.). 1992. *EURALEX '92 Proceedings I-II: Papers Submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland, 4–9 August 1992*: 71-76. Tampere: Universiteit van Tampere.
- Macmillan Dictionaries.** *From Corpus to Dictionary*. <http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/>. [Besoek 03-10-2017.]
- Macmillan Dictionary Blog.** *Stop the Presses — The End of the Printed Dictionary*. Aanlyn beskikbaar: <http://www.macmillandictionaryblog.com/bye-print-dictionary>. [Besoek 09-07-2017.]
- Macmillan Education.** *STOP THE PRESS: Dictionary No Longer a Page-turner*. Aanlyn beskikbaar: <http://www.macmillaneducation.com/MediaArticle.aspx?id=1778>. [Besoek 09-07-2017.]
- Mann, M.** 1990. A Linguistic Map of Africa. Hartmann, R.R.K. (Red.). 1990: 1-7.
- Meijs, W.J.** 1990. Morphology and Word-formation in a Machine-readable Dictionary: Problems and Possibilities. *Folia Linguistica: Acta Societatis Linguisticae Europaeae* 24(1–2): 45-71.
- Nkomo, D.** 2012. *Towards a Lexicographical Intervention in the Acquisition and Use of English in Zimbabwe*. Ongepubliseerde doktorsproefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Olivier, J.** 2016. Sesotho Online: Establishing an Internet-based Language Knowledge Community. *South African Journal of African Languages* 36(2): 141-152. DOI: 10.1080/02572117.2016.1252007.
- Otto, A.N.** 1989. *Kriteria vir 'n Afrikaanse aanleerderwoordeboek*. Ongepubliseerde D.Litt.-verhandeling. Stellenbosch: Universiteit van Stellenbosch.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.
- Prinsloo, D.J.** 2012. Electronic Lexicography for Lesser-resourced Languages: The South African Context. Granger, Sylviane en Magali Paquot (Reds.). 2012. *Electronic Lexicography*: 119-144. Oxford: Oxford University Press.
- Prinsloo, D.J. en G.-M. de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Prinsloo, D.J. en R.H. Gouws.** 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.
- Prinsloo, D.J., D. Prinsloo en J.V. Prinsloo.** 2015. *Sepedi-helper*. Aanlyn beskikbaar: www.sepedihelper.co.za.
- Rundell, M.** 2012. It Works in Practice but Will it Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Vatvedt Fjeld, Ruth en Julie Matilde Torjusén (Reds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August 2012, Oslo*: 47-92. Oslo: Departement Linguistiek en Skandinawiese Studies, Universiteit van Oslo.

- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- StatsSA.** 2017. *Living Conditions of Households in South Africa. An Analysis of Household Expenditure and Income Data Using the LCS 2014/2015*. Aanlyn beskikbaar: <http://www.statssa.gov.za/publications/P0310/P03102014.pdf>. [Besoek 09-05-2017.]
- Taljard, E., D.J. Prinsloo en I. Fricke.** 2011. The Use of LSP Dictionaries in Secondary Schools — A South African Case Study. *South African Journal of African Languages* 31(1): 87-109.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.
- Wójtowicz, B.** 2016. Survey of Swahili Dictionaries: The Macrostructure. *Studies of the Department of African Languages and Cultures* 50: 5-39.
- Wolvaardt, Jill.** 2017. South Africa's National Lexicography Units: Time for a Reboot? Bosch, S.E. en D. Nkomo (Reds.). 2017. *Abstracts of the 22nd International Conference of the African Association for Lexicography. Conference of the Language Associations of Southern Africa (CLASA), Rhodes University, Grahamstown, South Africa, 26–29 June 2017.*

Überlegungen zu einem neuen zweisprachigen phraseologischen Wörterbuch Deutsch–Rumänisch

Doris Sava, *Department für anglo-amerikanische und
germanistische Studien, Lucian-Blaga-Universität Sibiu,
Sibiu, Rumänien (doris.sava@ulbsibiu.ro)*

Zusammenfassung: Die herkömmliche rumänische zweisprachige Phraseographie mit Deutsch weist erhebliche Mängel auf, die sowohl die konzeptionelle Ausrichtung des Wörterbuchs als auch die Gestaltung der Wörterbuchartikel betreffen. Unter Berücksichtigung der Ergebnisse der bilingualen Speziallexikographie weltweit und der Brauchbarkeit vorhandener Wörterbücher aus heutiger Sicht, ist offensichtlich, dass die Ausarbeitung eines modernen phraseologischen Wörterbuchs, das zuverlässiger und benutzerfreundlicher als bisherige Werke ist, dringend zu fördern ist. Fasst man die Umgestaltung vorhandener Phraseolexika für rumänischsprachige Benutzer des Deutschen ins Auge, so muss das primäre Anliegen der Phraseographen auch darin bestehen, ein Nachschlagewerk auszuarbeiten, das bezüglich der Materialauswahl und der Kodifizierungspraxis höheren Ansprüchen genügen sollte. Der Beitrag zeigt daher Verbesserungsmöglichkeiten der rumänischen bilingualen phraseographischen Beschreibungspraxis mit Deutsch auf, wobei die vorgestellten lexikographische Beschreibungsmuster relevante Informationen zur Rezeption und Produktion der aufgenommenen Phraseologismen bietet.

Schlüsselwörter: LEXIKOGRAPHISCHE STANDARDS, PHRASEOLOGISMEN, LEXIKOGRAPHISCHE BESCHREIBUNG, MODELLVORSCHLAG, ZWEISPRACHIGES PHRASEOLOGISCHES WÖRTERBUCH DEUTSCH–RUMÄNISCH

Abstract: Considerations on a New Bilingual German–Romanian Phraseological Dictionary. The conventional Romanian bilingual phraseography with German contains significant shortcomings regarding both the conceptual orientation of the dictionary and the structure of the dictionary articles. Taking into account the results of specialised bilingual lexicography worldwide and the usefulness of existing dictionaries from a current perspective, it becomes obvious that it is imperative to promote the creation of a modern phraseological dictionary which will be more reliable and user-friendly than the existing works. When the restructuring of the existing phraseological lexica for Romanian-speaking users of the German language is considered, the main concern must be the creation of a reference book which will meet higher requirements regarding the selection of material and the codification practice. Thus, this article points out possible improvements to the Romanian bilingual lexicographical description of phraseological units with German, whereby the presented lexicographical description model offers relevant information for the reception and production of the presented idioms.

Keywords: LEXICOGRAPHICAL STANDARDS, IDIOMS, LEXICOGRAPHICAL DESCRIPTION, PROPOSED MODEL, BILINGUAL GERMAN–ROMANIAN PHRASEOLOGICAL DICTIONARY

1. Vorbemerkungen

Die phraseologische und metalexikographische Forschung hat lexikographische Standards für die Erfassung und Darstellung der Phraseologismen in der Wörterbuchpraxis festgelegt, die zu einer Verbesserung der Wörterbücher beigetragen haben.¹ Die Entwicklung moderner phraseologischer Nachschlagewerke mit Deutsch als Ausgangs- oder Zielsprache ist ein Desiderat. Der Beitrag geht von den Qualitätskriterien für zweisprachige Nachschlagewerke aus, und legt dar, inwieweit eine aus der Sicht der potenziellen Benutzer und der Spezifik phraseologischer Erscheinungen angemessene lexikographische Darstellung gewährleistet werden kann.² Von der Qualität eines Wörterbuchs zeugt weniger die Anzahl der aufgenommenen Lemmata, sondern vielmehr die adressatenspezifische Ausrichtung des Artikelprofils. Die Materialelektion, die Qualität der Wörterbuchangaben sowie Umfang und Konzeption der Umtexte sind weitere Bewertungskriterien, die es zu beachten gilt. Da die rumänische zweisprachige praktische Phraseographie mit Deutsch erhebliche Mängel aufweist, die sowohl die konzeptionelle Ausrichtung des Wörterbuchs als auch die Gestaltung der Wörterbuchartikel betreffen, muss auf die Dringlichkeit eines neuen zweisprachigen phraseologischen Wörterbuchs oder zumindest auf die Notwendigkeit einer Neubearbeitung gegenwärtiger bilingualer Phraseolexika hingewiesen werden. Berücksichtigt man die Ergebnisse der bilingualen praktischen Phraseographie andernorts, so ist zu bedauern, dass hierzulande keine ernsthaften Bemühungen um adäquatere Beschreibungsmöglichkeiten und Kodifizierungsverfahren existieren oder gefördert werden. Die Gründe für dieses Versäumnis sind vielfältig. Sie hängen auch damit zusammen, dass der reichen internationalen (auch germanistischen) wissenschaftlichen Auseinandersetzung mit der phraseologischen Kontrastivik und Phraseographie eine sehr bescheidene Anzahl von Publikationen aus rumänischer Sicht gegenübersteht und dass eine umfassendere Darstellung des rumänischen phraseologischen Systems lange Zeit ausblieb. Die Ermittlung der Vorkommensbereiche und Verwendungsweisen rumänischer Phraseologismen oder Frequenzuntersuchungen zur Produktivität einzelner Komponenten und Strukturmodelle sind Arbeitsfelder, die noch nicht gründlich erforscht wurden. Der Großteil der kontrastiven Untersuchungen aus der Perspektive des anvisierten Sprachpaares sind dem Bereich der interlingualen Äquivalenzermittlung gewidmet und bezwecken vornehmlich das Vorführen von Besonderheiten, die bei der Kontrastierung bestimmter struktureller Typen hervortreten.

Fasst man die Umgestaltung vorhandener Phraseolexika für rumänischsprachige Benutzer des Deutschen ins Auge, so muss das primäre Anliegen der Phraseographen darin bestehen, ein Nachschlagewerk auszuarbeiten, das bezüg-

lich der Materialauswahl und Kodifizierungspraxis höheren Ansprüchen genügt. Im Folgenden werden daher phraseologisch relevante Aspekte, die bei der Erstellung oder Revidierung von lexikographischen Werken zu berücksichtigen sind, angesprochen. Dabei sollen an einem Musterartikel Verbesserungsmöglichkeiten für die rumänische bilinguale phraseographische Beschreibungspraxis mit Deutsch exemplarisch aufgezeigt werden, wobei das vorgestellte lexikographische Beschreibungsmuster relevante Informationen zur Rezeption und Produktion der aufgenommenen Phraseologismen bietet.

Mit der Festlegung des Benutzerkreises und seiner Informationsbedürfnisse ist die Gestaltung der Wörterbuchstruktur und der Informationstypen verbunden, die in der Mikrostruktur vorkommen. Ein zweisprachiges phraseologisches Wörterbuch muss dem Benutzer u.a. folgende Informationen liefern: (1) In welchen Situationen wird der Phraseologismus gebraucht? (2) Können alle Äquivalente in der gleichen Situation gebraucht werden wie der Ausgangssprachliche Phraseologismus? (3) Welche Besonderheiten zeigen die Phraseologismen in der Ausgangssprache?

Bei der Erwartungshaltung des Wörterbuchbenutzers muss in Erwägung gezogen werden, dass der Wörterbuchbenutzer zum Wörterbuch greift, um einen Ausdruck in seiner Form bzw. in seinen Formen wie auch in seiner Bedeutung und Verwendungsweise im aktuellen Sprachgebrauch zu überprüfen. So ein Wörterbuch könnte ein passives Verständnis sichern, im Optimalfall auch eine aktive Verwendung bzw. auch zur Textproduktion anleiten und nicht nur rezeptiv ausgerichtet sein. Die lexikographischen Informationen müssen hinsichtlich der (tatsächlichen und potenziellen) Benutzungssituationen formuliert werden. Bei der Konzeption eines zweisprachigen phraseologischen Wörterbuchs mit Deutsch und Rumänisch wird von der Prämisse ausgegangen, dass das Nachschlagewerk sowohl für die Textrezeption als auch indirekt für die Textproduktion von rumänischen Deutschlernern sowie sekundär für die Textrezeption von deutschsprachigen Rumänischlernern genutzt werden kann.

Ein deutsch-rumänisches Wörterbuch kann von zwei verschiedenen Benutzergruppen zum Nachschlagen herangezogen werden. Daraus ergeben sich zwei unterschiedliche Benutzungssituationen, je nachdem, ob die Muttersprache des Benutzers die Ausgangssprache oder die Zielsprache darstellt, in der Informationen zu bestimmten Phraseologismen gesucht werden. Ein rumänischsprachiger Benutzer wird einen Phraseologismus in einem phraseologischen Wörterbuch Deutsch-Rumänisch nachschlagen wollen, (1) weil er die Bedeutung aus einem gegebenen gesprochenen/geschriebenen Text selbstständig nicht erschließen kann, auch wenn er die Bedeutung einzelner Konstituenten kennt. Er sucht eine Information über einen Phraseologismen in der Ausgangssprache Deutsch; (2) weil er bei einer Übersetzung in die Muttersprache Rumänisch ein Äquivalent sucht und das Wörterbuch eine zielsprachliche Entsprechung liefern sollte; (3) weil der rumänischsprachige Benutzer zwar die Bedeutung eines Phraseologismus kennt, jedoch unsicher ist, was seine ange-

messene kontextuelle Einbettung oder die vollständige/richtige lexikalische Form anbelangt und deshalb danach nachschlagen möchte. Ein Benutzer mit Deutsch als Muttersprache hingegen wird aus anderen Gründen ein phraseologisches Wörterbuch Deutsch–Rumänisch konsultieren wollen. Er verfügt bereits über eine muttersprachliche idiomatische Kompetenz in der Ausgangssprache und sucht gezielt nach Informationen über mögliche zielsprachliche Entsprechungen zu bestimmten Phraseologismen. Dabei kann mit folgenden Benutzungssituationen gerechnet werden: (1) der deutschsprachige Benutzer sucht ein Äquivalent für einen Phraseologismus, der in einem gesprochenen/geschriebenen Text in der Zielsprache Rumänisch vorkommt und benötigt eine Äquivalentangabe für die Hinübersetzung; (2) der Benutzer möchte einen fremdsprachlichen Text produzieren, weiß jedoch nicht, wie er den muttersprachlichen Inhalt phraseologisch in der Zielsprache ausdrücken kann.

Nach der in der zweisprachigen Lexikographie üblichen Wörterbuchtypologie soll das vorgeschlagene Wörterbuchprojekt als bidirektionales Konsultationswörterbuch (Nachschlage- bzw. Bestätigungsfunktion), primär rezeptiv aufgefasst werden, das jedoch auch weiteren Anforderungen nachkommen kann, wie z.B. der Lernfunktion. Es sollte Situationen der Textrezeption und der Herübersetzung abdecken, kann jedoch auch als aktives Wörterbuch aufgefasst werden.

2. Das Gebot der Kürze und die Verlockungen der Ausführlichkeit. Phraseologismen in der lexikographischen Praxis

2.1 Defizite der rumänischen bilingualen Phraseographie mit Deutsch im Überblick

In der lexikographischen Forschung wurde die Frage nach den Bedürfnissen des Benutzers zweisprachiger Wörterbücher intensiv diskutiert, vor allem die nach der angemessenen Auswahl und Darstellung der für die jeweilige Benutzergruppe relevanten Informationen. Die kritische Durchsicht der Wörterbucheinträge hat Überlegungen zu einer sorgfältigeren Beschreibung der Phraseologismen veranlasst. Die Bemühungen der Lexikographen sind auf eine möglichst umfassende Beschreibung der semantischen Komplexität, formalen Vielgestaltigkeit und pragmatischen Vieldeutigkeit ausgerichtet. Dabei treten deutlich die Schwierigkeiten hervor, die mit diesem Unterfangen verbunden sind, wie auch offensichtlich wird, dass ein theoretisch-methodisch fundiertes und auch praktisch umsetzbares lexikographisches Konzept notwendig ist, das auch die Forschungsergebnisse der theoretischen und praktischen, allgemeinen und kontrastiven Phraseologie zu berücksichtigen hat. Die dem Lexikographen von der theoretischen bzw. praktischen Phraseographie zur Verfügung gestellten Lösungen wurden unterschiedlich gewertet: als aufwendig, subjektiv und nicht konsequent durchführbar, wenn auch benutzerfreundlich.³ In einigen Beiträgen (z.B. 2010) bin ich auf die rumänische zwei-

sprachige praktische Phraseographie mit Deutsch als Ziel- und Ausgangssprache⁴ eingegangen, wobei ich die in Rumänien ausgearbeiteten Wörterbücher aus der Perspektive des Nichtmuttersprachlers und seiner Benutzerbedürfnisse bewertet habe. Die durchgesehenen Wörterbücher verzeichnen Phraseologismen alphabetisch, wobei jedem Lemma Äquivalentangaben mit usueller Variantenkennzeichnung zugeordnet werden: *a cădea/a da/a sări din lac în puț* — aus dem/vom Regen in die Traufe kommen (Mantsch/Anuței/Kelp 1979: 289) vs. *vom/aus dem Regen in die Traufe kommen* — *a da din lac în puț* (Roman 1993: 203 und 262). Insgesamt ist die lexikographische Versorgung in diesen Lexika als mittelmäßig bis schlecht zu beurteilen. Diese Einschätzung sei im Folgenden näher begründet.

Bei der Musterung beider Wörterbücher sind zahlreiche Ungenauigkeiten und Verstöße gegen die lexikographische Akribie auszumachen. Roman (1993) geht inkonsequent vor, wenn er im Vorspann Beispiele vorstellt (z.B. *von dem Regen in die Traufe kommen*; S. 5), die im Wörterbuchkörper mit Variantenkennzeichnung kodifiziert werden: *vom/aus dem Regen in die Traufe kommen* (1993: 203 und 262). Bereits im Vorspann begegnen bei Roman einige (Druck-)Fehler und zwar genau in dem Textabschnitt, der den Benutzerhinweisen gewidmet ist und wo der Autor die Prinzipien der lexikographischen Erfassung erläutert. Zur Veranschaulichung wählt Roman einen Phraseologismus, der zwei nominale Elemente aufweist, um darzulegen, unter welcher wendungsinternen Komponente der Phraseologismus gebucht wird. Der deutsche Phraseologismus *vom/aus dem Regen in die Traufe kommen* erscheint an zwei Stellen im Wörterbuch, wird aber nur unter dem ersten Substantiv mit seiner rumänischen Entsprechung angeführt und unter der weiteren Komponente *Traufe* — allerdings in der Form *Von dem Regen in die Traufe Kommen* (sic!) — mit dem Verweis „v. Regen“ [v. = rum. *vezi*; dt. siehe] versehen. Im Wörterbuch allerdings erscheint der Phraseologismus ohne Druckfehler. Gleichfalls inkonsequent gehen auch Mantsch/Anuței/Kelp (1979) vor. Vgl. hierzu die deutschen Entsprechungen zu *a cădea din lac în puț* und *a cădea/a da/a sări din lac în puț* — aus dem/vom Regen in die Traufe kommen (Mantsch/Anuței/Kelp 1979: 104 und 289); *colac peste pupăză* — aus dem Regen in die Traufe kommen (Mantsch/Anuței/Kelp 1979: 137); *a căuta pe dracul și a găsi pe tată-său* — vom/aus dem Regen in die Traufe kommen (Mantsch/Anuței/Kelp 1979: 173) und auch folgende Äquivalenzangaben: „für zwei arbeiten, wie ein Pferd arbeiten“ (S. 311), „schwer arbeiten, wie ein Pferd/Galeerensklave arbeiten, Blut und Wasser schwitzen, etwas ist ein saurer Verdienst“ (S. 233), „schufteten, sich abrackern, sich schinden, sich (ab)mühen, sich placken, wie ein Pferd/Ochse arbeiten, Pferdearbeit verrichten“ (S. 243) und „schwer arbeiten, sich placken, sich schinden, in den Sielen stecken, sich ins Joch spannen“ (S. 289). Gravierend ist insbesondere die Tatsache, dass manche Phraseologismen in verschiedenen Nennformen lexikographisch erfasst worden sind. Nichtmuttersprachliche Wörterbuchbenutzer benötigen ein Maximum an Informationen, wenn das zweisprachige Wörterbuch den Anspruch erhebt, ein Hilfsmittel für die Text-

produktion zu sein. Damit die Nennform eine sichere Textproduktion gewährleisten kann, muss durch die Wahl der Nennform klare Aussagen über mögliche Restriktionen in der Verwendung der Phraseologismen gemacht werden. Daher sollte eine benutzergerechte Nennformgestaltung dem Benutzer die wichtigsten Informationen zur Einbettung des Phraseologismus im Satz liefern und syntaktisch-semantische Angaben zu den Kontextpartnern (zu den obligatorischen und fakultativen Valenzen der Phraseologismen) als Informationstyp explizit und vollständig darbieten. Aus der Analyse der Nennformen geht hervor, dass der Wörterbuchbenutzer unterschiedliche Informationen darüber findet, welche Komponenten einen Phraseologismus ausmachen. So verzeichnen Mantsch/Anuței/Kelp (1979: 229) als deutsche Entsprechung für *acesta este adevărul gol-goluț* „das ist die nackte/reine Wahrheit“. 22 Zeilen darunter erscheint unter *a spune adevărul gol-goluț* das Äquivalent „die nackte/reine/ungeschminkte Wahrheit sagen“. Durch die Vernachlässigung der Information zur Valenzfähigkeit des Phraseologismus und die Reduzierung seiner Nennform kann der Wörterbuchbenutzer nicht immer einschätzen, ob als Handlungsträger eine Person oder auch eine Institution, Einrichtung usw. genannt werden kann. Die Nennformgestaltung sollte Restriktionen kenntlich machen und die korrekte Nennform sollte so formuliert werden, dass Angaben zur externen syntaktisch-semantischen Valenz auch berücksichtigt werden. Für die Klärung der Frage nach den fakultativen und obligatorischen Komponenten des Phraseologismus, die in die Nennform aufgenommen werden sollen, und für eine vollständige und korrekte Kodifizierung sind Recherchen der modernen elektronischen Textkorpora wie auch das Sammeln von Belegen aus modernen Quellen (z.B. aus der aktuellen Presselandschaft) hilfreich. Roman (1993) lässt die externe Valenz oft außer Acht und konkretisiert die Subjektergänzung in der Nennform nicht. Irrtümlicherweise verzeichnet Roman neben den in der lexikographischen Praxis üblichen Abkürzungen *etw.* (= etwas), *jd.* (= jemand) und *js.* (= jemandes) in dem Abkürzungsverzeichnis das Kürzel *jm.* mit der Erklärung *jemanden* (!). Die Abkürzung *jn.* (= jemanden) wurde nicht aufgenommen. Konsequenterweise begegnet man im Wörterbuch dem Element *jemandem* mal in der Abkürzung *jm.* oder *jn.*, sodass auch falsche Nennformangaben vorkommen: *jm. (blauen/eitlen) Dunst vormachen*; *jm. nicht um die Ecke trauen* (S. 60f.); *jm. das Fell über die Ohren ziehen* (S. 77); *jm. mit (eisiger) Kälte empfangen* (S. 134). Zu bemängeln ist auch, dass zwei oder mehrere Phraseologismen in einer einzigen Nennform zusammengefasst werden. Bei Mehrfachlemmatisierungen stimmen Nennformangaben, Markierungen und Äquivalenzangaben oft nicht überein. Als zielsprachliche Äquivalente fungieren Phraseologismen, freie Wortverbindungen, Umschreibungen oder Einwortlexeme, Sprichwörter oder Goethe-Zitate (!). Roman (1993) ordnet in fünf Fällen Sprichwörtern phraseologische Wortverbindungen (S. 216, 240, 267, 292) zu, wobei bekannte Sprichwörter (z.B. *Der Ton macht die Musik*; S. 260; *Übung macht den Meister*; S. 267) ohne den Hinweis *proverb* (dt. Sprichwort) erscheinen. Bei Mantsch/Anuței/Kelp (1979) werden 194 Sprichwörter ohne diesen Hinweis (vgl. S. 73, 82f., 105, 228,

326, 365 oder 613; *Aus den Augen, aus dem Sinn*; *Die Katze lässt das Mäusen nicht*; *Es ist nicht alles Gold, was glänzt*; *Aller Anfang ist schwer*; *Ende gut, alles gut*; *Eile mit Weile*; *Der Apfel fällt nicht weit vom Stamm*; *Kleider machen Leute*; *Wer A sagt, muss auch B sagen*) aufgenommen. Hier werden deutsche Sprichwörter in verschiedenen Varianten angeführt (vgl. z.B. S. 149, 218 oder 557) und auch bei Roman werden 39 deutsche Sprichwörter in variiertem Form dargeboten: z.B. *Über Nacht kommt Rat*; *guter Rat kommt über Nacht* (S. 179) oder *Auf Regen folgt Sonnenschein*; *nach Regen kommt Sonnenschein* (S. 203).

Die Sichtung dieser Wörterbücher zeigt, dass sich pragmatische Informationen eher auf stilistische Markierungen beschränken. Insgesamt kommen geschlechts- oder altersspezifische Markierungen äußerst selten vor. Oft in Form von Abkürzungen geboten, verzeichnen die Wörterbücher in den Vorworten nicht immer, was unter den betreffenden Stilangaben zu verstehen ist. In beiden Wörterbüchern können weitere Versäumnisse registriert werden, die den Wörterbuchbenutzer stören werden. Sowohl bei Mantsch/Anuței/Kelp (1979) als auch bei Roman (1993) erscheinen Abkürzungen, die im Verzeichnis nicht erfasst wurden. Bei Mantsch/Anuței/Kelp kommen im Wörterbuch Abkürzungen wie *prov.* (rum. *proverb*; dt. Sprichwort) oder *înv.* (rum. *învechit*; dt. veraltet) vor, die im Verzeichnis nicht erscheinen. Verwunderlich ist, dass die Autoren 37 Phraseologismen mit dem Hinweis *fig.* (rum. *figurat*; dt. bildhaft; übertragen, umgedeutet) versehen, diese Abkürzung jedoch in die Liste der Abkürzungen nicht aufnehmen. Bei Roman erscheinen sieben Abkürzungen für Markierungsprädikate oder Vorkommensbereiche, die im Vorspann nicht erfasst wurden (z.B. die Markierung *arg.* für rum. *argou*, *argotic*; dt. Gaunersprache). Ein Inhaltsverzeichnis als äußere Zugriffsstruktur zum gesamten Textverbund oder ein Index der Schlüsselwörter, der das Auffinden kodifizierter phraseologischer Einheiten ermöglichen soll, vermisst man in beiden phraseologischen Wörterbüchern.

Das lexikographische Bild rumänischer Phraseolexika verweist auf eine dem erfassten Sprachbestand nicht angemessene lexikographische Bearbeitung. Fasst man bestimmte lexikographische Standards ins Auge, so kann auch von einer sorgfältigen Ausführung der Umtexte bei diesem speziellen Wörterbuchtyp nicht gesprochen werden. Beide Wörterbücher liegen unter dem Gesichtspunkt der Gestaltung der Einleitung weit unter dem Stand eines benutzerfreundlichen Nachschlagewerks. Sie enthalten zwar im rumänischsprachigen Vorspann ein kurzes Vorwort, das neben den Bemerkungen zum Adressatenkreis und zu usuellen Benutzungssituationen auch Auskunft über den erfassten phraseologischen Wortschatzausschnitt liefert, aber insgesamt beschränkt sich der Vorspann auf eine kurze Benutzeranleitung und auf die Erläuterung der Abkürzungen im Artikeltext. Die Vorspannstrukturen beider Wörterbücher sind bezüglich der Anzahl und der Funktion ihrer Teile ähnlich. Außer einem Vorwort und einem Abkürzungsverzeichnis weist nur Mantsch/Anuței/Kelp (1979) in einem gesonderten Abschnitt explizite auf Benutzungshinweise hin, wie auch nur bei diesem Nachschlagewerk ein Verzeichnis der herangezogenen

genen Lexika und Quellen vorhanden ist. Mantsch/Anuței/Kelp legen einen ausführlicheren Vorspann als Roman vor, der auch klassifikatorische Schwerpunkte berücksichtigt und mit Beispielangaben untermauert. Positiv zu bewerten ist die dargebotene Bibliographie, die man bei Roman (1993) vermisst. Die Auflistung der Fachquellen und der zitierten Werke, der Verweis auf Parallelunternehmen oder die weiterführende Bibliographie sollen nicht als Sammelsurium zufälliger Titel fungieren, sondern auf die wichtigsten Titel begrenzt sein.

Die gesichteten Lexika bringen knappe Informationen zur Kodifizierungspraxis und wenige Hinweise zur lexikographischen Bearbeitung von speziellen Phraseologismustypen und den daraus sich ergebenden Konsequenzen. Da Phraseologismen mehrgliedrig sind, besteht eine grundsätzliche Schwierigkeit darin, zu entscheiden, unter welchem Stichwort ein Phraseologismus lexikographisch erfasst werden soll. Eine benutzerfreundliche Makrostruktur geht von formalen Kriterien aus, die bei der Kodifizierung von Phraseologismen auch strikt eingehalten werden sollten. In der Einleitung sollte ein präzises Anordnungsschema ausführlich vorgeführt und erklärt werden, sodass der Benutzer daraus erschließen kann, wie der jeweilige Phraseologismus im Wörterbuch zu finden ist. Die aufgezeigten Schwächen beeinflussen die Wörterbuchbenutzung in rezeptiven wie in produktiven Nachschlagesituationen. Sie sind auch durch die Spezifik phraseologischer Erscheinungen (u.a. semantisch-pragmatische Komplexität und polylexikalischer Charakter) erklärbar.

In den rumänischen Phraseolexika sind die speziellen Prinzipien der lexikographischen Erfassung des phraseologischen Inventars und damit auch die praktischen Bedürfnisse der Wörterbuchbenutzer ungenügend berücksichtigt. Die Konzeption der untersuchten Wörterbücher ist im Hinblick auf die lexikographische Makro- und Mikrostruktur zu wenig auf benutzerorientierte bzw. benutzungsbezogene Zielsetzungen ausgerichtet. Das bedeutet, dass die verschiedenen lexikographischen Angaben und die Wörterbuchartikel für eine Neuauflage anders strukturiert und die lexikographischen Informationen hinsichtlich der tatsächlichen und potenziellen Benutzungssituationen neu konzipiert und formuliert werden müssen. Es müssen nicht nur die Umtexte ausgeweitet und inhaltlich dem aktuellen Stand der Forschung angepasst werden, sondern auch die Auswahl, die Anordnung und lexikographische Bearbeitung der Phraseologismen in den herkömmlichen rumänischen phraseologischen Wörterbüchern müsste hinterfragt werden.⁵ Einige der hier aufgezeigten Versäumnisse hinsichtlich der lexikographischen Konzeption und Ausgestaltung der Wörterbuchartikel hängen *weniger* mit den lexikographischen Entscheidungen der Wörterbuchautoren zusammen, sondern sind in einem verlegerischen Kalkül begründet.

Für Neuauflagen oder für ausstehende Nachschlagewerke zum Sprachenpaar Deutsch und Rumänisch müssen die hier angesprochenen Mängel behoben werden, wobei eine stärkere Berücksichtigung der Benutzerbelange und der in der Forschung postulierten Anforderungen an die bilinguale phraseologische Praxis mit Deutsch zu erwarten ist. Für die praktische bilinguale

Phraseographie der Sprachen Deutsch und Rumänisch ist daher die Ausarbeitung eines modernen phraseologischen Wörterbuchs, das zuverlässiger und benutzerfreundlicher als bisherige Werke ist, *dringend* zu fördern.

2.2 Zur Konzeption eines neuen zweisprachigen phraseologischen Wörterbuchs Deutsch–Rumänisch

Das vorgeformte Sprachgut umfasst unterschiedlich gestaltete Einheiten und ist folglich sehr heterogen. Es nimmt eine Sonderstellung im Sprachsystem und -gebrauch ein, insofern es sich durch bestimmte inhaltliche, formale und pragmatische Besonderheiten von den übrigen Inventareinheiten der Sprache abhebt. Ausgewählte Typen von Phraseologismen sind für bestimmte Kommunikationsbereiche und Textsorten konstitutiv, was die Berücksichtigung ihres semantisch-pragmatischen Anwendungspotenzials bedingt und dessen aktive Aneignung und lexikographische Erfassung erheblich erschwert. Das rege Interesse an Fragestellungen der kontrastiven Phraseologie — nicht nur im binnendeutschen Sprach- und Kulturraum, sondern auch andernorts — ist auch in der Auseinandersetzung mit Problemstellungen zur Phraseographie verschiedener Sprachen zu suchen. Als Resultat zahlreicher theoretischer wie auch praktischer Auseinandersetzungen mit Fragestellungen aus dem Bereich Phraseologismen und Lexikographie ist u.a. die Notwendigkeit einer Neuorientierung gefordert worden, wobei nach Burger (1992) zweierlei abgewogen werden muss: (a) die gängige Wörterbuchpraxis und (b) die Erkenntnisse der Phraseologie-Forschung, da von hier aus konkrete Forderungen an die lexikographische Praxis gestellt wurden, die zu einer verbesserten Qualität der Wörterbücher beitragen sollen. Schemann (1989: 1026) erachtet es als die „vielleicht schwierigste Aufgabe“ des Verfassers eines phraseologischen Wörterbuchs, dem Benutzer all jene grammatischen Informationen zu liefern, die er benötigt, um einen phraseologischen Ausdruck in jedem möglichen Kontext korrekt zu gebrauchen.

Die moderne Metalexikographie versucht, durch die Bereitstellung eines geeigneten Methodenapparats und die Ausarbeitung neuer Kodifizierungsverfahren den Benutzerbelangen und der adressatengerechten Ausrichtung der Wörterbücher zu dienen, um künftige Wörterbücher oder neue Wörterbuchauflagen zu verbessern. Dabei ist darauf verwiesen worden, dass eine verbesserte Adressatenorientiertheit nicht nur durch einen sorgfältigen mikrostrukturellen Aufbau gesichert werden kann, sondern auch durch eine einheitliche lexikographische Bearbeitung.

Voraussetzung für die Zusammenstellung eines benutzerfreundlichen Nachschlagewerks sind theoretische und praktische Überlegungen zum Benutzerprofil, zu Auswahl- und Darstellungskriterien des kodifizierten Sprachmaterials, eine sorgfältige Erfassung der Gebrauchsbeschränkungen im aktuellen Sprachgebrauch wie auch eine verbesserte Äquivalentdarstellung. Für Neuauflagen vorhandener Wörterbücher zum Sprachenpaar Deutsch und

Rumänisch oder für die Konzeption geplanter ist die Ausarbeitung einer adäquaten Beschreibungsweise zwingend und eine eingehendere empirische Fundierung⁶ notwendig, um eine angemessene lexikographische Darstellung phraseologischer Einheiten zu gewährleisten. Aus der Perspektive dieses Sprachenpaares gestaltet sich dieses Vorhaben als schwierig, insofern eine Metasprache erst ausgearbeitet werden muss, da kein differenziertes und auf lexikographische Zwecke ausgerichtetes metalexikographisches Beschreibungsvokabular im Rumänischen bereitsteht. Eine Neubearbeitung ist auch dadurch erschwert, dass die Eintragungen relevante Informationen zur Rezeption und/oder Produktion der kodifizierten Phraseologismen enthalten müssen, die eine situations- und textsortenangemessene wie partnergerechte Verwendung des kodifizierten Sprachmaterials gewährleisten sollen, um eine aktive Aneignung des phraseologischen Inventars zu unterstützen.⁷

Für den Benutzer eines bilingualen phraseologischen Wörterbuchs sind bestimmte Informationen relevant: (a) die Lemmazuordnung; (b) die Nennformgestaltung; (c) die Vorkommensbedingungen; (d) die Äquivalenzdarstellung. Im Hinblick auf eine verbesserte Beschreibungspraxis des anvisierten Sprachenpaares sollte danach getrachtet werden, nicht nur zielsprachliche Äquivalente zu erfassen, sondern auch Charakteristika der kodifizierten Einheiten. Daher müssten Erweiterungen der Einträge um bestimmte Artikelpositionen eingeplant werden. Von grundlegender Bedeutung sind hierbei die Angaben zur sprachlichen Umgebung und die Kennzeichnung der stilistisch-pragmatischen Besonderheiten. Über das Verstehen einer festen Wortverbindung hinaus muss die lexikographische Beschreibung auch den korrekten Gebrauch der kodifizierten Phraseologismen bei der Textproduktion sicherstellen. Das Wörterbuchkonzept, das hier in Grundzügen vorgestellt werden soll, fasst idiomatisierte Phraseologismen ins Auge und richtet sich an einen breiten Adressatenkreis (z.B. Studierende, Lehrkräfte, Lehrbuchautoren, Übersetzer). Dabei wird von der Prämisse ausgegangen, dass das Nachschlagewerk sowohl für die Textrezeption als auch indirekt für die Textproduktion von rumänischen Deutschlernern sowie sekundär für die Textrezeption von deutschsprachigen Rumänischlernern genutzt werden kann. Bei der Konzeption eines Beschreibungsmodells mit Deutsch als Ausgangssprache für rumänische Benutzer wird eine Auffassung von Phraseologie im engeren Sinne vertreten, um eine möglichst einheitliche und umfassende Beschreibung der kodifizierten deutschen Phraseologismen mit ihren rumänischen Entsprechungen zu gewährleisten. Sie steht auch in Einklang mit den Empfehlungen der bilingualen Metalexikographie.⁸

Für die Ausarbeitung eines modernen und benutzerfreundlichen Beschreibungsmodells wurden die lexikographischen Standards, die die metalexikographische Forschung hinsichtlich Adressatengruppen und deren Nutzungsinteressen formuliert hat, berücksichtigt wie auch einige Verbesserungsvorschläge, die eine Erweiterung der Einträge um bestimmte Artikelpositionen gefordert haben. Für diesen speziellen Wörterbuchttyp hat Cheon (1998) Über-

legungen zum Ausbau der Mikrostruktur formuliert, während Mudersbach (1998) eine Beschreibung aufgrund eines universalen pragmatischen Modells vorschlägt, das das enthalten sollte „was ein Nichtmuttersprachler wissen muss, um ein Phrasem in einer Situation richtig gebrauchen zu können.“ Hier muss betont werden, dass der Einbezug von Phraseologen bei der Revidierung von Wörterbüchern notwendig ist, wie auch die Berücksichtigung der Ergebnisse der theoretischen und angewandten allgemeinen und kontrastiven Phraseologie, um eine optimale Lemmatisierung und Anordnung des kodifizierten phraseologischen Bestandes gewährleisten zu können.⁹

Die vorhandenen Nachschlagewerke zum Sprachenpaar Deutsch und Rumänisch, die in Rumänien erarbeitet worden sind, sind veraltet. Bei dem damaligen Stand der Phraseologie/Lexikographie — insbesondere der zweisprachigen — ist die mangelhafte lexikographische Bearbeitung phraseologischer Einheiten nachvollziehbar. Bei der Ausarbeitung eines Verbesserungsvorschlags¹⁰ zur phraseographischen Praxis des Sprachenpaares Deutsch und Rumänisch waren folgende Überlegungen ausschlaggebend: (1) das Erarbeiten eines einheitlichen Beschreibungsmodus; (2) die Präsentation und Gestaltung der Information, die ein semantisches-pragmatisches Verständnis gewährleisten sollte, im Hinblick auf die Erleichterung der Textproduktion und -rezeption; (3) die Ausarbeitung einer Metasprache, welche die Kriterien der Darstellung des phraseologischen Materials verdeutlichen und die Besonderheiten des aufgenommenen Inventars angemessen erfassen sollte. Bei der Erarbeitung eines bilingualen phraseologischen Wörterbuchs muss in Erwägung gezogen werden, dass für den Benutzer bestimmte Informationen relevant sind. Deshalb sollte der Wörterbuchvorspann — über das Verzeichnis der verwendeten Zeichen, Abkürzungen usw. hinaus — Angaben zum kodifizierten phraseologischen Inventar mit Hinweisen zum terminologischen und definitiven Ansatz, Informationen zur Selektion und Präsentation phraseologischer Einheiten, zur Kennzeichnung der Gebrauchsweisen und Angaben zur Äquivalenz-erfassung und -gestaltung enthalten. Um die aktive Aneignung von Phraseologismen zu unterstützen, wird der alphabetische Wörterbuchteil ergänzt durch einen Indexteil, der einen raschen Zugriff auf den gesuchten Phraseologismus ermöglicht und dem Wörterbuchbenutzer verdeutlicht, wie er das Nachschlagewerk konsultieren kann und soll. Dieser Indexteil umfasst folgende Teile: (a) ein alphabetisches Register, das alle im Wörterbuch verzeichneten und beschriebenen Phraseologismen in alphabetischer Reihenfolge aufführt und die entsprechende Seitenanzahl im Wörterbuch angibt. Das alphabetische Register ist praktisch für jede Benutzungssituation erforderlich und dient speziell den Zwecken des Nachschlagens bei der Textrezeption. Dieser Wörterbuchteil ermöglicht einen schnellen Zugriff auf unbekannte Phraseologismen. Er bietet eine Lernmöglichkeit im Sinne der Kontrolle und Wiederholung; (b) ein Register zielsprachlicher Äquivalente, falls der Benutzer eine dem rumänischen Phraseologismus entsprechende deutsche Wendung sucht; (3) ein Register alphabetischer Leitbegriffe ermöglicht über einen onomasiologischen Ansatz

den Zugang zu den rumänischen Phraseologismen. Dieser Wörterbuchteil kann mit bestimmten Einschränkungen die Funktion eines aktiven Wörterbuchs übernehmen. Konzeptionell würde ein moderneres phraseologisches Wörterbuch Deutsch–Rumänisch aus einer alphabetischen Darstellung des erfassten phraseologischen Wortschatzes und einem Indexteil bestehen.

Der Verbesserungsvorschlag wird daher folgende phraseographisch vorbelastete Problembereiche anvisieren: die Umtextgestaltung, die Nennformgestaltung, die Darbietung der Restriktionen und die Markierungspraxis. Die anvisierten Neuerungen der phraseographischen Praxis betreffen folglich: (1) die Orientierung im Wörterbuch; (2) die lexikographische Beschreibungspraxis, d.h. die lexikographische Erfassung und Darstellung phraseologischer Spezifik (Nennformgestaltung; Präsentation phraseologischer Variation; Berücksichtigung der externen Valenz; Erfassung spezifischer Gebrauchsmuster samt Darbietung der formal-semantischen und pragmatischen Restriktionen; verbesserte Markierungspraxis); (3) die primäre Berücksichtigung der Belange von Deutschlernenden.

Bei der Nennformgestaltung müssen dem Wörterbuchbenutzer Angaben geliefert werden, um die entsprechende phraseologische Einheit auch aktiv verwenden zu können, das bedeutet, dass morphosyntaktische Restriktionen kenntlich gemacht werden müssen.¹¹ Mit der Angabe der Nennform werden dem Benutzer die wichtigsten Informationen zur Einbettung des Phraseologismus im Satz geliefert. Um auch der Lernfunktion gerecht zu werden, muss die Nennform gewisse Voraussetzungen erfüllen. Dies bedingt, dass die Nennform Informationen über die Kontextpartner, d.h. über die obligatorischen und fakultativen Valenzen der Phraseologismen liefern soll. Die Nennform sollte in einer neutralen, nicht aktualisierten Form aufgeführt werden, wenn keine morphosyntaktischen Restriktionen vorliegen. Um vollständig zu sein, muss die Nennform mit den entsprechenden externen Valenzen ausformuliert werden. Das Indefinitpronomen *etwas* kann in der Nennform als metasprachliche Angabe fungieren, um die Valenz anzuzeigen, wobei *etwas* immer mit einem variablen Element in der Kommunikation zu ersetzen wären, zum anderen geht es um eine invariable Komponente der phraseologischen Einheit. Die Abkürzung des Indefinitpronomens *etw.* wird daher zur Markierung der unbelebten externen Valenz eingesetzt, sodass der Unterschied zwischen wendungsinternem Indefinitpronomen (z.B. *nach etwas aussehen*) und wendungsexterner Valenz (z.B. *etw. mit der Muttermilch einsaugen*) eindeutig ist. Elemente des Typs *jmdm.*, *jmdn.*, *jmds.*, *jmd.*, *etw.* gehören zu den obligatorischen Komponenten des Phraseologismus. Damit werden sowohl semantische Kriterien (u.a. Personen- und Sachbezug) als auch syntaktische Kriterien bei der Einbettung fester Wortverbindungen in den Ko(n)text berücksichtigt. Für Phraseologismen, die ein obligatorisches Subjekt als Komponente enthalten, hält es die lexikographische Forschung aus der Sicht des Fremdsprachenlerner angemessen, bei der lexikographischen Darstellung die Präsensform anzuführen. Dadurch erscheint die für den Lerner wichtige Ergänzung in Subjektposition in der

Nennform. Für die Frage nach den fakultativen und obligatorischen Komponenten des Phraseologismus, die in die Nennform aufgenommen werden sollen, können für eine vollständige und korrekte Kodifikation Recherchen der modernen elektronischen Textkorpora wie auch das Sammeln von Belegen aus modernen Quellen (z.B. aus der aktuellen Presselandschaft) hilfreich sein.

Bei der lexikographischen Darstellung der Phraseologismen empfiehlt sich, den sprachlichen und den situativen Kontext (pragmatische Angaben/gebrauchssemantische Beschreibung) zu berücksichtigen. Des Weiteren ist davon auszugehen, dass man den semantischen Inhalt eines Phraseologismus nicht ohne stilistische und pragmatische Informationen erfassen kann. Daher sind Gebrauchsmuster in einem phraseologischen Wörterbuch notwendig. Burger (1992: 49) schlägt standardisierte Formulierungen für die Erfassung pragmatischer Aspekte in der lexikographischen Praxis vor, die im Wörterbuchvorwort zu erläutern sind.

Die metalexikographische Fachliteratur hat eine Reihe von Überlegungen und Empfehlungen zu einer konsistenteren Markierungspraxis angestellt bzw. ein offenes Beschreibungsinventar vorgeschlagen.¹² Die Fachliteratur hat empfohlen, die Bedeutungserklärung durch pragmatische Hinweise zu ergänzen, um die für den Nichtmuttersprachler wichtigen Aspekte des Phraseologismgebrauchs (Situationsspezifika, Stilmarker, Sprecherintention, emotionaler Gehalt, Geltungsbereich) erfassen zu können. Vgl. hierzu folgende Angaben: Charakterisierung der Gebrauchssituation (\pm mündlich/schriftlich); Sprechereinstellung (z.B. euphemistisch/verhüllend, scherzhaft/humorvoll, pejorativ/ironisch/spöttisch/geringschätzig, vorwurfsvoll/verächtlich, grotesk/derb, abwertend, feierlich, beschönigend); Emotionalitätsgrad (z.B. \pm übertrieben/emotional); Stilschichtmarkierung (z.B. gehoben, umgangssprachlich, salopp); Charakterisierung der Sprecher-Hörer-Konstellation (\pm Vertrautheit, \pm formell, \pm hierarchisch, \pm symmetrisch).

Pragmatische Erkenntnisse haben zwar einen „breiten Eingang“ in die Theorie der Phraseographie gefunden, „es mangelt jedoch an der konsequenten praktischen Umsetzung der gewonnenen Postulate“ (Filatkina 2007: 152) in der allgemeinen und phraseologischen Wörterbuchpraxis. Angesichts der Tatsache, dass in den herkömmlichen Wörterbüchern die Semantik der Phraseologismen oft — und fast ausschließlich — durch deren denotative Bedeutung erfasst wird, hat sich in der metalexikographischen Diskussion verstärkt die Forderung nach einer sprachhandlungstheoretisch orientierten Beschreibung durchgesetzt. Mit Recht wurde gefordert, dass Angaben zu den Sprecher-/Schreibereinstellungen und den kommunikativen Funktionen von Phraseologismen in Wörterbüchern — unabhängig von ihrem Typ — unbedingt notwendig sind, um die hinsichtlich der lexikographischen Kodifizierung von Phraseologismen immer wieder beklagte Diskrepanz zwischen Sprachwirklichkeit und Wörterbuchpraxis zu mindern.¹³

Bei der Suche nach Äquivalenten sollte vom Begriff der funktionalen Äquivalenz ausgegangen werden. Dies impliziert, dass die ausgangssprachlichen Phraseologismen mit denen der Zielsprache in der gleichen kommunika-

tiven Situation gebraucht werden können. Besteht der Äquivalentteil aus mehreren Ausdrücken, so gilt die Regelung, dass derjenige Ausdruck, der der deutschen Ausgangseinheit hinsichtlich der Bedeutung und Gebrauchsmöglichkeiten am ehesten entspricht, an erster Stelle erscheint. Wenn es zu einem deutschen Phraseologismus keinen adäquaten Ausdruck gibt, müssen im Äquivalentteil ergänzend Erläuterungen geboten werden. Diese erscheinen in Petit und stehen nach dem Äquivalent. Da ein Fremdsprachler nicht immer entscheiden kann, welche Konstituente bei der Lemmaansetzung in Frage kommt, wird im Wörterbuch mit Verweisangaben gearbeitet. Die Pfeile können nach der aktuellen Suchrichtung nach oben bzw. nach unten zeigen. Aufgrund vielfacher Fehler — wie auch aus Platzgründen — wäre von einer Mehrfachlemmatisierung abzuraten.

Im Folgenden wird die konzeptionelle Ausrichtung der Wörterbuchartikel für ein neues phraseologisches Wörterbuch Deutsch–Rumänisch vorgestellt. Die Einführung pragmatischer Parameter, die für das richtige Verständnis und den korrekten Gebrauch eines Phraseologismus ausschlaggebend sind, d.h. eine erweiterte Mikrostruktur, würde dem Benutzer Informationen bieten, die in den angenommenen Benutzungssituationen relevant sind. Zusatzangaben wie präzisierende und pragmatische Erläuterungen sichern eine differenziertere Beschreibung der Besonderheiten und der Gebrauchsbedingungen. Die verbesserte Kennzeichnung der Verwendungsweisen und eine differenziertere Markierungspraxis sind für die Qualität des mikrostrukturellen Profils des Nachschlagewerks ausschlaggebend.

Das vorgeschlagene Artikelprofil zur Verbesserung der rumänischen bilingualen Wörterbuchpraxis weist daher folgende lexikographische Parameter auf:

- Nennformangabe / rum. forma uzuală;
- Bedeutungserläuterung / rum. semnificație;
- Angaben zur sprachlichen Umgebung / rum. caracterizare [externe Valenz / rum. valența externă; z.B. *jmd.*; freie Angabe / rum. element facultativ; Negationsform(en) / forma de negare; z.B. *nicht*];
- Variationsangaben / rum. forme de varietate;
- pragmatische Angaben / rum. profil pragmatic [Charakterisierung der Gebrauchssituation; ± mündlich / schriftlich / rum. indicații contextual-pragmatice privind specificul situației de comunicare; ± oral / scris; Einstellung- / Bewertungsindikatoren / rum. evaluaire; Emotionalitätsgrad / rum. gradul de emoționalitate; Stilmarker / rum. nivelul stilistic; Charakterisierung der Sprecher-Hörer-Konstellation; z.B. ± Vertrautheit, ± formell, ± hierarchisch, ± symmetrisch / rum. relația vorbitor-receptor];
- kommunikativer Wert / rum. valoarea comunicativă;
- phraseosemantische Einordnung / rum. câmp frazeosemantic;
- Äquivalenzangabe(n) / rum. echivalent(e): *fig.* (rum. *figurat*; dt. bildhaft; übertragen); Stilmarker / rum. nivelul stilistic;
- Verweis / rum. trimitere la.

Dieser Modellvorschlag ermöglicht durch die eingebrachten Angaben Ungenauigkeiten oder Fehlinterpretationen beim Benutzer weitgehend auszuschließen. Wie dieses Beschreibungsmodell, das eine verbesserte lexikographische Darstellung des aufgenommenen Sprachmaterials in einem neuen phraseologischen Wörterbuch Deutsch–Rumänisch anvisiert, in die lexikographische Praxis umgesetzt wird, soll im Folgenden an einem Probeartikel aufgezeigt werden. Dabei wird exemplarisch auf die Bearbeitung des Phraseologismus *vom/aus dem Regen in die Traufe kommen* eingegangen, um auch zu zeigen, wie einige Mängel der herkömmlichen rumänischen bilingualen Wörterbuchpraxis mit Deutsch (vgl. 2.1) behoben werden können. Die Eintragungen zu diesem Phraseologismus in den gängigen Wörterbüchern zu diesem Sprachenpaar sind äußerst knapp: *vom/aus dem Regen in die Traufe kommen* — a da din lac în puț (Roman 1993: 203) und *a cădea/a da/a sări din lac în puț* — aus dem/vom Regen in die Traufe kommen (Mantsch/Anuței/Kelp 1979: 289). Das verbesserte Artikelprofil für ein neues phraseologisches Wörterbuch bietet dem rumänischen Wörterbuchbenutzer unter dem Lemma *Regen* folgende Informationen:

jmd. kommt vom/aus dem Regen in die Traufe: a da de un rău mai mare în încercarea de a scăpa de un alt rău; a ajunge dintr-o situație critică în alta și mai critică; ± comunicare orală / + nemulțumire / + emoționalitate; *fam.*; + familiaritate / ± formal / ierarhic; DEZILUZIE / RESEMNARE / SUPĂRARE; câmp frazeosemantic: NEPLĂCERE / DECEPTIE / DEZAMĂGIRE; [fig]; *fam.* a cădea / a da / a sări din lac în puț; din rău în mai rău; a căuta pe dracul și a găsi pe tată-său; trimitere la: Traufe↑

Dieser Modellvorschlag bietet nicht nur zielsprachliche rumänische Äquivalente an, sondern erfasst auch Besonderheiten der kodifizierten deutschen phraseologischen Einheiten. Die Benutzerfreundlichkeit wird nicht nur durch den Einbezug pragmatischer Parameter bei der Gebrauchsmusterbeschreibung gewährleistet, sondern auch durch den Zugriff auf eine flexiblere, differenziertere Markierungsweise und die Zuordnung ausgangssprachlicher Fügungen zu einem phraseosemantischen Feld, womit die Suche des Benutzers nach dem geeigneten funktionalen Äquivalent vereinfacht wird.¹⁴

Die Anzahl der Angaben, die in einem Wörterbuch vorkommen, hängt von verschiedenen Faktoren (Aktiv-Passiv-Prinzip, Wörterbuchumfang, Benutzerbedürfnisse) ab. Dabei ist zu beachten, dass je stärker die aktive Funktion berücksichtigt wird, desto vielfältigere Informationen zum Kontext und zum Gebrauch geboten werden müssen. Über Äquivalenzangaben hinaus sollten in einem bilingualen aktiven Wörterbuch pragmatische Zusatzinformationen wie auch Informationen zur syntaktischen Einbettung geliefert werden.

Dieser Alternativvorschlag stellt bestimmte Anforderungen an seine Benutzer; er ist daher eher für einen Benutzer konzipiert worden, der sich der

Mühe unterzieht, einen kompletten Wörterbuchausschnitt durchzuarbeiten und dabei auch die Hinweise zur Benutzung in der Einleitung nachzulesen. Die empirische Benutzungsforschung hat gezeigt, dass die Wörterbuchbenutzer die in den Nachschlagewerken dargebotenen Informationen nicht immer verstehen oder angemessen verwerten können.¹⁵ Gleichfalls muss darauf hingewiesen werden, dass nicht kühne Entwürfe, sondern zuverlässige Informationen für die Konzeption neuer Lexika bzw. bei der Planung von Neuauflagen ausschlaggebend sind. Man kann auch Wiegand (1998a: 193) nur zustimmen, wenn er feststellt:

Es ist sicherlich kein Verdienst, lexikographische Methoden und die aus deren Anwendung resultierende Praxis zu kritisieren und sie z.B. als unangemessen zu beurteilen [...]; es reicht m.E. auch nicht aus, lediglich mehr oder weniger allgemeine Vorschläge zu machen oder weitgehende Forderungen zu formulieren, ohne Rücksicht darauf, ob sie in der Wörterbucharbeit praktikabel sind. Wer Lexikographen und deren Produkte kritisiert, sollte — wenigstens exemplarisch — eine alternative Praxis *vormachen* [Hervorhebung im Original], und zwar jene, die er selbst für die angemessenere hält.

Die Erstellung von aktiven phraseologischen Wörterbüchern, die mehr Informationen über eine beschränkte Auswahl von Phraseologismen bieten, ist eine zentrale Aufgabe der modernen Phraseographie.¹⁶ Für Neuauflagen vorhandener Wörterbücher oder für die Konzeption neuer Werke muss aus der Perspektive des Nichtmuttersprachlers und seiner Benutzerbedürfnisse für die rumänische phraseographische Praxis mit Deutsch eine verbesserte, benutzerfreundliche linguistische und metalexikographische Behandlung der Phraseologismen gesichert werden. Eine verbesserte Adressatenorientiertheit kann durch einen mikrostrukturellen Ausbau und die Ausarbeitung einer speziell auf die Bedürfnisse der Benutzer zugeschnittenen Mikrostruktur gesichert werden. Nicht nur durch eine verbesserte Konzeption der Artikelkörper, die dem modernen Stand der phraseologischen und phraseographischen Forschung entsprechen, sondern auch durch den Ausbau und die Umgestaltung der Umtexte wird die Benutzerfreundlichkeit zusätzlich erhöht. Dabei ist das Heranziehen zuverlässiger Vorgängerwerke, Fachquellen und Textkorpora grundlegend wie auch die Ausarbeitung einer für phraseologische Inventareinheiten adäquaten Beschreibungsweise.

Mit der Behebung der hier aufgezeigten Versäumnisse hinsichtlich der lexikographischen Konzeption und Ausgestaltung der Wörterbücher würde nicht nur ein Beitrag zur aktiven Phraseographie geleistet werden. Die Ausarbeitung eines wissenschaftlich und empirisch abgesicherten lexikographischen Modellvorschlags, der sich konzeptionell auf die Ergebnisse der theoretischen und praktischen Phraseologie/Phraseographie stützt, um verlässliche Auskunft bei der Produktion/Rezeption von Texten zu bieten, würde auch die Erforschung der Phraseologie dieser Sprachen im internationalen Vergleich erleichtern.

3. Schlussbemerkungen

Vorliegender Beitrag hat Fragen der Erarbeitung von zweisprachigen Wörterbüchern sowie lexikographische Standards diskutiert und deren Berücksichtigung/Verwirklichung in der rumänischen praktischen Lexikographie ausdrücklich gefordert, um dem Benutzer ein brauchbares Hilfsmittel an die Hand zu geben oder dem Übersetzer zuverlässigere Äquivalenzzuordnungen zu ermöglichen. Um der zweisprachigen lexikographischen Praxis wertvolle Anregung zu vermitteln, wurde die Einbindung der aus (sprach)praktischer und phraseologischer Sicht relevanten Angaben zu den Gebrauchsbeschränkungen und -bedingungen an einem Modellvorschlag aufgezeigt.

Der auf die herkömmliche rumänische bilinguale lexikographische Praxis mit Deutsch ausgerichtete Beitrag hat gezeigt, welche Domänen phraseographischer Relevanz anzusetzen sind, um wissenschaftlich und empirisch abgesicherte phraseologische Nachschlagewerke ausarbeiten zu können. Als Ergebnis der kritischen Durchsicht gängiger Wörterbücher zu diesem Sprachenpaar und bei der Prüfung ihrer konzeptionellen Ausrichtung lassen sich Defizite festhalten, die für die Orientierung in rezeptiven wie in produktiven Nachschlagesituationen ausschlaggebend sind. Insgesamt kann eine benutzerunfreundliche und unangemessene Präsentation des phraseologischen Materials — Inkonsequenzen bei der Festlegung der Nennform, der Handhabung lexikalischer Varianten oder bei Äquivalenzangaben — ausgemacht werden wie auch zahlreiche Verstöße gegen die lexikographische Akribie.

Angesichts lexikographischer Leistungen der bilingualen Speziallexikographie weltweit und der Brauchbarkeit vorhandener Wörterbücher zum Sprachenpaar Deutsch und Rumänisch aus heutiger Sicht ist ersichtlich, dass die phraseographische Erfassung der Sprachen Deutsch und Rumänisch ein bis dato stark vernachlässigtes Feld der praktischen Phraseographie ist, das es zu fördern gilt.

Da die gängigen rumänischen bilingualen Wörterbücher mit Deutsch einer Aktualisierung bedürfen und auf dem Wörterbuchmarkt längst vergriffen sind, ist die Entwicklung moderner bilingualer phraseologischer Nachschlagewerke mit Deutsch als Ausgangs- oder Zielsprache eine noch nachzuholende Aufgabe der rumänischen Phraseographie. Mit dem hier vorgestellten Alternativvorschlag war das Bestreben verbunden, die lexikographische Praxis anzuregen und zu verbessern, wobei auch einer adäquaten lexikographischen Darstellung phraseologischer Einheiten Rechnung getragen wurde. Eine verbesserte Erfassung der Gebrauchsweisen der Phraseologismen würde auch den Rückstand der rumänischen bilingualen Wörterbuchpraxis mit Deutsch aufholen und implizite auch eine erhöhte Qualität des Wörterbuchs und dessen Benutzerfreundlichkeit sichern.

Unter Bezugnahme auf den aktuellen Stand der rumänischen lexikographischen Praxis mit Deutsch ist ersichtlich, dass die Erstellung eines modernen Nachschlagewerks mit Deutsch als Ausgangs- oder Zielsprache eine *dringende* Aufgabe der rumänischen Phraseographie darstellt.

Anmerkungen

1. Zu allgemeinen Fragen der lexikographischen Behandlung von Phraseologismen in Wörterbüchern vgl. u.a. Barz/Henning/Korhonen (2005), Burger et al. (2007), Blanco (2009) oder Blanco et al. (2010).
2. Vgl. dazu z.B. Bergenholtz/Tarp (2003).
3. Vgl. hierzu die Ausführungen bei Schemann (z.B. 1989, 1991), Martin (2001) oder Ptashnyk (2003).
4. Sava (2010). Vgl. Mantsch/Anuței/Kelp (Rumänisch–Deutsch; 1979) und Roman (Deutsch–Rumänisch; 1993).
5. Obwohl der Wörterbuchbenutzung in der lexikographischen Diskussion ein breiter Raum eingeräumt wurde, sind Umfang und Ausgestaltung der Umtexte in der deutschsprachigen Wörterbuchlandschaft sporadisch Gegenstand metalexikographischer Reflexionen gewesen. Die kritischen Anmerkungen zur phraseographischen Wörterbuchpraxis, die in der Forschung anzutreffen sind, betreffen vornehmlich Probleme der Lemmatisierung, Nennformgestaltung und Gebrauchsmarkierung phraseologischer Einheiten.
6. Hierfür sind u.a. die Textkorpora des Instituts für deutsche Sprache in Mannheim zu empfehlen.
7. Fragen der lexikographischen Bearbeitung von Phraseologismen aus bilingualer Perspektive sind Diskussionsgegenstand u.a. bei Filipenko (2002, Deutsch–Russisch), Martin (2001, Deutsch–Spanisch), Petelenz (2001, Deutsch–Polnisch), Cheon (1998, Deutsch–Koreanisch). Manche Autoren erläutern lexikographische Standards der bilingualen Phraseographie am Material phraseologischer Untergruppen (Farø 2002; Deutsch–Dänisch/Somatismen) oder rücken linguistisch–lexikographisch relevante Kodifizierungsbereiche in den Mittelpunkt der Betrachtungen (z.B. pragmatische Angaben in einem zweisprachigen Lernerwörterbuch Deutsch–Bulgarisch, Drumeva, 2001). Mit der Spezifik der bilingualen Wörterbuchpraxis befassen sich auch die Beiträge von Petkov (2001) oder Korhonen (z.B. 2004), während eine eingehende Besprechung zur Theorie und Praxis der Phraseographie Schemann (1991) leistet. Der Sammelband von Blanco et al. (2010) dokumentiert den aktuellen Forschungsstand und gibt wichtige Impulse für die Weiterentwicklung der ein- und zweisprachigen Phraseographie des 21. Jahrhunderts.
8. Die neuere Phraseologieforschung fordert die Einbeziehung peripherer phraseologischer Bereiche (z.B. Funktionsverbgefüge, Routineformeln) und befürwortet eine weite Phraseologieauffassung. Für die Lemmaauswahl sollten Textkorpora, allgemeine und spezielle Wörterbücher herangezogen werden aber auch verschiedene Quellen nichtlexikographischer Art wie Lehrbücher.
9. Vgl. dazu Földes (z.B. 1996) oder Ptashnyk (2003).
10. Dazu auch Sava (2015).
11. Vgl. hierzu auch Martin (2001: 142).
12. Näheres dazu bei Hausmann (1989: 649).
13. Vgl. hierzu u.a. Blanco et al. (2010).
14. Dazu u.a. Dobrovols'kij (1997).
15. Zur Benutzung von Wörterbüchern vgl. Wiegand (1998) oder Engelberg/Lemnitzer (2009).
16. Als innovativer Ansatz für die Erforschung fester Wortverbindungen gilt die Methode der statistischen Kookkurrenzanalyse, die in den 1980er-Jahren am Institut für deutsche Sprache

in Mannheim entwickelt wurde. Steyer präsentiert in mehreren Arbeiten (z.B. 2000 und 2004) das linguistische Modell, die Korpusmethodik und die lexikographischen Perspektiven dieses Ansatzes.

Das Acknowledgement

Dieser Artikel entstand im Rahmen des von der Lucian-Blaga-Universität Sibiu/Hermannstadt geförderten Projekts LBUS-IRG-2017-03 (Laufzeit: 2017–2019).

Literatur

- Barz, Irmhild, Henning Bergenholtz und Jarmo Korhonen (Hrsg.).** 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. Finnische Beiträge zur Germanistik 14. Frankfurt [etc.]: Peter Lang.
- Blanco, Carmen Mellado (Hrsg.).** 2009. *Theorie und Praxis der idiomatischen Wörterbücher*. Tübingen: Max Niemeyer.
- Blanco, Carmen Mellado, Patricia Buján, Claudia Herrero, Nely Iglesias und Ana Mansilla (Hrsg.).** 2010. *La fraseografía del S. XXI. Nuevas propuestas para el español y el alemán*. Romanistik 6. Berlin: Frank & Timme.
- Bergenholtz, Henning und Sven Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes. Journal of Linguistics* 31: 171-196.
- Burger, Harald.** 1992. Phraseologie im Wörterbuch. Überlegungen aus germanistischer Perspektive. Eismann, Wolfgang und Jürgen Petermann (Hrsg.). 1992. *Studia phraseologica et alia. Festschrift für J. Matešić zum 65. Geburtstag*: 33-51. Specimina Philologiae Slavicae. Supplementband 31. München: Sagner.
- Burger, Harald, Dmitrij Dobrovol'skij, Peter Kühn und Neal R. Norrick (Hrsg.).** 2007. *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* [...]. Berlin/New York: Walter de Gruyter.
- Cheon, Mi-Ae.** 1998. Zur Konzeption eines phraseologischen Wörterbuchs für den Fremdsprachler: Am Beispiel Deutsch-Koreanisch. *Lexicographica. Series Maior* 89. Tübingen: Max Niemeyer.
- Dobrovol'skij, Dmitrij.** 1997. Prototypentheoretische Ansätze bei der Beschreibung der Idiomatik. Barz, Irmhild und Marianne Schröder (Hrsg.). 1997. *Nominationsforschung im Deutschen. Festschrift für W. Fleischer zum 75. Geburtstag*: 157-165. Frankfurt [etc.]: Peter Lang.
- Drumeva, Stanislava.** 2001. Pragmatische Angaben in einem zweisprachigen Lernerwörterbuch mit Deutsch und Bulgarisch. Iгла, Birgit, Pavel Petkov und Herbert Ernst Wiegand (Hrsg.). 2001: 21-40.
- Engelberg, Stefan und Lothar Lemnitzer.** 2009. *Lexikographie und Wörterbuchbenutzung*. 4. Aufl. Tübingen: Stauffenburg.
- Farø, Ken.** 2002. Somatismen als Problem der dänischen und deutschen Lexikographie. Gottlieb, Henrik, Jens Erik Mogensen und Arne Zettersten (Hrsg.). 2002. *Symposium on Lexicography X. Proceedings of the Tenth International Symposium on Lexicography May 4–6, 2000, at the University of Copenhagen*: 107-124. *Lexicographica. Series Maior* 109. Tübingen: M. Niemeyer.

- Filatkina, Natalia.** 2007. Pragmatische Beschreibungsansätze. Burger, Harald, Dmitrij Dobrovol'skij, Peter Kühn und Neal R. Norrick (Hrsg.). 2007: 132-158.
- Filipenko, Tat'jana V.** 2002. Beschreibung der Idiome in einem zweisprachigen Idiomatik-Wörterbuch (Deutsch–Russisch). Vollstedt, Marina (Hrsg.). 2002. *Das Wort*. Germanistisches Jahrbuch GUS: 43-62.
- Földes, Csaba.** 1996. Gesucht ... und gefunden? Idiomlexika und Deutsch als Fremdsprache. *Fremdsprache Deutsch* 15: 64-67.
- Hausmann, Franz Josef.** 1989. Die Markierung im allgemeinen einsprachigen Wörterbuch: eine Übersicht. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.). 1989: 649-657.
- Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.).** 1989–1991. *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie* [...]. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1–5.3. Berlin/New York: Walter de Gruyter.
- Igla, 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*. Germanistische Linguistik 161–162/2001. Hildesheim/Zürich/New York: Georg Olms.
- Korhonen, Jarmo.** 2004. Duden 11 — Nutzungserfahrungen aus der DaF-Perspektive. Steyer, Kathrin (Hrsg.). 2004. *Wortverbindungen — mehr oder weniger fest*. Jahrbuch des Instituts für Deutsche Sprache 2003: 360-392. Berlin: Walter de Gruyter.
- Mantsch, Heinrich, Mihai Anuței und Helmut Kelp.** 1979. *Dicționar frazeologic român–german*. București: Științifică și Enciclopedică.
- Martin, Luis.** 2001. *Phraseologie im zweisprachigen Wörterbuch Deutsch–Spanisch*. Hispano-Americana. Geschichte, Sprache, Literatur 28. Frankfurt [etc.]: Peter Lang.
- Mudersbach, Klaus.** 1998. Ein Vorschlag zur Beschreibung von Phrasemen auf der Basis eines universalen pragmatischen Modells. Wiegand, Herbert Ernst (Hrsg.). 1998. *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*. Lexicographica. Series Maior 84: 319-347. Tübingen: Max Niemeyer.
- Petelenz, Krzysztof.** 2001. Standardisierung der Lexikoneinträge für ein neues großes deutsch-polnisches und polnisch-deutsches Wörterbuch. Wiegand, Herbert Ernst (Hrsg.). 2001. *Studien zur zweisprachigen Lexikographie mit Deutsch VII*. Germanistische Linguistik 159–160/2001. Hildesheim/Zürich/New York: Georg Olms.
- Petkov, Pavel.** 2001. Zum Problem der Äquivalenzbeziehung und der lexikographischen Lücke in zweisprachigen Wörterbüchern. Igla, Birgit, Pavel Petkov und Herbert Ernst Wiegand (Hrsg.). 2001: 73-81.
- Ptashnyk, Stefaniya.** 2003. Wie hart ist die phraseographische Nuss? Überlegungen zu phraseologischen Wörterbüchern aus der Sicht der ukrainischen Germanistik. Wierlacher, Alois, Konrad Ehlich, Ludwig M. Eichinger et al. (Hrsg.). 2003. *Jahrbuch Deutsch als Fremdsprache. Intercultural German Studies*. Bd. 29: 31-54. München: Iudicium.
- Roman, Alexandru.** 1993. *Dicționar frazeologic german–român*. București: Teora.
- Sava, Doris.** 2010. Defizite der rumänischen bilingualen Phraseographie mit Deutsch. *Germanistische Beiträge* 26: 185-222.

- Sava, Doris.** 2015. Linguistische Zugänge zur phraseographischen Beschreibungspraxis. Einsatzmöglichkeiten für die rumänische bilinguale Phraseographie mit Deutsch. *Forschungen zur Volks- und Landeskunde* 58: 159-176.
- Schemann, Hans.** 1989. Das phraseologische Wörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.). 1989: 1019-1032.
- Schemann, Hans.** 1991. Die Phraseologie im zweisprachigen Wörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.). 1991: 2789-2794.
- Steyer, Kathrin.** 2000. Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikographische Möglichkeiten. *Deutsche Sprache* 28: 101-125.
- Steyer, Kathrin.** 2004. Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikographische Perspektiven. Steyer, Kathrin (Hrsg.). 2004. *Wortverbindungen — mehr oder weniger fest*. Jahrbuch des Instituts für Deutsche Sprache 2003: 87-116. Berlin: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998a. Pragmatische Informationen in neuhochdeutschen Wörterbüchern. Ein Beitrag zur praktischen Lexikologie. Wiegand, Herbert Ernst (Hrsg.). 1998. *Studien zur neuhochdeutschen Lexikographie I*: 139-271. Hildesheim/Zürich/New York: Georg Olms. (= Germanistische Linguistik 3-4/79).

Lost and Found: The Value of a Little Known Bilingual Dictionary Towards the Intellectualization of Ndau*

Emmanuel Sithole, *Rhodes University, Grahamstown,
South Africa (emmanesu@gmail.com)*

and

Dion Nkomo, *School of Languages, Rhodes University, Grahamstown,
South Africa (d.nkomo@ru.ac.za)*

Abstract: This article critically evaluates the *ChiNdau–English and English–ChiNdau Vocabulary: With Grammatical Notes*, a bilingual dictionary published in 1915 by the American Board Mission (Rhodesian Branch), with a view of determining its suitability for use as a point of departure for modern Ndau lexicography. More than a century after its publication, it remains the only dictionary in the language. The language has been treated as a dialect of Shona for close to a century, until its emergence as one of the country's sixteen officially-recognized languages in the 2013 *Constitution of Zimbabwe*. This landmark development requires practical work that can transform Ndau into a fully-fledged and intellectualized language. Lexicography is one of the key intellectual enterprises that can contribute in this regard. It is therefore the contention of this article that the existing dictionary be considered as a vital point of reference for future lexicographic work in Ndau. An analysis of various aspects of the dictionary indicate that, notwithstanding some limitations, this dictionary indeed managed to set some standards that may be incorporated in current and future lexicographic works in this less documented language.

Keywords: DICTIONARIES, MISSIONARY LEXICOGRAPHY, CHINDAU, CHINDAU VOCABULARY, NDAU, NDAU LEXICOGRAPHY, NDAU ORTHOGRAPHY, LANGUAGE INTELLECTUALIZATION, MARGINALIZED LANGUAGES, OFFICIALLY-RECOGNIZED LANGUAGES, ZIMBABWEAN LEXICOGRAPHY

Opsomming: Verlore en gevind: Die waarde van 'n taamlik onbekende tweetalige woordeboek rakende die intellektualisering van Ndau. In hierdie artikel word die *ChiNdau–English and English–ChiNdau Vocabulary: With Grammatical Notes*, 'n tweetalige woordeboek wat in 1915 deur die American Board Mission se Rhodesiese tak gepubliseer is, krities beoordeel ten einde die toepaslikheid daarvan vir gebruik as 'n vertrekpunt vir die moderne

* 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.

Ndaou-leksikografie te bepaal. Meer as 'n eeu ná die publikasie daarvan is dit steeds die enigste woordeboek in die taal. Die taal is vir ongeveer 'n honderd jaar behandel as 'n dialek van Shona totdat dit in 2013 in die *Grondwet van Zimbabwe* as een van die land se sestiende amptelik erkende tale verklaar is. Hierdie belangrike gebeurtenis vereis praktiese werk ten einde Ndaou in 'n volledige en geïntellektualiseerde taal te omskep. Leksikografie is een van die deurslaggewende intellektuele ondernemings wat hiertoe kan bydra. Dit is dus die betoog van hierdie artikel dat die bestaande woordeboek beskou word as 'n onmisbare uitgangspunt vir toekomstige leksikografiese werk in Ndaou. 'n Ontleding van verskeie aspekte van die woordeboek toon dat hierdie woordeboek ten spyte van sommige beperkings inderdaad daarin geslaag het om sekere standaarde te stel wat in huidige en toekomstige leksikografiese werke in hierdie minder gedokumenteerde taal ingesluit kan word.

Sleutelwoorde: WOORDEBOEKE, SENDINGLEKSIKOGRAFIE, CHINDAU, CHINDAU-WOORDESKAT, NDAU, NDAU-LEKSIKOGRAFIE, NDAU-ORTOGRAFIE, TAALINTELLEKTUALISERING, GEMARGINALISEERDE TALE, AMPTELIK ERKENDE TALE, ZIMBABWIESE LEKSIKOGRAFIE

1. Introduction

Ndaou is a language that straddles Zimbabwe's south-eastern districts and central Mozambique with a combined total population of almost 1,6 million speakers (Instituto Nacional de Estadística 2010; Zimbabwe National Statistics Agency 2012). It has five mutually intelligible varieties namely, Shanga and Danda, spoken in Mozambique, Garwe and Tonga, spoken exclusively in Zimbabwe, and Ndaoundau/Ndaou which straddles the arbitrarily drawn international border between the two countries (MacGonagle 2007). Despite constitutional recognition in both countries, the language is still patently marginalized as it is restricted to private and community domains where it mediates communication in interpersonal, religious and cultural domains. Despite the vital parent-to-child transmission of Ndaou, the language confronts a genuine displacement risk by powerful contact languages like English, Shona and Portuguese which Ndaou speakers now find convenient to adopt and use even in their home and community domains. The language remains under-described and under-documented, with very little church literature that was published by missionaries between 1903 and 1928 (Sithole 2017). This article, therefore, seeks to offer a critical discussion of how the *ChiNdaou-English and English-ChiNdaou Vocabulary: With Grammatical Notes*, henceforth the *Ndaou Vocabulary*, can be used as a basis for modern Ndaou lexicography and the general intellectualization of Ndaou.

The term *intellectualization* has a long history dating back to the Prague School of Linguistics in the 1930s where it referred to the development of language to "express the interrelationships and complexity of thought processes" (Havranek 1964: 6) of a linguistic community. In contemporary scholarship, the term has been given intellectual currency by Philippines scholars, who include Bonifacio Sibayan who defines "an intellectualized language as one that func-

tions in the controlling domains of language" (Sibayan 1999: 229). Thus language intellectualization refers to the development and use of a language in high-status domains such as education, media, science, technology, and economy among others. In the context of Ndaou, this necessitates urgent interventions to codify, elaborate and document it so that it can function in intellectual domains consistent with the constitutional stipulations in Zimbabwe.

Most of the earliest literature in African languages published by missionaries is now little known as the majority of mother-tongue speakers, researchers and academics have become oblivious of its existence. Accordingly, this article regards them as lost. For Ndaou, some of the materials, such as the *Ndaou Vocabulary*, were never disseminated widely while others were destroyed by fire alongside a printing plant at Chikore Mission in 1916. Only some few selected texts were re-published later (Doke 1931a: 132). For instance, the *Ndaou Vocabulary* has been reproduced in microfilm format and is now available at a number of university libraries in South Africa. Thus the dictionary has been found and, together with other literary materials, needs to be re-examined in order to determine its contemporary relevance and potential use in current and future language intellectualization endeavours.

The history of the development of languages in general confirms the salience of old materials in providing a basis for language intellectualization. Old materials such as dictionaries are relevant and influential in the codification of marginalized languages such as Ndaou. For illustration, Samuel Johnson's (1755) *A Dictionary of the English Language* was useful as a basis for the compilation of the *Oxford English Dictionary* (Nkomo 2017). It is thus reasonable for academics to subject such materials to critical scrutiny. For example, Gouws (2016) examines "a few lesser known Afrikaans dictionary monuments" (Gouws 2016: 355). In the same way, this article demonstrates that *Ndaou Vocabulary* can effectively support scholarly efforts of documenting, preserving, maintaining and intellectualizing Ndaou. A number of pertinent issues regarding the dictionary, such as orthography, lemmatization, definitions and lexical coverage are examined in order to get an insightful perspective for future lexicographic work in the language.

2. Functions of dictionaries in marginalized languages

To date lexicography as a discipline and practice boasts of a glorious historical timespan of over 4000 years with all sorts of dictionary typologies — normative, synchronic, diachronic, pedagogic, monolingual, bilingual and multilingual — being published in different forms over that *longue durée*. Both the so-called civilized and uncivilized worlds have continued to find the motivation to produce lexicographic products of diverse types because of the fundamental role of dictionaries in society, particularly their problem-solving function (Béjoint 2010; Hartmann 2001; Yong and Peng 2007). Dictionaries execute pertinent linguistic, communicative and cognitive roles in society. Yet even more

compelling for marginalized languages is the language planning role played by dictionaries towards the emergence of national languages from European vernaculars together with the rise of modern European states during the Renaissance period (Mugglestone 2015; Nkomo 2017).

African languages are marginalized in higher order domains owing to colonial language policies. However, it is notable that a few languages like Swahili have morphed into powerful languages hogging the public domains in society — thanks to deliberate, strategic and timely political interventions by committed stakeholders. In this respect, the focus of this article narrows to marginalized languages, such as Ndaou in Zimbabwe, that are still restricted to religious and cultural domains. They are in such a precarious position because of serious challenges involving, but not limited to, assimilation into dominant languages, widespread antipathy against their use in key domains as well as the absence of appropriate materials to facilitate their use in such influential domains. To reverse this situation, this article maintains the scholarly view that dictionaries play a pivotal role in language development (Landau 2001; Hadebe 2006). Dictionaries are important in setting a basis for language codification, language standardization, language planning, language documentation and language resuscitation.

For a marginalized language like Ndaou, print dictionaries, online dictionaries, computer dictionaries and cellphone dictionaries are critical for its revitalization, growth and intellectualization. In an uncodified and unstandardized language, earlier lexicographic works can inform better documentation and standardization because their data could be reutilized for many other language-related activities including literacy, grammar, literature and linguistic description. Appropriate and informed use of dictionaries helps in mediating communication between various stakeholders in public domains. That way, dictionaries can meaningfully facilitate the acquisition and dissemination of information and knowledge, thereby satisfying current and future needs of users. From that perspective, it is hoped that lexicography can assist in transforming the status of Ndaou into a functional language that executes utilitarian and cognitive functions in education, media, technology, administration, legislature and the judiciary.

3. The *Ndaou Vocabulary*

It is not very clear who compiled and edited the *Ndaou Vocabulary*. Several versions of the dictionary in existence bear different names of missionaries who worked in the Ndaou speaking communities in Zimbabwe and Mozambique. Among these missionaries are J.P. Dysart, G.A. Wilder and A.J. Orner. For example, there is a version that is attributed to Wilder, another one that is credited to Dysart and Orner, and another version that simply bears the name of the publisher, the American Board Mission. Two things that are certain are that the dictionary was published by the American Board Mission at Mt.

Selinda, Zimbabwe, and that it was published in 1915, since those details appear consistently in all the versions. Although the content of the dictionary is the same, this article is based on the version attributed to J.P. Dysart and A.J. Orner. Its compilation had an external motivation (Gouws 2005) of serving "new missionaries, civil servants and other white residents of the country" (Dysart and Orner 1915: 4). It was intended to "be of real service to those who may use it in the evangelization and civilization of the Ndau people" (Dysart and Orner 1915: 4). That way, it would serve both pedagogical and communicative functions. In line with this purpose, the dictionary provides some brief notes on Ndau grammar, as clearly expressed in its title. The remainder of the article discusses the salient features of the dictionary in order to assess its contribution towards the codification and standardization of Ndau. The discussion will focus on the lexical coverage, data presentation, lemmatization, definitions and orthographic choices made in the dictionary. Such a discussion hopes to inform lexicographic discussions around the compilation of future dictionaries not only in Ndau but also in other marginalized languages.

3.1 Lexical coverage

The *Ndau Vocabulary* has 4000 lemmas across two alphabetical sections. It follows the traditional **Aa**-to-**Zz** listing approach with the Chindau–English section occupying 45.3% (63 pages) while the English–Chindau section takes up 39.5% (55 pages). The front matter which comprises of the preface and grammatical notes has 13.7% (19 pages) while the back matter which includes names of birds, fish and trees has only 0.7% (1 page). The 4000 wordlist in the dictionary is not only a good starting point for future lexicographic endeavors but also for linguistic and sociolinguistic purposes. Lexicographers could use the wordlist to guide the compilation of future monolingual and bilingual dictionaries while linguists and sociolinguists could use it as a basis for the study of linguistic notions such as phonology, semantics, lexical variation and language change, given the paucity of texts in the language.

As a cross border language, entries for the dictionary were drawn from the five dialects of Ndau already identified above. Entries from the five different dialects were labelled using either the shortened forms of the dialect, for example, **Chid** for **Chiduma** (Tonga), or the abbreviations for the place where the dialect is spoken, for example, **L.B** for **Lower Buzi** (Shanga) (Dysart and Orner 1915: 20). The inclusion of vocabulary from the different dialects captures the heterogeneous nature of Ndau. In future Ndau lexicography, dialect labeling as done in the *Ndau Vocabulary* should be maintained not only for showing the dialect scenario in the language but also towards contributing to the overall intellectualization agenda of Ndau. It would also avoid criticism and rejection problems akin to *Duramazwi ReChiShona*, the monolingual Shona dictionary, which some reviewers and users vowed not to use, insisting that it was biased towards Karanga, the mother-tongue of the editor-in-chief (Chimhundu 2010: 110).

The dictionary also accommodated a significant amount of borrowings from contact languages such as English, Afrikaans, Portuguese and Shona. Despite being a general-purpose dictionary with an overly Christian inclination, the dictionary also covers vocabulary from other important domains of society. In view of that, the dictionary strives to maintain a fair lexical balance by including data from diverse primary and secondary sectors. Such a positive attribute can be used as a stepping stone towards developing future monolingual and bilingual dictionaries in the language which generally describe the language as it is used across all major domains (Zgusta 1971). While the compilers acknowledge the presence of inconsistencies, "faults, errors and incompleteness" (Dysart and Orner 1915: 3) on lemma selection and the overall dictionary contents, it is proffered that the *Ndaau Vocabulary* can be very useful towards providing a foundation for the compilation of future dictionaries. It sets a norm or a standard upon which Ndaau can be described for linguistic purposes. The dictionary can also be employed as a point of reference for further studies on Ndaau vocabulary, grammar, terminology and lexicography. These activities assist towards corpus planning which is necessary for the intellectualization of the language.

3.2 Presentation of data

This section is a critical examination of both the macrostructural and microstructural aspects of the *Ndaau Vocabulary*. It discusses how the compilers handled lexicographic aspects such as orthography, lemmatization, definitions, and grammar to establish how they can assist in developing Ndaau. Such aspects are essential as they determine the extent to which dictionaries meet user needs. Largely, such aspects interlink with the size, volume and entire structure of the dictionary. It is therefore imperative to make a critical engagement with such aspects to identify how they can inform future lexicographic activities in Ndaau. From a practical point of view, such issues are important as they help towards the standardization of the language as users often consult dictionaries to learn how to spell, pronounce and use words correctly (Hadebe 2006; Chimhundu 2010). It is therefore argued that the indicated aspects can determine the overall accessibility, acceptability and use of any dictionary compiled to meet communicative and cognitive needs of speakers in a society.

3.2.1 Orthographic issues

The compilers of the *Ndaau Vocabulary* deserve commendation for devising a spelling system that adheres to the orthographic principle of accuracy. According to Capo (2002: 21) accuracy is a scientific orthography development practice where there is a one-to-one correspondence between sound and symbol. Lexicographers who compiled the dictionary used this principle to represent salient phonological aspects of Ndaau such as aspiration, laterality and phonotactics.

As Doke (1931b: 147) emphasizes, aspiration is an outstanding and widespread feature that deserves to be represented in Nda. The compilers represented aspirated voiceless stops <th, p^h> as in **pump^ha** (make a false allegation) and **nt^hekwe** (snuff box) and laterality with <hl> as in **h^hlu** (giraffe). Regrettably, the compilers were not consistent as observed in lemmas like **k^hum^hucha** (translate), **mukeyo** (drinking pot), **kebe** (melon) being lemmatized instead of **k^hum^hucha**, **muk^heyo** and **k^hebe**. The compilers also distinguished the voiced labialized alveolar fricative <zv> with the voiced alveolar fricative <z> as in **zina** (name) and **zina** (dance) respectively while also using <v> to represent the voiced dentilabial fricative <vh> and <v> voiced labiodental fricative <v> as in **vanga** (scratch hard) and **vangeri** (gospel).

However, after Doke's (1931a) designation of Nda as a dialect of Shona, some of its peculiar features such as clicks, laterals, aspiration were sacrificed. Interestingly, Doke's landmark one-sound-one symbol principle that resulted in special symbols borrowed from the International Phonetic Alphabet (IPA) was later abandoned and replaced in successive revisions. From this perspective, it can be argued that the *Nda Vocabulary* laid a scientific foundation in orthography development which, when improved, can greatly assist towards the intellectualization of the language in functional domains. Its accurate symbols on aspiration, laterality and phonotactics among others can be incorporated in new revisions of Nda orthography to fully match Nda sounds to adopted symbols.

3.2.2 Lemmatization

The *Nda Vocabulary* adopts a full word-lemmatization approach in the **Aa-to-Zz** list. Full words are entered as the canonical forms of headwords without removing prefixes or extensions. This includes all word categories such as nouns, verbs, adverbs, interjections and prepositions. Only adjectives are lemmatized using the stem-based approach. Below is an examination and exemplification of how different speech categories are entered in the dictionary.

3.2.2.1 Lemmatization of nouns

The word lemmatization approach resulted in nouns being written directly as they are spoken. The initial letters of headwords were used to decide the alphabetical listing of nouns. The following is a description of how nouns are presented:

- The regular plural prefix is given after the noun, for example, **besha**, ma, cold in head.
- When the prefix is not used with the noun as in 'i-/dzi-' class, the plural pronoun is given in parentheses as in **mbudzi**, (dzi), goat.

- Where the plural form is irregular the singular pronoun is given in parentheses followed by the plural noun with its pronoun in parenthesis as in **rukuni**, (ru), (Plur. huni-dzi), stick of fire-wood.
- Where a noun occurs in the singular or plural form only, it is followed by its pronoun and indicated as in **nyama**, (I sing. only), meat.
- Nouns of the 'u' and 'ku' classes, which have one form only, are followed by the pronoun in parenthesis as in **uchi**, (hu) honey; **kurgya**, (ku), food.
- Nouns of the 'mu-/mi-' class beginning with 'mw' are followed by both the singular and plural pronouns when used in both singular and plural as in **mwoyo**, (u, i), heart.
- Where letters in nouns of the 'ri-/ma-' class change in the plural they are given in parenthesis as in **banga**, ma (p), knife (Dysart and Orner 1915: 19).

A critical look into how nouns are presented reveals that the compilers mixed different strategies in the lemmatization of nouns. The compilers used the full word lemmatization approach by entering singular nouns only as headwords. As singularity and plurality in Ndaou is mostly indicated through prefixes, the lemmatization of singular noun headwords accompanied by their regular plural form prefixes became a space- and time-saving lexicographic technique that avoided lexical redundancy in the dictionary. It also links singular and plural nouns, for example, **beshu** (cold in head) and **mabeshu** (mucus) to grasp singularity and plurality rules in Ndaou. But for purposes of future lexicographic endeavours in Ndaou, it could be necessary to improve the lemmatization of nouns to make it more consistent. For instance, the description of **bare** (salt-pan) as exclusively a singular noun is misleading because there is **mapare** (salt-pans). It is evident that lemmas such as **benzi** (fool) and **dahwa** (night hawk) change into **mapenzi** (fools) and **matahwa** (night hawks) respectively when they are pluralized and not ***madahwa** and ***mabenzi** (which are grammatically incorrect) as implied in the morphophonemic rules governing the 'ri-/ma-' nouns.

The full word lemmatization approach used in the dictionary is usually the standard approach in agglutinating languages such as Ndaou where conjunctive writing is usually preferred. This approach provides an easy and straightforward access to the lemma because it is given the full nominal form. Users do not encounter challenges of looking up the stem which in some cases like in class 9/10 is evidently elusive. It is also easy to provide the category of the lemmas while also reassuring users that they are dealing with the correct lemma unlike using the stem lemmatization approach (Prinsloo 2011: 183-184).

3.2.2.2 Lemmatization of verbs

The *Ndaou Vocabulary* consistently uses the full word lemmatization in verbs. In the quest to indicate grammatical information, it indicates 'v,' for 'verb', and

'adv.,' 'adj.,' 'prep.,' 'conj.,' and 'interj.,' representing 'adverb,' 'adjective,' 'preposition,' 'conjunction,' and 'interjection' respectively as shown below.

- dai**, v., be like this.
- , conj., as; if.
- , verbal part. Indicates "let" as in-Ex. dai ndiende.
- , ma (t), branch of tree.

The adopted labeling style performs a key role of distinguishing different senses and categories of words. The lemma above has different categories and multiple meanings. Depending on the context, the lemma **dai** is used as a verb, verbal particle, conjunction and also as a noun which in its plural form is **matai** (branches). This clarifies the different senses in which the word is used. Despite its advantages, a word-based approach to the lemmatization of verbs is not feasible in a paper dictionary because verb stems can combine with numerous extensions. Unlike in English, agglutinating African languages like Ndaus verbs acquire prefixes, concords and extensions, which pose serious challenges in lemmatizing them. Some disadvantages of full word lemmatization in the *Ndau Vocabulary* are shown in examples below:

- amba**, v., begin; attack.
- ambana**, v., quarrel.
- ambira**, v., offend.
- kuambana**, (ku), quarrel.

In relation to verbs, the word lemmatization approach encourages repetition; promotes the scattering of related words; and wastes space unnecessarily. It is doubtless that the alphabetization of verbs using their accompanying prefixes does not only physically separate verbs but also differentiates them with nominal derivatives such as **(ku)ambana** among others. In addition to that, adopting this approach promoted the reduplication of similar words in the *Ndau Vocabulary* as seen in Table 1 below:

Table 1: Lexical redundancy in the *Ndau Vocabulary*

Verb lemma	Gloss	Verb/noun lemma	Gloss
Buda	come out, go out	kubuda	projection
Chengedza	deceive	kuchengedza	deceit
Dakara	be happy	kudakara	happiness
Dai	be	kudai	as, even, so, like
Edza	tempt	kuedzwa	temptation
Gwimba	groan	kugwimba	groan
Irikidza	praise	kuirikidza	praise
Jeka	shine, light	kujeka	light
Angarara	be naughty	kuangarara	naughtiness

Table 1 shows the nominalization of verbs by prefixing them with the class 15 (ku-), but there is an interrelatedness between the meanings of verbs and nouns. This concurs with Prinsloo's (2011: 184) submission that at times nominal and verbal meanings are closely related which causes classification problems. These double, triple or quadruple entries in the *Ndaou Vocabulary* redundantly increased the total number of headwords. Adopting the stem lemmatization strategy could have reduced the number of lemmas by grouping verbs and their related senses under one stem as this is a modern and economical lexicographic practice (De Schryver and Wilkes 2008). For example, **amba**, **ambana**, **ambira** could have entered under **amba**. This could also have effectively addressed the challenge posed by verbal and nominal words, for example, nominal derivatives such as **kuambana** could have been lemmatized alphabetically as full words under **ku-**.

3.2.2.3 Lemmatization of adjectives

Adjectives were lemmatized according to their stems. This is seen in **doko** (small), **dodori** (little, small amount), **tsvuku** (red), **chena** (white), **svipu** (black), **rebu** (tall, long) and **refu** (long, tall, high). Stem lemmatization is clearly a good lexicographic practice that saves time and space while also avoiding repetition. The compilers also managed to respond to dialectal variation by including dialectal variants for example, **refu** (Shanga) vs **rebu** (Ndaoundau/Ndaou); **doko** (Tonga, Garwe, Shanga) vs **dodori** (Danda, Ndaoundau/Ndaou) as headwords. While the inclusion of variants as headwords was sociolinguistically and linguistically sound, it is, however, lexicographically less helpful as such lemmas are semantically similar. Apart from representing their individual dialects and the obtaining phonological differences, there is no useful information on them. This results in unnecessary repetition and redundancy, for example, **rebu**, adj., long; tall; high and **refu**, adj., long; tall; high.

While the decision to represent dialectal variation in the dictionary is sociolinguistically laudable, it is important that future lexicographers need to be careful not to include simple variants such as **refu** and **rebu** shown above. The two headwords are essentially allophones of the same word differing phonologically and dialectally. In modern lexicographic practice they could and should be entered as one word as they are lexically or semantically similar. It would suffice to use synonyms, for example, **mbishi** (Shanga) and **gan'a** (Ndaoundau/Ndaou) as headwords since they are lexically different. This is advantageous as it depicts the breadth and richness of Ndaou's vocabulary.

3.2.2.4 Lemmatization of loanwords

In an attempt to describe Ndaou as it was spoken, borrowed words were also lemmatized. This is in keeping with the lexicographic principle that dictionaries should be descriptive to document and express the current language as it is

spoken. The majority of loanwords are rephonologised to reflect native Ndaun speakers' pronunciation. The labels (Z) for Zulu, (E) for English, (D) for (Dutch), (Mash.) for Mashonaland, and (P) for Portuguese show the place or language from which the lemma is borrowed. Examples below show the loanwords in the dictionary as well as the languages from which they were borrowed:

bururu (isibululu [Zulu]; puff adder)
ngubo (ingubo [Zulu]; dress)
baptidza (English; baptize)
mukristu (English; Christian)
chapewa (chapeu [Portuguese]; hat)
kapita (cabeca [Portuguese]; head boy)
sono (Sondag [Afrikaans]; Sunday)
chari (sjaal [Afrikaans] shawl)

The compilers of the *Ndaun Vocabulary* deserve commendation for handling loanwords in the dictionary considering that they are normally a problem in lexicography (Hadebe 2006). Naturally, words from other languages such as English and Portuguese do not conform to the orthographic and phonological conventions of Ndaun. For instance, lemmas such as **sono**, **kristu**, and **baptidza** exemplified above usher in new phonetic, phonological, morphological and semantic elements into the receptor language. However, elements such as "nt, kr, st, pt and pr" frustrate the phonological rules of the language necessitating their separation by vowels, as in **khirisitu** and **bhabhatidza** and so on. However, the compilers failed to insert vowels to conform to the Ndaun spelling system.

Sociolinguistically, the inclusion of loanwords concurs with Kwama-ngamalu's (1997: 89) observation that no language is lexically self-sufficient as languages always supplement their vocabularies with loanwords from other languages. This means that Ndaun lexicon can easily grow through adopting and adapting words from contact languages. Examples of adoptives found in the dictionary include **chibuko** (mirror), **jaha** (boy) from Zulu; **chapewa** (hat) from Portuguese; **chiamburera** (umbrella), **ginandera** (granadilla) from English, **puraze** (prazero), **tafura** (table) from the then Dutch (now Afrikaans) and **nyakwaa** (smooth) from Shona. Labelling the source of borrowed words is critical in establishing a lexicographic foundation for treating loanwords in future Ndaun dictionaries.

The inclusion of loanwords in the *Ndaun Vocabulary* highlights sociolinguistic and diglossic situations in which Ndaun exists. For example, words originally from ex-colonial languages demonstrate language power politics where Ndaun is a subordinate language to English, Portuguese and Dutch as its speakers often borrow from them for prestige reasons. In the same vein, words borrowed from African languages such as Shona and Zulu expose the contemporary and historical contact scenario with Ndaun. For instance, Ndaun has a lin-

guistic relationship with Shona. Also, the inclusion of lexical items of Nguni origin points to a historical and linguistic connection between Zulu and Ndaou. Among other historians, MacGonagle (2007) states that Soshangane's Gaza-Nguni group which was fleeing the Mfecane wars conquered the Ndaou during the 1820s. The military, political, social and cultural contact, including inter-marriages, between the original Ndaou group and Soshangane's people resulted in a permanent accommodation of over 2000 Nguni lexical items in Ndaou (Doke 1931a; Fortune 1990).

3.2.3 Definitions

Most traditional bilingual dictionaries by missionaries in Zimbabwe were normally translations of lemmata in both languages (Mpofu-Hamadziripi et al. 2013). An example is Hannan's (1979) *Standard Shona Dictionary*. The *Ndaou Vocabulary* is in the same category implying that Ndaou-English section headwords are translated into English while the reverse is also true. Apart from being mere translations, the dictionary has neither real definitions, nor contextual information nor synonyms of entries. In keeping with the dictionary's purpose, the compilers managed to provide reasonably adequate equivalences between the two languages.

The *Ndaou Vocabulary* provides equivalences (which are treated as definitions in this article) that largely conform to the defining principles of cultural acceptability, accessibility and user-friendliness. Firstly, translators-cum-definers were careful in using non-technical language, which made word meanings easily accessible to learners in a general dictionary. Secondly, the compilers attempted to depict cultural and religious tolerance and resisted Eurocentric biases typical in literature published during the colonial period. For example, traditional religious terms such as **mudzimu** (personal ancestral spirit), **n'anga** (doctor), **hakata** (lots, bones for divining), **chiremba** (doctor skillful with hakata) could have been referred to as elements of either heathenism or paganism or both as in other early dictionaries in African languages. Such tendencies have been noted in other dictionaries that are products of missionary lexicography in Africa (Moropa and Kruger 2000). This makes the *Ndaou Vocabulary* a relevant and socio-culturally acceptable lexicographic work even among contemporary mother-tongue users. Thirdly, the equivalences were not circumlocutive but very concise which made the dictionary easily express meanings to intended users. This satisfied the principle of brevity, which encourages straightforward definitions and translations.

3.2.4 Grammatical information

As an underdescribed language, there are no grammatical works in Ndaou, which makes the *Ndaou Vocabulary* the only reference thus far. While the compilers advise that "the grammatical notes are not intended to be complete or

take the place of a grammar" (Dysart and Orner 1915: 3), the dictionary provides essential grammatical information that helps non-mother tongue speakers of Ndau to learn the language. It explains the morphophonemic rules governing the description and classification of nouns, pronouns, adjectives, verbs, verb phrases and inflections. For instance, it shows the rules determining the singularity/plurality of nouns in the (ri-/ma-) class, such as **banga**, ma (p), knife and (mu-/mi-) class like **munda**, (mi) field as well the concordance rules governing the language. This helps in showing existing word and lexical relationships in the language as it is shown in the dictionary.

While the importance and effectiveness of such an incomplete Ndau grammar to non-mother tongue learners of Ndau could be tremendously questionable as there is a whole inconclusive debate relating to whether or not grammatical knowledge is necessary for foreign language learning, it is nevertheless very important to the standardization of the language. Acquiring the knowledge of grammar is critical to establish a basis upon which rules for writing, reading and mastering Ndau can be standardized. From the perspective of Ndau's intellectualization, it is important to emphasize that grammatical information is essential. This means that improving grammatical rules contained in the dictionary is critical towards informing further linguistic studies on Ndau.

3.4 Ndau lexicography and the road ahead

The way ahead for Ndau lexicography requires careful planning to produce a variety of dictionaries that serve as utility products and contribute towards the intellectualization of the language. Building on the century old, little known or lost and found dictionary, scholars and lexicographers, particularly mother-tongue lexicographers, should design sound strategies and plans for the development of future dictionaries. Positive gains on orthography development; dialect unification and representation; cultural, religious and gender sensitivity; grammar; and phonology should be built on as explained in foregoing sections. Such strides should help create a foundation upon which future documentation of Ndau can be premised. This is more particularly advantageous in lexicography where a starting point has already been established.

The planning of future mono-/bi-/multilingual, specialized, general and pedagogical dictionaries in the language can meaningfully begin by using the *Ndau Vocabulary* as the basis. Researchers can begin by studying the 4000 headwords in the dictionary to establish whether they are still in use before collecting new words and defining them appropriately, taking cognisance of contemporary contexts.

Furthermore, data in the general-purpose dictionary can be converted to serve a specialized pedagogical function. This means that a children's dictionary can be built from a quantitative reduction of the words in the current dictionary. The basic vocabulary list from the dictionary can be selected, described

and explained in a simple language consistent with children's level of linguistic competence. In a children's dictionary where language instruction for communication purposes is pursued it might be desirable to improve the data presentation approach. This can be very useful in concentrating essential information that can be constructively used for educational purposes. This is so despite the fact that scholars such as Hernandez (1989: 152) insist that reducing the size of a dictionary is a deplorable practice in the modern world of school dictionaries. Tarp (2011: 224) also echoes this view when he intones that "when a dictionary is the result of a purely quantitative reduction, it can no way be considered a work conceived for users." This is correct, taking into account that, for example, the original target users of the dictionary and native Ndaau children's linguistic needs are different. However, it can still be possible and necessary with proper planning to adapt and produce a dictionary whose contents accurately address its user needs. For example, the reduction of old lemmas in the *Ndaau Vocabulary* can be complemented by adding new vocabulary collected from every day communicative situations and define them appropriately to identify with children's communicative and cognitive needs. In that way, a new children's dictionary would be an improved lexicographic work, which would have used *Ndaau Vocabulary* as a basis.

In view of the foregoing, it is important that the development of Ndaau lexicography should assist immensely in the intellectualization of the language in Zimbabwe. Nkomo (2012: 223) is critical of the comprehensive lexicographic processes in Zimbabwe for their exclusive focus on Shona and Ndebele. He proposes a well-conceived comprehensive lexicographic process for Zimbabwean lexicography that also includes English as well as minority languages that were already gaining traction in education using successive language-in-education policies. In that case, an appropriate lexicographic intervention process is imperative in all the 16 officially-recognized languages with specific focus on minority languages such as Ndaau. The constitutional recognition of previous minority languages to have official parity with majority languages such as Shona and Ndebele emphasizes the necessity of a paradigm shift to embrace multilingualism in the country by technically developing all the languages to function in important high-status domains. This bestows relevance on the crafting a comprehensive language policy whose tenets and stipulations are mirrored and implemented through an appropriate lexicographic intervention process in Zimbabwe. Taken together, this implies that the time is ripe to not only take stock and a critical evaluation of existing dictionaries such as the *Ndaau Vocabulary* but to also think carefully about specific dictionary projects in the concerned languages.

3.5 Conclusion

Over a century ago, the compilers of the *Ndaau Vocabulary* made it clear that the dictionary was never intended to be a complete or final grammatical or lexico-

graphic work for the language. Instead, they acknowledged its 'incompleteness in the lists of words', definitions and other errors (Dysart and Orner 1915: 4) which need proper and careful attention in future lexicographic publications. Most importantly, they were cognizant of their own shortcomings but undertook to compile a work which would "serve as a basis for comparison" with other works in other Bantu languages "by those who were competent to make such a study." Such awareness indicates that the compilers never intended to forestall and preclude future and deeper studies in Ndaue alone or as part of the larger Bantu language family. Instead, they encouraged further studies in Ndaue by encouraging constructive criticism in the "spirit that seeks to improve the work already done" (Dysart and Orner 1915: 4). In tandem with such a selfless and visionary exhortation, this article focuses on how the *Ndaue Vocabulary* could help create a foundation for future linguistic and lexicographical debates and practices. The dictionary can inform language documentation in the form of lexicography, terminology, grammar, linguistic comparison and language development. This makes it strategically positioned to support and champion the intellectualization of Ndaue from a restricted community language to a vehicular language functioning in high-status domains. With Ndaue now being one of the officially-recognized languages of Zimbabwe, there has not been a better time to build on the century old dictionary towards the intellectualization of the language.

References

- Béjoint, H.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Capo, H.B.C.** 2002. The Pan-dialectal Approach to Orthographic Conventions: The Case of the Gbe Languages of West Africa. Prah, K.K. (Ed.). 2002. *Writing African: The Harmonization of Orthographic Conventions in African Languages*: 5-25. Cape Town: The Centre for Advanced Studies in African Society (CASAS).
- Chimhundu, H.** 2010. Harmonization and the Handling of Language Variation in Shona. Lexicography. Chimhundu, H., W. Magwa and A. Chebanne (Eds.). 2010. *Harmonization of Shona-Nyai varieties*: 105-122. Cape Town: CASAS.
- De Schryver, G.-M. and A. Wilkes.** 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 827-836. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Doke, C.M.** 1931a. *Report on the Unification of the Shona Dialects Carried Out under the Auspices of the Government of Southern Rhodesia and the Carnegie Corporation*. Hertford: Stephen Austin and Sons.
- Doke, C.M.** 1931b. *A Comparative Study in Shona Phonetics*. Johannesburg: University of Witwatersrand Press.
- Dysart, J.P. and A.J. Orner.** 1915. *ChiNdaue-English English-ChiNdaue Vocabulary: With Grammatical Notes*. Mt. Selinda: American Board Mission, Rhodesian Branch.
- Fortune, G.** 1990. From Zulu to Ndaue: A Change of Medium. *South African Journal of African Languages* 10(4): 354-364.

- Gouws, R.H.** 2005. Lexicography in Africa. Brown, K. (Ed.). 2005. *Encyclopedia of Language and Linguistics*: 95-101. Second edition. Oxford: Elsevier.
- Gouws, R.H.** 2016. Enkele minder bekende Afrikaanse woordeboek-monumente. *Tydskrif vir Geesteswetenskappe* 56(2-1): 355-370.
- Hadebe, S.** 2006. The Standardisation of the Ndebele Language through Dictionary-making. Harare/Oslo: The ALLEX Project.
- Hannan, M.** 1979. *Standard Shona Dictionary*. Second edition — Reprint with Addendum. Harare: College Press.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Essex: Pearson Education.
- Havranek, B.** 1964. The Functional Differentiation of the Standard Language. Garvin, P. (Ed.). 1964. *A Prague School Reader on Esthetics, Literary Structure and Style*: 3-16. Washington DC: Georgetown University Press.
- Hernández, H.** 1989. *Los diccionarios de orientación escolar. Contribución al estudio de la lexicografía monolingüe española*. Tübingen: Max Niemeyer.
- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton/T. & T. Longman et al.
- Kwamangamalu, N.M.** 1997. Language Frontiers, Language Standardization, and Mother Tongue Education: The Zaire–Zambia Border Area with Reference to the Bemba Cluster. *South African Journal of African Languages* 17(3): 88-94.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York/Cambridge: Cambridge University Press.
- MacGonagle, E.** 2007. *Crafting Identity in Zimbabwe and Mozambique*. Rochester: University of Rochester Press.
- Moropa, K. and A. Kruger.** 2000. Mistranslation of Culture-specific Terms in Kropf's *Kaffir–English Dictionary*. *South African Journal of Languages* 20(1): 70-79.
- Mpofu-Hamadziripi, N., A. Ngunga, E.N. Mberi and F. Matambirofa.** 2013. *A Descriptive Grammar of Shona*. Harare: Sable Press.
- Mugglestone, L.** 2015. *Samuel Johnson and the Journey into Words*. Oxford: Oxford University Press.
- Nkomo, D.** 2012. *Towards a Lexicographical Intervention in the Acquisition and Use of English in Zimbabwe*. Unpublished D.Litt. Dissertation. Stellenbosch: Stellenbosch University.
- Nkomo, D.** 2017. Dictionaries and Language Policy. Fuertes-Olivera, P.A. (Ed.). 2017. *The Routledge Handbook of Lexicography*: 152-165. London: Routledge.
- Prinsloo, D.J.** 2011. A Critical Analysis of the Lemmatisation of Nouns and Verbs in IsiZulu. *Lexikos* 21: 169-193.
- Sibayan, B.P.** 1999. The Intellectualization of Filipino. *International Journal of the Sociology of Language* 88(1): 69-82.
- Sithole, E.** 2017. *From Dialect to 'Official' Language: Towards the Intellectualization of Ndaou in Zimbabwe*. Unpublished D.Phil. Thesis. Grahamstown: Rhodes University.
- Tarp, S.** 2011. Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. *Lexikos* 21: 217-231.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam/Philadelphia: John Benjamins.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography

Sven Tarp, *Department of Afrikaans and Dutch,
University of Stellenbosch, South Africa and Centre for Lexicography,
Aarhus University, Denmark (st@asb.dk)*

Kasper Fisker, *Ordbogen A/S, Odense, Denmark
(kfi@ordbogen.com)*

and

Peter Sepstrup, *Ordbogen A/S, Odense, Denmark
(pse@ordbogen.com)*

Abstract: Dictionaries are increasingly integrated into other tools designed to assist the reading, writing and translation of texts. *Write Assistant* is a newly developed tool aimed at assisting people writing in a second language. It feeds on big data taken in from corpora and digital dictionaries. The paper discusses the philosophy behind the tool, the techniques applied, its empirical basis and functionality, as well as the extent to which it helps its users. It shows how the tool makes it possible to shorten and even skip some phases in the traditional information-search process and allows its user to maintain the focus on the message to be written without the need to consult external information resources. The paper shows how the underpinning technology gives birth to a new type of dictionary that is context-aware and provides a more personalised user service. But it also indicates that future dictionaries need to be conceptionally adapted to the specific tool in order to optimize the service. All this poses new challenges to lexicography.

Keywords: *WRITE ASSISTANT, INFORMATION TOOL, INTEGRATED DICTIONARIES, L2 WRITING, INFORMATION-SEARCH PROCESS, EMPIRICAL RESOURCES, CORPORA, LANGUAGE MODEL, CONTEXT-AWARE DICTIONARIES, DICTIONARY CONCEPT*

Opsomming: L2-Skryfhulpmiddels en konteks-sensitiewe woordeboeke: Nuwe uitdagings vir die leksikografie. Woordeboeke word toenemend geïntegreer in ander hulpmiddels wat ontwerp is om te help met die lees, skryf en vertaling van tekste. *Write Assistant* is 'n nuut ontwikkelde hulpmiddel wat gerig is daarop om mense wat in 'n tweede taal skryf te assisteer. Dit is gebaseer op groot data wat verkry is uit korpora en digitale woordeboeke. In hierdie artikel word die filosofie agter die hulpmiddel, die tegnieke wat toegepas word, die empiriese basis en die funksionaliteit daarvan bespreek, sowel as die mate waartoe dit die gebruikers daarvan help. Daar word aangetoon hoe die hulpmiddel dit moontlik maak om sommige fases in die funksionele inligtingsoektogproses te verkort en selfs te weg te laat, en hoe dit die gebruiker toelaat om fokus op die boodskap wat geskryf moet word, te behou sonder dat dit nodig is om eksterne inligtingsbronne te raadpleeg. In hierdie artikel word aangedui hoe die onderliggende teg-

nologie lei tot die skep van 'n nuwe soort woordeboek wat konteks-sensitief is en wat 'n meer persoonlike gebruikersdiens verskaf. Dit dui egter ook daarop dat toekomstige woordeboekbehoefes konseptueel by die spesifieke hulpmiddel aangepas moet word om die diens te optimaliseer. Dit hou alles nuwe uitdagings in vir die leksikografie.

Slutelwoorde: *WRITE ASSISTANT*, INLIGTINGSHULPMIDDEL, GEÏNTEGREERDE WOORDEBOEKE, L2-SKRYFWERK, INLIGTINGSOEKTOGPROSES, EMPIRIESE HULPBRONNE, KORPORA, TAALMODEL, KONTEKS-SENSITIEWE WOORDEBOEKE, WOORDEBOEKKONSEP

1. Introduction

To raise new questions, new possibilities, to regard old questions from a new angle, requires creative imagination and marks real advances in science. (Einstein and Infeld 1938: 92)

Over the years, dictionaries have been designed and compiled to meet a big variety of human needs detected in society. One of their many functions has been to assist users who write texts in a foreign language. As Chon (2009) reports, people engaged in this activity frequently consult a combination of L1–L2, L2 and L2–L1 dictionaries when they experience different types of problem and require information in order to solve them. The tradition of consulting a combination of monolingual and bilingual dictionaries has continued from the printed to the digital world where, surprisingly, only a relatively few dictionaries, like *WorldReference*, so far have been conceived as an integrated package that can meet both monolingual and bilingual user needs in connection with L2 text production. Although this last class of dictionary undoubtedly represents an important step forward for lexicography, the time now seems ripe to take another and even bigger step forward in terms of a more personalized and efficient user service. The technological preconditions for this already exist. It is merely a question of taking full advantage of the available computer and information technologies.

One of the many current challenges is to rethink the whole process which people traditionally follow when they look for information. This process is more or less as described by Bergenholtz et al. (2015):

1. a person experiences an information need in a specific situation,
2. becomes aware of the need,
3. decides to take action and consults an information bearer, e.g. a dictionary,
4. carries out the consultation,
5. evaluates the result, i.e. if it is satisfactory,
6. returns to the situation where the information need originally occurred,
7. and uses the retrieved information to solve the problem that gave rise to the need.

This process is quite complex and demanding. In a lexicographical perspective, it requires, among other things, that the person experiencing the need is aware of this need, can define its character, has a dictionary at hand and knows how to consult it and find the relevant data, is capable of retrieving the required information from these data, is sufficiently prepared to evaluate the retrieved information, and knows how to apply it. Besides, the process is in any case *time-consuming* and *disturbing* as it takes focus away from the main activity which the person is performing, especially if it is a communicative one like text production in a non-native language where the message to be transmitted is the central issue. Nesi (2015: 584) rightly observes that "people typically consult maps, encyclopaedias and dictionaries while they are doing something else". This being the case, any consultation of an external information resource inevitably represents an interruption of the activity in question. It may be assumed that most users of these resources just want to go back, as quickly as possible, to what they were doing in order to maintain the focus.

As an additional complication, writers looking for assistance in one of the many dictionaries available on the Internet are at risk of encountering unforeseen temptations. Many 21st century users, especially young people, expect lookups in such dictionaries to be free. Publishing houses all over the world are therefore struggling to find a sustainable business model, a challenge that have led some of them to resort to advertising as a survival strategy. This may be good for business but not necessarily for their users. Research into the phenomenon indicates that as much as five percent of all visitors to websites using advertising get tempted by the colourful banner ads and click through in order to get further information; see Robinson et al. (2007). Even those who do not follow this road paved with good intentions may still be distracted by the many temptations. When they finally return to the task they were performing they will probably have lost their focus and maybe even forgotten why they started the consultation in the first place.

Hence, from a user perspective the challenge is to design an information tool that makes it possible to shorten or even skip some of the above-mentioned phases and steps in order to reduce the overall consultation time, maintain the writing flow, and avoid that the person in question loses his concentration. In this respect, the key words are *time*, *flow* and *focus*, to which should be added *quality* of the provided information.

In the following sections, we will look at a tool that intends to take up the gauntlet, namely the *Write Assistant* developed by Ordbogen A/S, a Danish IT company providing language services and online dictionaries, both general and specialized. We will discuss the philosophy behind the tool, the techniques applied, its empirical basis and functionality, the extent to which it actually helps its users, as well as the new challenges posed to lexicography in this connection. But before doing so we will briefly look at the relationship between lexicography and technology as well as the requirements to assist L2 text production.

2. Lexicography and technology: Recent developments

Historically, there is an intimate relationship between lexicography and technology as shown by Hanks (2010), and Rundell and Kilgarriff (2011), among others. Technology has strongly influenced the design, compilation, presentation, distribution, availability, and usage of dictionaries. New disruptive inventions like the printing and computer technologies have led to genuine paradigm shifts with far-reaching consequences for the discipline. Fuertes-Olivera (2016) has described the current situation within lexicography as a "Cambrian explosion". In spite of the undeniable progress that can be observed, lexicography is in many aspects still in the process of fully adapting to the computer and information technologies put at its disposal. Among the main strategic challenges is to work out a lexicographical response to the increasing societal demand for a more personalised service; see Rundell (2010) and Tarp (2011).

Although lexicography can be viewed as a discipline in its own right, it is traditionally characterized by a big interdisciplinary spirit; see Nielsen (2017) and Tarp (2017). This is even more pronounced today where no serious lexicographical project can be carried out without the combined knowledge of lexicography, information technology, linguistics, and other disciplines relevant to the project. No single expert from any of these fields can produce a high-quality lexicographical product on his own. Whoever takes the initiative has to cross the disciplinary border and include knowledge and skills from the relevant fields. The integrated information tools discussed below would never see the light of the day without the combined efforts of IT experts, lexicographers, corpus linguists, Internet designers, etc. In this respect, the central role of lexicography is determined by its long experience and ability to specify user needs, define the corresponding lexicographical data, and establish the best ways to access these data. This role goes far beyond traditional lexicography and dictionary-making as predicted by Tarp (2009).

One of the most promising developments in recent years is the growing challenge to the traditional stand-alone dictionary, whether printed or digital, and the gradual integration of lexicographical products into other digital information tools designed to assist the reading, writing and translation of texts (see for instance Verlinde et al. 2010, Paquot 2012, Verlinde and Peeters 2012, and Granger and Paquot 2015). Another important development affects the very concept of consultation where an increasing number of lookups are made automatically, a phenomenon defined by Tarp (2008: 123) as "passive searching" in contrast to the "active searching" performed by the users themselves. These kind of "passive" lookups frequently take place without the users even being aware that they are consulting a dictionary. Although there is no reliable statistics, today most dictionary consultations are probably made in the various types of integrated information tool and even automatically.

These two innovations point in the right direction. The integration of dictionaries into other tools makes allowance for reduced "information costs" in

terms of the time spent in the consultation and in the processing and application of the retrieved information; see Nielsen (2008). In addition, the introduction of "passive searching" in dictionaries and other information resources tends to neutralize the well-known problem that "most users get tired of consulting a dictionary whenever they encounter an unfamiliar word" (Verlinde and Peeters 2012: 158). Both innovations represent a step forward towards an improved service that leaves the users of these tools with more time to focus on their primary activity, be it reading, writing or translation.

When it comes to writing aid, the tools designed to provide this service can be conceived with different functionalities depending on the underpinning philosophy. The Dutch–French–English *Interactive Language Toolbox*, for instance, is primarily conceived as a language-learning tool and therefore it "does not correct the submitted text" but "only identifies syntactic and lexical patterns that may contain errors" in order to encourage the user "to reflect critically on his writing" (Verlinde 2011: 282-283).

Against this background, a distinction can be made between *detective*, *corrective*, and *predictive* writing assistants. The former refers above all to various types of spelling and grammar checker, but efforts are also being made to develop tools that can check other linguistic categories such as collocations; see Wanner et al. (2013). The main advantage of detective writing assistants is that they can call the users' attention to problems (and needs) of which they may not be aware. By contrast, corrective writing assistants, for example the one provided by Microsoft, do not only detect possible errors in the already written text but also come up with alternative solutions. Finally, predictive tools intervene directly in the writing process with suggestions, either on how to complete a word when the first letters have been typed, or on which word(s) could be next in the sentence. Many people know such interactive and predictive tools from their smartphones and tablets where they have the potential to speed up a writing process usually performed with only one or two fingers.

3. Information needs in L2 text production

There is a relatively big body of lexicographical literature dealing with the needs people may experience when writing in a second language as well as the response which dictionaries should provide to these needs (see for instance Rundell 1999, Tarp 2004, Bogaards 2005, Chon 2009, and Lew 2016). This literature contains valuable ideas which will be taken into account in the following reflections. Among the mentioned lexicographers there is, for instance, a general understanding that neither monolingual nor bilingual dictionaries are capable of meeting all the users' needs on their own. Only a combination of dictionaries can achieve this. But what combination? And which lexicographical data should these dictionaries offer in order to serve users engaged in L2 text production? An intent to answer these questions will be made in this section. The discussion will be based on the idea that user needs are first of all

determined by the situation in which they occur and then shaped by the relevant characteristics of the person who experiences them; see Fuertes-Olivera and Tarp (2014: 48-57). The starting point for a determination of these needs is therefore an analysis of the L2 writing situation and the different types of problem a writer may encounter in this situation.

If we exclude the already committed mistakes which the writer is not aware of, the most common problems and corresponding information needs are probably the ones included in the following Top Ten where 'word' is used as a generic term that also includes compounds:

Class of problem		→	Information needs
1	The writer knows an L2 word but cannot remember it.	→	L2 equivalents provided in an L1-L2 solution
2	The writer does not know the word to be used in L2.	→	L2 equivalents provided in an L1-L2 solution with meaning differentiation
3	The writer does not know a collocation to be used in L2.	→	L2 collocations provided in an L1-L2 solution
4	The writer knows an L2 word but is not sure if it can be used with the concrete meaning.	→	Explications
5	The writer knows an L2 word but is not sure if it can be used in a concrete context.	→	Stylistic, pragmatic, cultural or language-political information
6	The writer knows an L2 word but is not sure how it is spelled.	→	Information about orthography (or autocorrection)
7	The writer knows an L2 word but is not sure how it is inflected.	→	Information about inflected forms
8	The writer knows an L2 word but is not sure how to combine it with other words.	→	Explicit and/or implicit information about syntactic properties.
9	The writer knows an L2 word but is not sure how to construct collocations with this word.	→	Collocations with the L2 word in question
10	The writer knows an L2 word but wants to vary the language and use another word.	→	Synonyms and/or antonyms

Schema 1: Frequent problems and information needs in L2 writing

It is a matter of course that the above Top Ten does not give a full picture of all the complex problems and needs in connection with L2 writing, for example the ones related to phraseology and various sorts of fixed expressions. Yet it represents undoubtedly some of the most relevant problems and needs although they cannot be seen isolated from the concrete pair of languages in question. Gender is, for instance, not relevant in modern English but it certainly is in languages like German, Spanish and Italian. To this can be added other classes of problem which people may experience when writing L2 texts in other languages, especially the ones outside the Indo-European family.

It is worth noting the expression "is not sure" used in relation with various classes of problem (4–9). This expression covers two variants of the same problem, namely 1) that the writer does not know the linguistic item in question, and 2) that the writer knows it but just cannot remember it and therefore needs a *reminder*, a very frequent experience for L2 learners who typically master a much bigger passive than active vocabulary and grammar.

As can be observed, only problems of the three first classes listed above require a bilingual L1–L2 solution whereas the remaining problems require a solution based upon L2, i.e. either a monolingual L2 or a bilingual L2–L1 one. Inexperienced writers and learners at the beginner's level may typically encounter more problems of the first three classes than experienced writers and advanced learners, but all of them may from time to time experience all ten classes of problem. The traditional separation of learners into beginners, intermediate and advanced does therefore not seem to be relevant in this connection although it could be argued that children having problems in L2 writing may demand less complex information than adult writers; see Nomdedeu and Tarp (2017).

4. *Write Assistant*

In the following, we will discuss *Write Assistant* developed by Ordbogen A/S. This tool is aimed at giving instantaneous, high-tech response to the problems referred to in the previous section. As such, it is not designed as an independent text-processing program, but as an *application* which currently can only be used together with Microsoft's Office Package although nothing prevents it from being adapted to other types of software when needed. This means that the writer using the application simultaneously can benefit from the advantages of the Office Package, including the Spelling and Grammar Checker. *Write Assistant* is therefore not designed mainly with detective and corrective functions, but above all with a combination of *predictive* and *translative* functions as we will see below.

Once downloaded, the application can only be activated when Word is the front-most window. It will then appear as an icon in the centre of a Word document, an icon that has to be moved outside the document in order to get activated. From then on, it will work in all Office programs, including Outlook,

PowerPoint, Explorer, etc. In its current version (June 2017), *Write Assistant* is only available for Danish native speakers writing in English but the technology developed allows the incorporation of any new pair of languages within a few weeks provided the right empirical resources are at hand (see below). It is foreseen that the tool will be available in dozens of languages in the nearby future, and to this should be added a series of versions specifically adapted to the terminology of concrete companies. A description and analysis of the tool — including its philosophy, empirical basis, language model, functionality, usefulness, possibilities as well as limitations — is therefore relevant far beyond the Danish borders.

4.1 Underpinning philosophy and empirical basis

Write Assistant distinguishes itself from machine-translation programs that provide more or less adequate solutions as well as from stand-alone dictionaries and corpora that require the users' active decision to be consulted. The tool is not a "one-stop shop" (Bowker 2012: 381), but rather a *delivery-on-demand service*. It is not "a set of components which customers can mix and match according to their needs" (Rundell 2007: 50), but rather a set of components which the tool handles in order to provide *customized service* to its users according to their likely needs in each concrete case.

Write Assistant does not claim to meet all information needs occurring during the L2 writing process. The application is not designed to deal with *what* knowledge is transmitted during this process but only with *how* it is transmitted. If students, for instance, need information on the close relationship between Johann Wolfgang von Goethe and Alexander von Humboldt in order to write an essay on this subject, they will have to look for it elsewhere. Even so, and as its name suggests, *Write Assistant* is not conceived to provide solutions to the writers' communicative needs, but "only" *assistance*. The users of this product are still expected to play an active and decisive role and take full responsibility for the quality of the final L2 text. As an integrated and interactive application, the main purpose of *Write Assistant* is to offer instantaneous assistance which the writers can ignore or accept without wasting time on consulting external information resources, without too many interruptions in the writing process, and without losing their concentration and focus.

Write Assistant is based on big data taken in from two main empirical resources, namely a big L2 corpus and one or more digital dictionaries, among them at least one L1–L2 and another L2-based dictionary. When the first English edition is published (at the end of 2017), the corpus in question will be the British National Corpus with 100 million words. The reason why the application feeds on an existing corpus instead of using data directly from the Internet has to do with the quality of these data. Although continuously actualized and much bigger than any other corpus, Internet contains too much "noise" in the form of unedited texts and misspelled words. *Write Assistant* needs high-qual-

ity data in order to provide high-quality service, and therefore even existing corpora have to be further "cleaned" and items with less than ten occurrences deleted as a means to avoid as many spelling mistakes as possible.

Like the corpora, the dictionaries to be used by *Write Assistant* as data resources have to be available on a digital platform, preferably in the form of so-called Models T Ford that have been designed from scratch for the electronic media (Tarp 2011: 60). In the first edition to be published, the lookups will be made in the biscopal online Danish–English/English–Danish dictionary already handled by Ordbogen A/S and widely used by its subscribers. As such, it represents a good starting point although the incorporation of a monolingual English dictionary is also considered. However, and as will be discussed below, it cannot be excluded that a future need for optimizing the quality of *Write Assistant* may require the design of new, or partially new, lexicographical products that have their data adapted to the very specific requirements of this application.

In summary, it can be said that, apart from the technology developed as an indispensable precondition for any success, the quality of *Write Assistant* depends above all on the quality of the corpora and the dictionaries used as its empirical basis. At this point, all that is required to adapt the application to any pair of languages is the existence of big data in the form of digital corpora and dictionaries. If this empirical basis is at hand, *Write Assistant* can be made available in any language in a very short span of time.

4.2 Language model

In its current beta version, *Write Assistant* performs lookups in either an English *language model* or the biscopal Danish–English/English–Danish dictionary mentioned above; hence, it does not look up directly in the monolingual English corpus. The language model is the result of an automatic analysis and restructuring of the words contained in the corpus. It has been designed to receive *tuples* of up to four words and give each of them a score. We will exemplify this with the sequence "I was drawn forward with the prospect of employment" taken from Samuel Johnson's *The Plan of an English Dictionary* (1747). This sequence consists of a total of six 4-word tuples:

"I was drawn forward"

"was drawn forward with"

"drawn forward with the"

"forward with the prospect"

"with the prospect of"

"the prospect of employment"

Based on the frequency with which they occur in the corpus, each of these six tuples will be given a score that indicates the probability of the respective string of four words appearing together in the English language. The language model based on these principles has three functions: 1) the completion of words, 2) the prediction of the next word in the sentence, and 3) the prioritization of translation candidates.

The *completion of words* will be illustrated with a concrete example. If the writer, for instance, has typed "The man was married to a beautiful wo", the model will single out "to a beautiful" and "wo". These two strings are called *context* and *prefix*, respectively, where the latter should not be confused with the traditional linguistic term. By means of a sorted list of all the words contained in the language model, *Write Assistant* now queries all words beginning with the prefix "wo". This could be *work, woman, world, word*, etc. These possible completions of "wo" will then be added to the context "to a beautiful" resulting in the following 4-word tuples:

"to a beautiful work"

"to a beautiful woman"

"to a beautiful world"

"to a beautiful word"

etc.

Each of these tuples will be given a score from the model proportional to the number of occurrences in the corpus. The completions in the ten tuples that gets the highest score will be selected and presented to the user in a prioritized order. In the concrete example discussed here, it will be of no surprise that the most likely completion of "wo" is "woman". However, this should not be understood as the final solution, but just a suggestion, as the writer may want to express something else (e.g. dedicate a poem "to a beautiful world") and thus has to play an active role and take responsibility for the word eventually chosen.

If the language model cannot give a score to the 4-word tuple shown to it because it does not contain any such combination of words, the first of the four words will be cut out and the 3-word result presented to the model, and so forth. Hence, "to a beautiful woman" will be reduced to "a beautiful woman", then to "beautiful woman", and finally just to "woman". Each of these quadrigrams, trigrams, bigrams, and unigrams will be multiplied by a specific weight in order to get a score. It goes without saying that the shorter the tuple the smaller the weight used as multiplier.

The *prediction of the next word* in the sentence may, from a user's point of view, represent a welcome assistance. Yet from the programmer's point of view it is just a special case of the word completion explained above. To show this

we will take the example where the writer has typed "When I worked on the oil platform I had a terrible". The language model will now single out the context "had a terrible" and the prefix " " where the latter consists of an empty string. It will then take all the words from the sorted list mentioned above, add them to the context, and give each of the resulting 4-word tuples a score. Similar to the previous example, the additions in the ten tuples that have the highest score will be selected and presented to the user in a prioritized order. In this particular case, the ten words are *time*, *start*, *experience*, *impact*, *effect*, *accident*, *season*, *fight*, *relationship*, and *and*. One of these words may satisfy the writer's needs in a concrete context. If not, an alternative strategy has to be chosen, for example to type an L1 word. The language model will then ensure that the L2 equivalents offered to the user are furnished in a prioritized order.

The following example illustrates the *prioritization of translation candidates*: A writer has typed "The man was gift". The word *gift* is a Danish polyseme and will automatically be looked up in the Danish-English dictionary that contains a number of translation candidates such as *poison*, *venom*, *married*, *toxins*, etc. These words are now used to replace *gift* with the following result:

"The man was poison"
"The man was venom"
"The man was married"
"The man was toxins"
etc.

As was the case above, each of these 4-word tuples will then receive a score from the language model. Fortunately, "The man was married" is the one with the highest score, which suggests that *married* is the most likely translation of *gift* in this concrete context. Once given a score, the translation candidates will be presented to the user in the most likely order.

Hence, what has been generated is a *context-aware dictionary*, i.e. a completely new type of dictionary that marks a further step towards the ideal of a more personalized lexicographical product.

4.3 Functionality and examples

When the user opens a new document — or begins a new paragraph or sentence — with *Write Assistant* running, a small box will pop up with the ten most frequent starter words. If the writer wants to use one of these words, all he has to do is to click on it. The first word will then be marked in green (see Figure 1). The writer has now two options, i.e. either to use this word or to scroll down to the right word using the Down Arrow key. When one of the suggestions is marked in green, the normal function of the return key will be deactivated and this key can instead be used to enter the highlighted word

directly into the text, thus skipping one of the steps in the traditional information search process discussed in Section 1. Once the word has been added, *Write Assistant* will move forward so that the left edge of the prefix aligns with the left edge of the ten suggestions most likely to become the next word, and so forth (see Figure 2). This also takes advantage of the way the eye works.

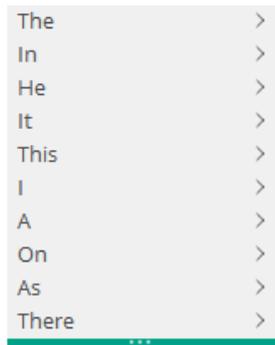


Figure 1: Starter words

If the writer does not opt for any of the suggested words and starts typing another one, a box with the ten most likely completions will pop up as explained in Section 4.2. Figure 2 shows the example with "The man was married to a beautiful wo" used previously. However, if a third letter is added and the prefix instead is "wom", this would narrow the field of possible candidates and raise the probability of having the required word among the ones with the highest score.



Figure 2: The tuple "to a beautiful wo" + completions

The application may also suggest a word with which the writer is not completely familiar, or not familiar at all. In this case, the only thing required in order to get additional assistance is to click the arrow right of the suggestion. *Write Assistant* will then perform an automatic lookup in the English–Danish dictionary and present the result in a new window containing a dictionary article that provides meaning explications, among others. This function can be exemplified with the sequence "I want to give a bidrag", where *bidrag* is a Danish polyseme with various English equivalents such as *donation*, *contribution*, *subscription*, etc. (see Figure 3).

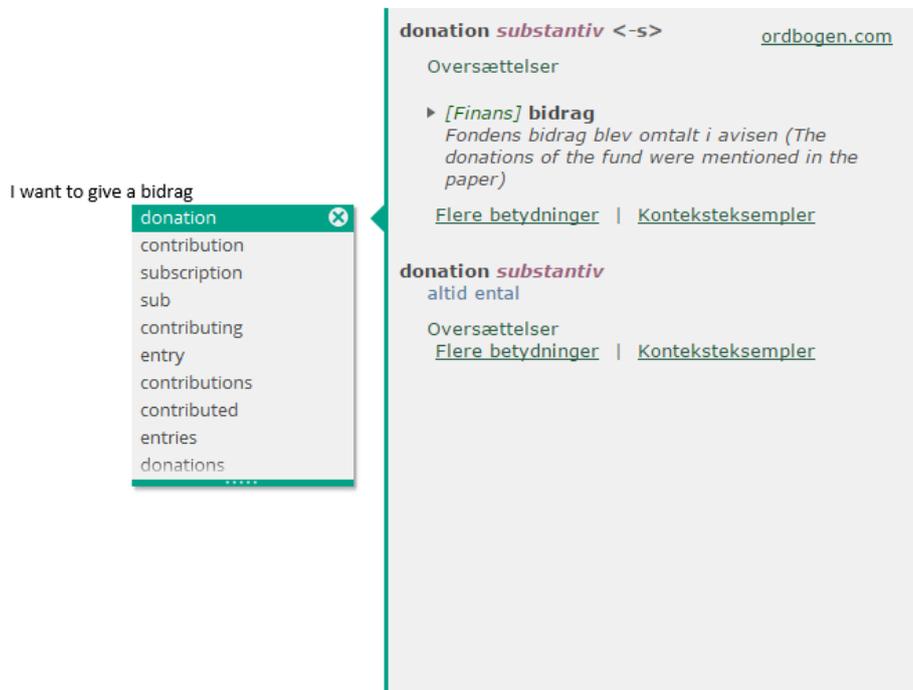


Figure 3: The tuple "to give a bidrag" + default article

The default article reproduced in Figure 3 furnishes only one sense of *donation*. However, it offers various options to the user who can click on *Flere betydninger* (More senses) or *Konteksteksempler* (Context examples) for additional data (or even on *ordbogen.com* to be referred to the specific article in the online dictionary). If the writer opts for *Flere betydninger*, he will get instantaneous access to more senses as shown in Figure 4.

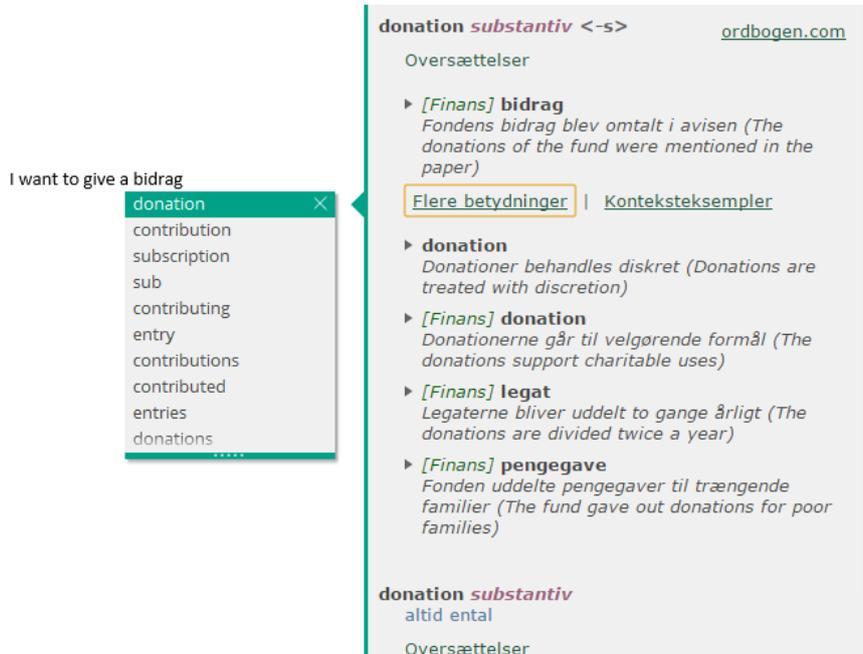


Figure 4: The tuple "to give a bidrag" + unfolded article

If the problem is that the writer does not know the meaning of the suggested word, or has doubts about it, the information need arising from this uncertainty has all chances of being met immediately. However, the problem may also be that the writer does not know how to use the suggested word together with other words. In this case, further assistance can be achieved by clicking on *Konteksteksempler*. The black data hitherto hidden under this metatext will then be visible in the form of various example sentences uploaded from the English–Danish dictionary and originally selected in a corpus (see Figure 5). This would allow the writer to detect a number of useful collocations, phrases, and syntactic properties, among others, although these data are only provided implicitly. It is worth noting that the metatexts *Flere betydninger* and *Konteksteksempler* are offered with hidden data as default, and that these data can only be unfolded through the user's decision to act. In this way, the increasingly problematic phenomenon of information and data overload, i.e. the "wall of text" that instantly turns people away from webpages, is reduced considerably; see Gouws and Tarp (2018).

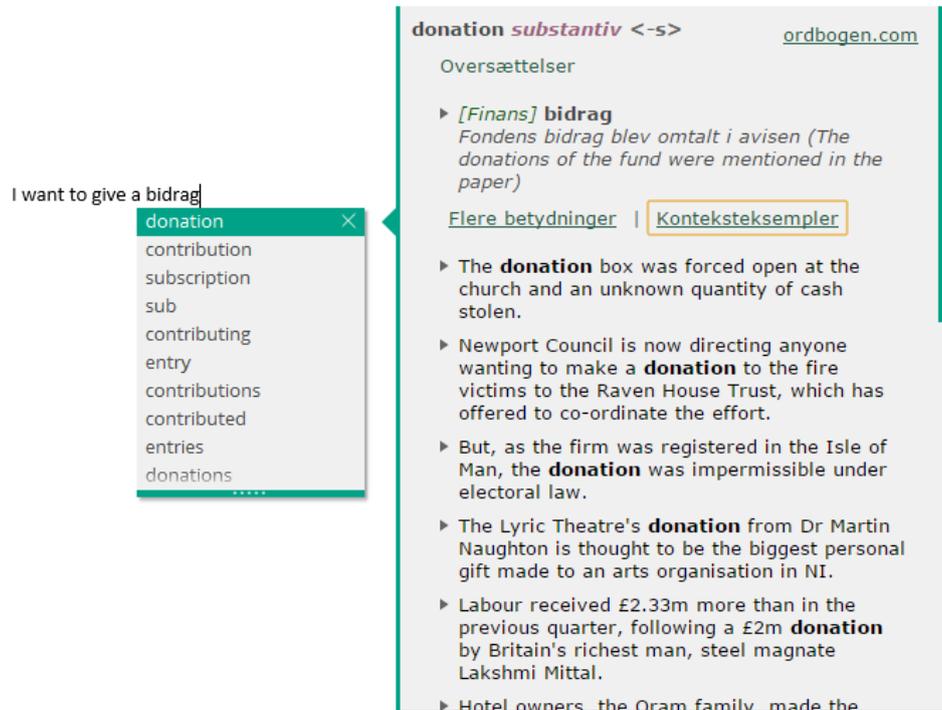


Figure 5: The tuple "to give a bidrag" + example sentences

Depending on their L2 proficiency level, people writing texts in this language will more or less often have to make use of their mother tongue in order to find or remember the L2 words to be used. Let us take the example where a Danish native speaker wants to say something nice about a woman and does not know, or is not sure of, the English word that can be employed to express this idea. He therefore types the following sequence: "She is very smuk". *Smuk* is a polysemous Danish word with a relatively big number of English equivalents, of which various are more or less synonymous. *Write Assistant* will immediately look up in the Danish-English dictionary and present ten of these equivalents to the writer in a specific, prioritized order (see Figure 6).



Figure 6: The tuple "She is very smuk" + equivalents

However, if the writer instead wanted to express something similar about a man, the words typed down could be: "He is very smuk". The application would then once more look up in the dictionary and suggest the ten most likely translations of *smuk* (see Figure 7).

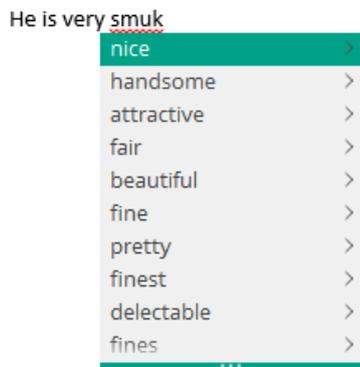


Figure 7: The tuple "He is very smuk" + equivalents

A comparison between Figure 6 and 7 provides another example of the context-awareness of the underlying dictionary. Although identical in this case, the ten equivalents suggested are not presented to the user in the same order. Some of these translation candidates are very close synonyms and the writer's use of one or another would not make a big difference. However, if we look at *beautiful* and *pretty*, these two words are presented as the most likely equivalents of *smuk* in relation to *she* whereas they are relegated to the fifth and seventh position in relation to *he*. By contrast, *nice* and *handsome* are the two highest ranked candidates in relation to *he*, whereas they are number three and seven in con-

nection with *she*. This difference is undoubtedly relevant inasmuch as it reflects real language use as it has been collected in the corpus. In any case, if the writer is still not sure which word to use, all that he has to do is to click the arrow right of the suggestion in order to consult the explications of the respective words and, if it is deemed necessary, click through to the example sentences available under *Konteksteksempler*.

From a philosophical point of view, it is interesting to observe how the relationship between man and machine changes when the user decides to click through to a dictionary article or click on one of the metatexts. Until then, the writer has only been typing letters and words whereas the tool has suggested possible solutions which the writer can ignore or accept with a simple movement of the fingers. Nothing else has been required from the writer. Although interacting with the tool, its user has basically played a passive role in terms of the assistance provided. By contrast, when the user clicks the arrow right of the suggestion, or click on the metatexts, then he turns into an active player that takes a conscious decision to look for further information. This situation is very much similar to the consultation of traditional stand-alone dictionaries, but with one important difference. Whereas the latter have to be taken down from the shelf or accessed on a separate website, the dictionaries integrated into *Write Assistant* are already there with an invitation to be consulted.

4.4 Completely or partially solved information needs

In this section, we will try to answer the probably most essential question, namely to which extent *Write Assistant* actually meets its users' information needs when they experience problems in relation to text production in a foreign language. As a reference we will take the ten typical information needs listed in Schema 1.

(1) When writers think they know an L2 word but cannot remember it, they will need an L1–L2 solution providing *L2 translations* of L1 words. In the previous section (Figure 6 and 7), we saw how the application fully meets this requirement, at least for simplex words. But what about compound words? In an assistant for Danes writing English texts, this word class does not constitute a big problem. Most Danish compounds are written as a single word, and as long as they are selected as lemmata in the L1–L2 dictionary, their English equivalents will automatically be presented to the users when required. However, as most English compounds consist of two or more single words with space in between, a slightly different technical solution would be needed if the tool was designed to assist English native speakers writing Danish.

(2) When writers do not know the English word to be used, they will need an L1–L2 solution providing L2 translations as well as *meaning differentiation*. This requirement is also fully met by *Write Assistant*, but in a way that differs from the traditional solution in dictionaries where meaning differentiators, as a rule,

are visible simultaneously. In this application the ten most likely equivalents are furnished in a prioritized order (a big advantage in comparison to traditional dictionaries), but meaning differentiation has to be accessed for each equivalent separately. In this case, the *explications* of the respective L2 words (now presented as equivalents) serve as meaning differentiators (see Figure 4).

(3) When writers do not know a collocation to be used in L2, they will need an L1–L2 solution furnishing *L2 translations* of L1 collocations. In its current version, *Write Assistant* does not offer such a solution. It works relatively well when the collocations are straightforward with similar bases and collocates according to Hausmann's (1985) 2-word collocation theory. If the Danish writer, for instance, types the verb *børste*, the application will suggest *brush* as the first translation candidate, and if the writer then continues with *mine tænder* (my teeth) there would be no big problem. But if L2 was Spanish where the equivalent collocation is *lavarme los dientes* (wash my teeth), there would be no direct solution. But even if there is a more or less similar L1 collocate this would not necessarily mean that the solution was easy. Let us take the Danish collocation *blande kort* (shuffle the cards) as an example. When the writer types the highly polysemous word *blande*, e.g. in the sequence "John was the next to blande", a number of possible English translations of *blande* will be proposed by the tool. However, *shuffle* will only pop up as the fifth most likely candidate after *mix*, *blend*, *mixed* and *dilute* (see Figure 8). The risk of choosing the wrong English equivalent would probably be high if the writer is not very conscious about the problem and looks for additional assistance in the integrated dictionary where *shuffle* is indicated as the right verb in connection with *cards*.

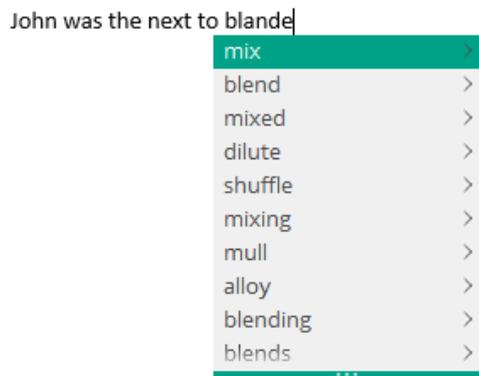


Figure 8: The tuple "the next to blande" + equivalents

(4) When writers know an L2 word but are not sure whether it can be used with the concrete meaning, they will need *explications* of the L2 words. This requirement is fully met as can be seen in Figure 9.

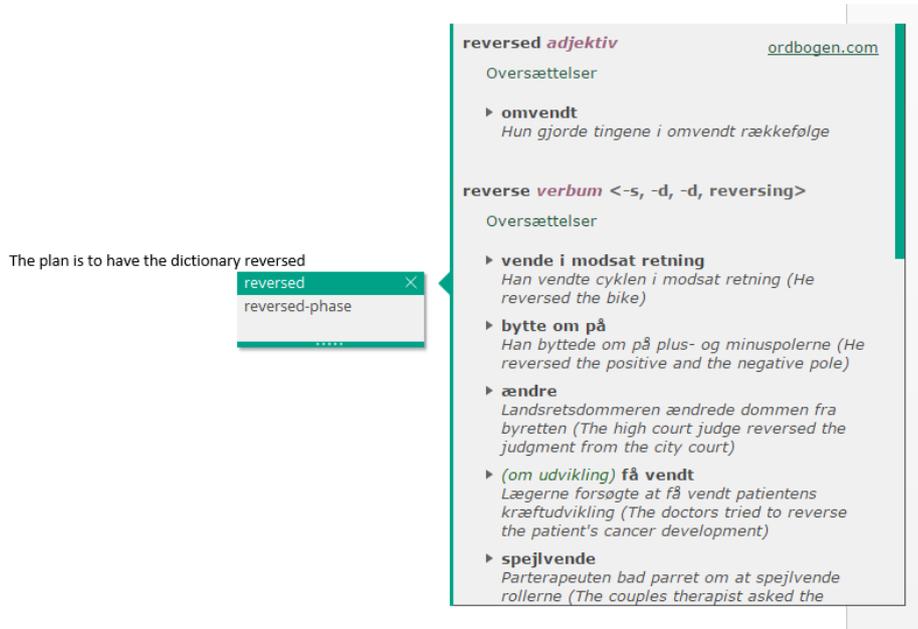


Figure 9: The tuple "have the dictionary reversed" + article

(5) When writers know an L2 word but are not sure whether it can be used in a concrete context, they will need *stylistic, pragmatic, cultural* or *language-political* information. This requirement is only met to a limited degree. The L2–L1 dictionary integrated into the tool does not offer explicit data of this type. But the information may, to a certain extent, be deduced from the example sentences available under *Konteksteksempler*. As a special case, *Write Assistant* will also be available in versions specifically adapted to concrete companies that have their own language policies in terms of the terminology to be used by their employees. These special versions of the application will therefore be designed to reflect and transmit the relevant language-political information.

(6) When writers know an L2 word but are not sure how it is spelled, they will need information about *orthography* or *autocorrection* when typing L2 words. This requirement is fully met. On the one hand, the tool offers correctly spelled completions once the first letters are typed; and on the other hand, it runs together with Microsoft's Grammar and Spelling Checker that will detect possible spelling mistakes and suggest corrections. Alternatively, the user could simply write an L1 word in order to get the rightly spelled L2 word.

(7) When writers know an L2 word but are not sure how it is inflected, they will need information about *inflected forms*. This requirement is fully met. The possible completions, next-words, and equivalents suggested by the application are all inflected in order to fit into the sentence, but only by statistics in the

language model, i.e. without applying grammatical rules. If additional information is required, the full inflectional pattern of the word in question can be accessed in the L2–L1 dictionary (see Figure 9). In any case, Microsoft's Grammar and Spelling Checker will also be there as a more or less effective grammatical safety net.

(8) When writers know an L2 word but are not sure how to combine it with other words, they will need explicit and/or implicit information about *syntactic properties*. This is one of the most complex issues in lexicography. Like almost all existing dictionaries *Write Assistant* does not come up with a completely satisfactory response but it provides at least some assistance. As can be seen in Figure 10, the L2–L1 dictionary offers some syntactic "rules" that are subsequently exemplified. Additional assistance can also be found in the example sentences (in Figure 5), but the users have to deduce the underpinning rules themselves when they want to construct their own sentences. However, neither the explicit nor the implicit data are addressed to specific senses, a weakness that leaves room for mistakes. What is required to satisfy the needs of different user types is a combination of explicit data (rules) and implicit data (example sentences), all of it addressed to specific senses of the word. This requirement is only partially met by the application in its current version.

The new government benefited

benefited

benefit *verbum* <-s, -ed el. -ted, -ed el. -ted, -ing el. -ting> ordbogen.com

Oversættelser

- ▶ **gavne**
Det her gavner ikke virksomheden (This does not benefit the company)
- ▶ **være til fordel for**
Hun mente, at det måtte være til fordel for ham (She believed that it would benefit him)
- ▶ **benefit somebody gavne nogens sag**
Det her gavner hans sag (This benefits him)
- ▶ **benefit from nyde godt af**
Hun nyder godt af hans rigdom (She benefits from his wealth)
- ▶ **benefit from drage fordel af**
Hvordan drager vi fordel af deres konkurs? (How can we benefit from their bankruptcy?)
- ▶ **benefit from få gavn af**
Hvordan får firmaet gavn af det her? (How does the company benefit from this?)

[Konteksteksempler](#)

Figure 10: The tuple "The new government benefited" + article

(9) When writers know an L2 word but are not sure how to construct collocations with this word, they will need *collocations* with the L2 word in question. This requirement overlaps with the one discussed under Point 3. The main difference is that the writers in this case know at least one of the words composing the collection. The problem could therefore be solved by consulting an L2-based dictionary that contains collocations. As can be seen in Figure 11, the dictionary integrated into the tool does offer some explicit collocations apart from the ones that can be deduced from the example sentences. Besides, if the writers know the first word in the collocation and type this word, then it is possible that the second word in the collocation will be among those suggested by the tool, at least if it is a frequently used collocation. If this is not so, or if it is only the second word that is known to the writer, then an L1–L2 solution is required as in the example with *shuffle the cards* discussed above. Such a solution is currently not provided either. Hence, *Write Assistant* has not yet come up with a completely satisfactory response to the challenges posed by collocation writing in L2.

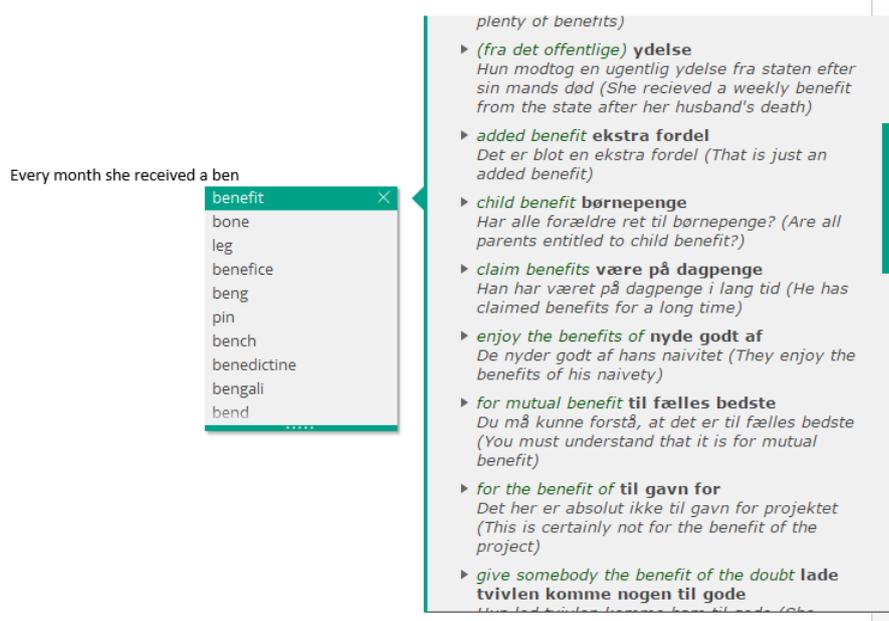


Figure 11: The tuple "she received a ben" + article

(10) When writers know an L2 word but wants to vary the language and use another word, they will need *synonyms* and/or *antonyms*. This requirement is only met to a very modest degree. When the tool suggests the possible next words to be written in the sentence, there are sometimes synonyms among the proposed words but this solution can, by definition, never be systematic inasmuch as the words presented to the user by the language model do not depend

on their semantic structures but on the likeliness of the respective tuples to be found in L2. Thus, a more appropriate solution would be that the underlying dictionary offered synonyms and antonyms, a service it could easily render.

In short, it can be concluded that the first version of *Write Assistant* fulfils half of the ten requirements to a tool designed to assist L2 writing, whereas it meets the remaining requirements to different degrees. The limitations can be due to the technology developed, the design of the interface, or the quality of the empirical basis. Among the five requirements that have not found a satisfactory solution, only those related to L2 collocations can be explained by technological limitations. The challenge is therefore to further develop *Write Assistant* to the extent that it does not only complete or predict words based on the previous 3-word context, but also looks backwards and proposes corrections of already typed words in order to offer the correct collocations. Such a solution would also be relevant to multiword compounds and terms, and would, simultaneously, raise the context-awareness of the integrated dictionary.

The remaining problems detected above — i.e. those related to the treatment of synonyms, antonyms, syntax, style, pragmatics, and culture — require a different solution inasmuch as they have to do with the design of the visualized dictionary articles and as well as the content of the underlying dictionary used as empirical resource. These problems are all of a lexicographical nature and will briefly be discussed in Section 5.

4.5 Usefulness and writing flow

In the introduction to this article we referred to the traditional information-search process as it has been described by Bergenholtz et al. (2015). We stressed the need for a tool that could assist L2 writing in such a way that some of the phases and steps in this process could be shortened or even skipped in order to save time and maintain the writing flow without losing the focus on the message to be transmitted. The analysis which we subsequently made of *Write Assistant* indicates that it is a qualified candidate to become a tool with such properties. As can be expected from beta versions of complex digital products, it has been born with some childhood diseases but they all seem to be curable.

The tool appears to be very simple and easy to use from a functional point of view. Its users can work with their normal keyboard and navigate among the suggested words and perform supplementary lookups in the integrated dictionaries by means of either the mouse or the Alt and Arrow keys. Following the same philosophy, the return key has been reprogrammed so it can now be used to enter words directly into the text once they are marked in green. All this creates the technical conditions for an improved writing flow.

Writers may easier become aware of their information needs when they start typing and the application immediately comes up with suggestions for completions and next words. If they are happy with one of these suggestions,

the only action required from them is to mark the word in question and press the return key. The same is the case when they type an L1 word and instantly get a number of L2 candidates. If further information is required about any of the suggested words, a simple finger movement is all it takes to visualize a dictionary article right away. Once the result is deemed to be satisfactory and the return key pressed, the word in question will automatically be added to the text and solve the problem that originally gave rise to the information need.

As can be seen from the presentation above, some of the traditional phases and steps in the information-search process have become less complex while others have been skipped because the need to access external information resources has been reduced considerably. In fact, the users of the application can go a long way without consulting such resources. They can save precious time and concentrate on the message to be written.

By contrast, while *Write Assistant* undoubtedly makes the writers' job easier, it does not take any responsibility away from them. The users of this tool are still expected to critically evaluate the retrieved information, decide which words to use, and take responsibility for the final text. The users, and nobody else, are the sole authors of the texts produced with the assistance of the tool. The latter is only their handy assistant, but not their co-author.

5. Challenges to lexicography

The existence of computers and huge databases, the programming of the language model, and the functional and user-friendly design make up the necessary technological and technical conditions for an L2 write assistant like the one described here. However, these conditions do not by themselves guarantee the quality of the application. Once they have been created, the quality of the product depends first and foremost on the quantity and quality of the empirical basis from where the data are taken in, that is, the corpus and the dictionaries.

The requirement to the corpus is basically that it should be relatively big, well-composed, up to date, and sufficiently "clean" in order to reduce the risk of spelling mistakes. But what about the dictionaries? Many of the problems detected in Section 4.4 have to do with the underlying dictionaries. A more detailed analysis of the lexicographical data would confirm this tendency. The criticism of the quantity of L1 words with L2 equivalents, the quantity and quality of the latter, the quality of the definitions of the L2 words, the existence or not of other relevant data categories as well as the quality of these — all this is to a large extent a criticism of the dictionaries used as empirical resources.

It is not difficult to find (or compile) dictionaries that contain most of the missing data categories discussed in Section 4.4. With the possible exception

of appropriate cultural notes, the other items — i.e. stylistic and pragmatic labels, synonyms and antonyms, collocations as well as explicit and implicit syntactic data — are already available in many existing dictionaries, at least in a language like English. However, all these items must be adapted to the specific technical and functional requirements of the tool, an adaptation that may have consequences for the overall dictionary concept and the production methods in new dictionary projects. All this poses new challenges to lexicography.

The usage of already existing dictionaries may create some problems that affect the overall quality of the tool. Today, most dictionaries are still conceived to provide assistance to both text production and text reception without the necessary differentiation. The design and presentation of the different lexicographical items used for both purposes may therefore represent a sort of compromise which is not necessarily the most adequate for L2 writing. Besides, there may be other problems in terms of scopus and directionality. The dictionary used to support the *Write Assistant* is *biscopal*, i.e. it consists of a set of bilingual dictionaries in both language directions with the special requirement that all L2 words in the L1–L2 dictionary must have an equivalent entry in the L2–L1 dictionary. The biscopal solution seems to be the most appropriate in connection with L2 writing, and it is also the one recommended for this purpose by most modern lexicographers (see for instance Lew and Adamska-Salaćiak 2015). Even so, it is important that the dictionary is designed from scratch as a *monodirectional* dictionary. The reason for this fundamental conceptual requirement is that every single version of *Write Assistant* has only one user group in terms of mother tongue, namely native speakers of L1, and that any bidirectionality may interfere inconveniently with the lexicographical data presented to this specific user group.

Here we will briefly discuss *meaning items* as just one example of how lexicographical data that have been taken in from a set of not fully adapted bilingual dictionaries may not be the best solution. In traditional bilingual dictionaries there are two main classes of meaning item in terms of their purposes: 1) the one that is used to differentiate between L2 equivalents in the L1–L2 dictionary, and 2) the one that is used to explain the meaning of an L2 word in the L2–L1 dictionary. Different techniques are applied in each case. In L2–L1 dictionaries, where focus is on text reception, the items most frequently used to explain the meaning of L2 words are L1 equivalents, whereas the items used to differentiate between L2 equivalents in L1–L2 dictionaries are typically L1 synonyms and paraphrases.

In Section 4.4, we saw how the same data category played the role as both meaning explanation and meaning differentiator. This double role may create some inconveniences if it has not been foreseen and taken into account in the moment when the concept for the underlying dictionaries was decided. Figure 12 and 13 illustrate the problem.

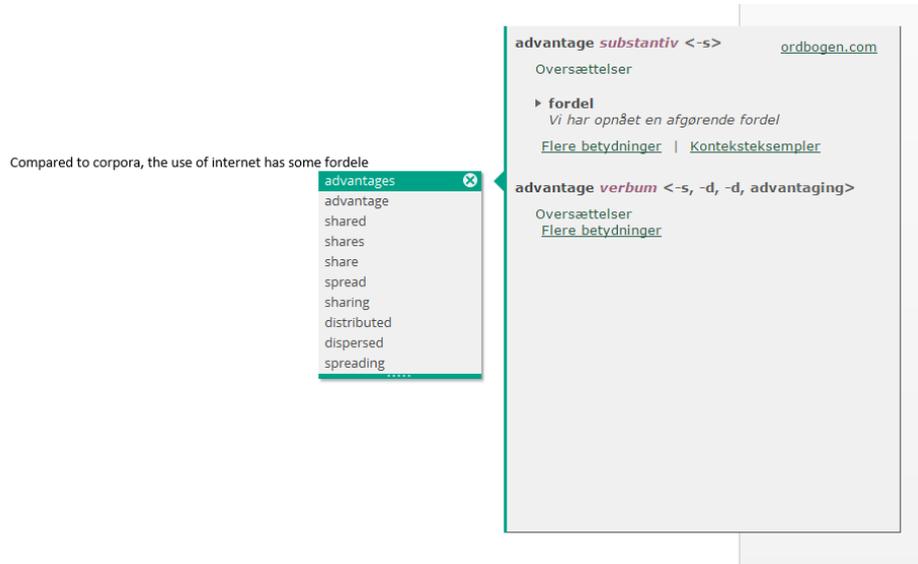


Figure 12: The tuple "Internet has some fordele" + article

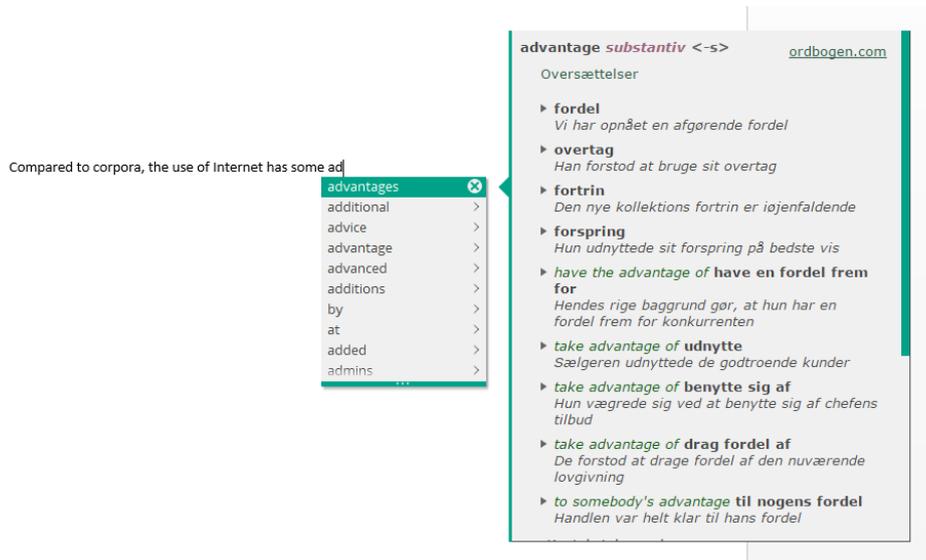


Figure 13: The tuple "Internet has some ad" + article

Although taken from the same underlying dictionary, the two default articles reproduced in Figure 12 and 13 are different because only one meaning item is needed as a *differentiator* in the former, whereas several items are required to

explain the meaning of *advantages* in the latter. In both cases the meaning items consist of an L1 word exemplified by an L1 sentence (e.g. **fordel** — *Vi har opnået en afgørende fordel*). This type of item may be useful when the writers have doubts about the meaning of an L2 word. But they will probably get more confused than enlightened when an L1 word (in this case *fordel*) is used to differentiate between its own equivalents, although the subsequent example sentence may remedy the problem to a certain extent. A different solution is therefore required, preferably a short L1 definition that would be helpful both as an explanation and a differentiator.

Meaning items are no exception. Other classes of lexicographical data also need to be scrutinized and adapted to the special requirements of *Write Assistant* and similar tools as a precondition for optimizing the quality of the service provided.

6. Perspectives

In the first half of 2017, *Write Assistant* was tested among a small group of Danish upper high school students, and it was also demonstrated to English teachers and students at various Chinese universities. The feedback was generally very positive in terms of its overall usefulness but different opinions were expressed concerning the possible consequences for foreign-language learning. For instance, fear was expressed that future language students may become too dependent on the tool. This may be so. Similar fear was voiced when the calculator was introduced in math teaching. Today it is obvious that many people are highly dependent on this tool, but it is also a fact that it allows them to perform more complex calculations than ever before without committing too many mistakes.

The use of *Write Assistant* and similar tools requires consciousness of the role of man and machine in modern communication. It is still man who is the sole responsible for both the content and form of the message to be written. Technology is only there to assist, not to take over. If this is understood, tools like *Write Assistant* will definitely be helpful in L2 writing. By contrast, the students who currently continue to use machine translation uncritically in spite of repeatedly being warned by their teachers, will probably not be the ones who benefit mostly from these tools.

There is little doubt that the new technology is there to stay. High-tech tools designed to assist the writing, reading and translation of texts will be an integrated part of our lives in the years to come. People will become increasingly dependent on them whether we like it or not. Lexicography can either adapt to this reality or die. Lexicographers are therefore challenged, not only to raise new questions and possibilities, but also to regard old questions from a new angle. There is no perspective in transforming the discipline into a Knight of the Woeful Countenance.

Acknowledgement

Thanks are due to the China University of Mining and Technology (Beijing) for supporting the work on this article.

References

- 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.
- Bogaards, P.** 2005. Dictionaries and Productive Tasks in a Foreign Language. *Kernerman Dictionary News* 13: 20-23.
- Bowker, L.** 2012. Meeting the Needs of Translators in the Age of e-Lexicography: Exploring the Possibilities. Granger, S. and M. Paquot (Eds). 2012. *Electronic Lexicography*: 379-397. Oxford: Oxford University Press.
- Chon, Y.V.** 2009. The Electronic Dictionary for Writing: A Solution or a Problem? *International Journal of Lexicography* 22(1): 23-54.
- Einstein, A. and L. Infeld.** 1938. *The Evolution of Physics*. New York: Simon & Schuster.
- Fuertes-Olivera, P.A.** 2016. A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. *International Journal of Lexicography* 29(2): 226-247.
- 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.** 2018. Information and Data Overload in Lexicography. *International Journal of Lexicography*. [Advance publication: <https://academic.oup.com/ijl/advance-articles>.]
- Granger, S. and M. Paquot.** 2015. Electronic Lexicography Goes Local: Design and Structures of a Needs-driven Online Academic Writing Aid. *Lexicographica* 31(1): 118-141.
- Hanks, P.** 2010. Lexicography, Printing Technology, and the Spread of Renaissance Culture. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*, Leeuwarden, 6–10 July 2010: 988-1016. Ljouwert: Fryske Akademy.
- Hausmann, F.J.** 1985. Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. Bergenholtz, H. and J. Mugdan (Eds.). 1985. *Lexikographie und Grammatik*: 118-129. Tübingen: Max Niemeyer.
- Johnson, S.** 1747. *The Plan of an English Dictionary*. <https://andromeda.rutgers.edu/~jlynch/Texts/plan.html>. Accessed on 31 May 2017.
- Lew, R.** 2016. Can a Dictionary Help You Write Better? A User Study of an Active Bilingual Dictionary for Polish Learners of English. *International Journal of Lexicography* 29(3): 353-366.
- Lew, R. and A. Adamska-Salaciak.** 2015. A Case for Bilingual Learners' Dictionaries. *ELT Journal* 69(1): 47-57.
- Nesi, H.** 2015. The Demands of Users and the Publishing World: Printed or Online, Free or Paid for? Durkin, P. (Ed.). 2015. *The Oxford Handbook of Lexicography*: 579-589. Oxford: Oxford University Press.
- Nielsen, S.** 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.
- Nielsen, S.** 2017. Lexicography and Interdisciplinarity. Fuertes-Olivera, P.A. (Ed.). 2017. *Routledge Handbook of Lexicography*: 93-104. London: Routledge.

- Nomdedeu Rull, A. and S. Tarp.** 2017. Hacia un Modelo de Diccionario en Línea para Aprendices de Español como LE/L2. *Journal of Spanish Language Teaching*. [To appear.]
- Paquot, M.** 2012. The LEAD Dictionary-cum-writing Aid: An Integrated Dictionary and Corpus Tool. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 163-185. Oxford: Oxford University Press.
- Robinson, H., A. Wysocka and C. Hand.** 2007. Internet Advertising Effectiveness. The Effect of Design on Click-through Rates for Banner Ads. *International Journal of Advertising* 26(4): 527-541.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-53.
- Rundell, M.** 2007. The Dictionary of the Future. Granger, S. (Ed.). 2007. *Optimizing the Role of Language in Technology-enhanced Learning*: 49-51. <https://hal.archives-ouvertes.fr/hal-00197203/document/>. Accessed on 31 May 2017.
- Rundell, M.** 2010. What Future for the Learner's Dictionary? Kernerman, I. and P. Bogaards (Eds.). 2010. *English Learners' Dictionaries at the DSN 2009*: 169-175. Jerusalem: Kdictionaries.
- Rundell, M. and A. Kilgarriff.** 2011. Automating the Creation of Dictionaries: Where Will It All End? Meunier, F., S. de Cock, G. Gilquin and M. Paquot (Eds.). 2011. *A Taste for Corpora. In Honour of Sylviane Granger*: 257-282. Amsterdam/Philadelphia: John Benjamins.
- Tarp, S.** 2004. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.
- 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.** 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 17-32. Bern: Peter Lang.
- Tarp, S.** 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 54-70. London/New York: Continuum.
- Tarp, S.** 2017. Lexicography as an Independent Science. Fuertes-Olivera, P.A. (Ed.). 2017. *Routledge Handbook of Lexicography*: 19-33. London: Routledge.
- Verlinde, S.** 2011. Modelling Interactive Reading, Translation and Writing Assistants. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 275-286. London/New York: Continuum.
- Verlinde, S., P. Leroyer and J. Binon.** 2010. Search and You Will Find. From Stand-alone Lexicographic Tools to User Driven Task and Problem-oriented Multifunctional Leximats. *International Journal of Lexicography* 23(1): 1-17.
- Verlinde, S. and G. Peeters.** 2012. Data Access Revisited: The Interactive Language Toolbox. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 147-162. Oxford: Oxford University Press.
- Wanner, L., S. Verlinde and M. Alonso Ramos.** 2013. Writing Assistants and Automatic Lexical Error Correction: Word Combinatorics. Kosem, I., J. Kallas, P. Gantar, S. Krek, M. Langemets and M. Tuulik (Eds.). 2013. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17-19 October 2013, Tallinn, Estonia*: 472-487. Ljubljana/Tallinn: Institute for Applied Slovene Studies/ Eesti Keele Instituut.

Approaches to the Treatment of Zero Equivalence in a Bilingual Dictionary

Marjeta Vrbinc, *Department of English, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia* (marjeta.vrbinc@ff.uni-lj.si)

and

Alenka Vrbinc, *Faculty of Economics, University of Ljubljana, Ljubljana, Slovenia* (alenka.vrbinc@ef.uni-lj.si)

Abstract: The article discusses the treatment of zero equivalence in an English–Slovene dictionary (ESD). The absence of equivalents in the TL is marked by two symbols: Ø (a complete absence of any equivalent) and # (equivalence at the level of the entire message rather than at word level). Sixty-five lemmata in the ESD contain a slashed zero, a hash or both, but one and the same entry can contain more than one symbol. Detailed results are presented by parts of speech of the lemmata, starting with a numerical analysis, which is followed by analysis of the content of illustrative examples. Then follows a detailed discussion of lemmata expressing pragmatic meaning in the SL, lemmata with lexico-grammatical, grammatical and lexical differences between the SL and the TL as well as lemmata with a number of SL senses included under one sense in the ESD. In the ESD, the problem of zero equivalence is mostly solved by the inclusion of translated examples of use. Another method employed in the ESD is the use of short descriptions of the function of the lemma or one of its senses. In this way, a more successful and thorough retrieval of information can be achieved with every look-up.

Keywords: BILINGUAL DICTIONARY, CONTEXT, COTEXT, DESCRIPTIONS OF FUNCTIONS, EXAMPLES OF USE, GRAMMATICAL DIFFERENCES, LEXICAL DIFFERENCES, LEXICO-GRAMMATICAL DIFFERENCES, PRAGMATIC SENSES, ZERO EQUIVALENCE

Opsomming: Benaderings tot die hantering van zero-ekwivalensie in 'n tweetalige woordeboek. Die hantering van zero-ekwivalensie in 'n Engels–Slovenese woordeboek (ESD) word in hierdie artikel bespreek. Die afwesigheid van ekwivalente in die doeltaal (DT) word aangedui deur twee simbole: Ø ('n totale afwesigheid van enige ekwivalent) en # (ekwivalensie op die vlak van die volle boodskap eerder as op die vlak van 'n woord). Vyf-en-sestig lemmas in die ESD bevat 'n nul met 'n skuinsstreep daardeur, 'n hutsmerker of albei, maar een inskrywing kan meer as een simbool bevat. Die resultate word uitvoerig voorgehou volgens die woordsoort van die lemmas, en begin met 'n syferkundige ontleding, gevolg deur 'n ontleding van die inhoud van die toeliggende voorbeelde. Daarna volg 'n gedetailleerde bespreking van lemmas wat 'n pragmatiese betekenis uitdruk in die BT, lemmas met leksiko-grammatikale, grammatikale en leksikale verskille tussen die BT en die DT, sowel as lemmas met 'n aantal BT-betekenisgegroeper onder een betekenisonderskeiding in die ESD. Die probleem van zero-ekwivalensie word meestal in die ESD opgelos deur die insluiting van vertaalde gebruiksvoorbeelde. 'n Ander metode wat in die

ESD gebruik word, is die gebruik van kort beskrywings van die lemma se funksie of van een van sy betekenisse. Op hierdie manier kan 'n meer suksesvolle en deeglike onttrekking van inligting gedoen word tydens elke opsoeksessie.

Slutelwoorde: TWEETALIGE WOORDEBOEK, KONTEKS, KOTEKS, BESKRYWINGS VAN FUNKSIES, GEBRUIKSVORBEELDE, GRAMMATIKALE VERSKILLE, LEKSIKALE VERSKILLE, LEKSIKO-GRAMMATIKALE VERSKILLE, PRAGMATIESE BETEKENISONDERSKELDINGS, ZERO-EKWIVALENSIE

1. Introduction

The most salient element of a lexicographic description, regardless of the type and scope of a dictionary and its intended users, is the semantic component. Bilingual lexicographers are therefore expected to find equivalents in the target language (TL) that correspond semantically to the source language (SL) lexical items not only in one particular context but more universally (Adamska-Sałaciak 2010: 388; Atkins 1992/1993: 44f). Lexicographers, however, often come across cases when they fail to find suitable equivalents. This can happen with any lexical item; this is why the lexicographers must try to find and use means other than lexical equivalence (Zgusta 1971: 323). The provision of dictionary equivalents in the TL often depends on co-text: i.e., the linguistic environment of a word, or context, i.e., the non-verbal environment in which a word is used. There is a strong relationship between language and context, and seeking equivalence of meaning is in fact seeking equivalence for the situational context (Hu 2010: 324). It should be stressed that carefully selected co-text or context provided in a mono- as well as in a bilingual dictionary in the form of illustrative examples plays a very important role, since examples disambiguate and/or specify the meaning of the lexical item in question (Zgusta 1971: 337), and "any semantic phenomenon, whether in the field of designation, connotation, or the range of application can be clarified by means of examples" (Zgusta *ibid.*: 340).

When talking about equivalents in two languages, we actually refer to what Adamska-Sałaciak (2010: 389) explains as a particular sense of a lexical item X in the SL being equivalent to a particular sense of a lexical item Y in the TL (for more on the problem of equivalence see Adamska-Sałaciak 2010, 2011; Atkins and Rundell 2008: 468; Wiegand 2005). The relation between the SL lexical item and the TL lexical item is regarded as the equivalent relation (Gouws 2002: 195-196). Equivalent relations are generally of three types, which are variously referred to as absolute equivalence, partial equivalence and explanatory equivalence by Zgusta (1971: 312-325); full equivalence, partial equivalence and non-equivalence by Wiegand (2002); as full equivalence (congruence), partial equivalence (divergence) and zero equivalence (surrogate equivalence) by Gouws (2002: 196); and multiple equivalence, zero equivalence and partial equivalence by Šipka (2015: 51). Adamska-Sałaciak (2011: 4), however, enumerates four varieties of equivalence: cognitive, translational, explanatory and

functional. According to Gouws (2002: 196), full equivalence implies that the SL and the TL lexical items are equivalent lexically, pragmatically and semantically, which means that the TL lexical item can replace the SL lexical item in all contexts and uses. From the lexicographic perspective, full equivalence is considered relatively unproblematic, but as Zgusta (1971: 312) points out, absolute equivalents are quite rare. The equivalent relation that is most common in bilingual dictionaries is partial equivalence, where the semantic component of the dictionary entry consists of several TL equivalents that cover the entire spectrum of meaning of the SL item (cf. Zgusta *ibid.*: 315). These equivalents usually share certain semantic components, i.e., they are considered to be partial synonyms. Partial equivalence is characterized by the fact that the "polysemy structure of a source language lexical item does not correspond with that of the target language on the systemic level" (Wiegand 2002: 243). Gouws (2002: 197-199) further distinguishes between lexical divergence, semantic divergence and polydivergence. The last and the most intriguing semantic relation is that of surrogate or zero equivalence, characterized by a lack of equivalent in the TL. Gouws (2002: 200) points out that "[s]urrogate equivalents belong to different categories and their choice is determined by the nature of the lexical gap". Among solutions which can be adopted in the case of zero equivalence, Svensén (2009: 261, 274-275) enumerates direct borrowing, loan translation, new coinage, encyclopedic explanations, etc., whereas Šipka (2015: 52-53) stresses the importance of explanatory glosses combined with adopting an SL word in lexicography and possibly translation and cultural notes in language teaching. Since zero equivalence in the case of culture-specific words can be solved in a bilingual dictionary in the ways mentioned by Svensén and Šipka, culture-specific words are not the subject of our research.

Apart from zero equivalence in culture-specific lexical items, other cases of zero equivalence can be found. Although this may happen with any lexical unit, Zgusta (1971: 323-325; 1984: 149) enumerates other types of lexical items that often lack equivalents in other languages: onomatopoeia, interjections, functional words, particles and similar lexical units. Here, functional or grammatical words should be especially highlighted, since these often prove to be problematic when trying to establish equivalence. This can be explained by the distinct function performed by grammatical words in a sentence and by the fact that they are used differently in different languages. Among the grammatical words that are problematic from the dictionary-making perspective, Atkins and Rundell (2008: 164-165) enumerate the following: prepositions, conjunctions, pronouns (demonstratives, possessives, quantifiers), auxiliary verbs and the modals, determiners including the definite and indefinite articles, numerals, negatives, and predeterminers. Many of these grammatical words are also identified in our study as being difficult to treat in an English-Slovene dictionary. This can be explained by the fact that no conceptual content can be established in the TL, but as Šipka (2015: 69) points out "the operation that is performed by a lexeme in the SL can be rendered in the TL by some kind of

grammatical manoeuvre or any other non-lexical means". He therefore stresses that in this case, we can talk about zero equivalence at the juncture of lexical and grammatical elements.

The aim of this study is to investigate how to compensate for absent equivalents in the TL in a decoding English–Slovene dictionary (hereafter referred to as ESD) which is in its final stages of completion. Since the most important function of a bilingual dictionary is to offer equivalents that can be used in translation from the SL into the TL, special attention is paid to different methods employed by the lexicographers compiling a bilingual dictionary to provide acceptable solutions when dictionary equivalent(s) do not exist in the TL as well as to those cases which are solved neither by a descriptive equivalent nor by a loan word. The ESD contains many innovative features that are untypical of existing English–Slovene dictionaries. Among the innovations, two symbols, i.e., Ø and #, should be mentioned, both of them being employed to mark the absence of equivalents in the TL. The importance of explicitly marking zero equivalence in a bilingual dictionary is also stressed by Wiegand (2002: 248-249), who proposes the use of the symbol "0". It should, however, be stressed that cases of true zero equivalence are rare in the ESD, because the lexicographers tried to mark the fewest possible lemmata or their senses with the symbol Ø indicating a complete absence of any equivalent, but instead used the hash sign. The symbol # implies no equivalence at the word level, but if the untranslatable SL lexical item is used in an example illustrating its use, be it a sentence or a truncated phrase, it can be rendered into the TL, which means that equivalence is reached at the level of the entire message. This clearly indicates the importance of the role played by co-text and context in retrieving semantic information on the SL lexical items.

2. Methodology

The ESD investigated in this research contains 53,233 lemmata and 16,274 secondary lemmata (i.e., phraseological units included in the idioms section and multi-word verbs included in the phrasal verbs section) and belongs to the group of more comprehensive bilingual English–Slovene dictionaries; at the same time, it is the only pedagogically- and didactically-oriented bilingual dictionary compiled so far in Slovenia.

Since the study focused on zero equivalence, the entire dictionary was taken as a base for extracting cases of zero equivalence. Zero equivalence is marked in two ways: with a slashed zero (Ø), which indicates a complete absence of a dictionary equivalent in the TL; and with a hash sign (#), which indicates the absence of equivalence at word level but not at the level of the entire message.

The material collected was first analysed as to the part of speech of the lemmata to see whether any part of speech stands out as regards zero equivalence. Each lemma or sense of the lemma with zero equivalence was then

investigated more thoroughly, with the aim of identifying any potential features that may be considered a reason for zero equivalence and possible solutions offered by the lexicographers. Special attention was also paid to examples of use and their translations into the TL. Moreover, a content analysis of examples of use was made to see whether these examples share any features that could potentially be regarded as a reason for zero equivalence.

3. Results

The analysis of the material gathered showed that 65 lemmata in the ESD contain a slashed zero, a hash or both. The total number of hash signs used in the ESD is 92, but closer observation shows that one and the same entry can contain more than one hash, which means that the total number of lemmata with a hash sign is lower, since hash signs are found in 41 lemmata. The same holds true for the slashed zero, which can be found in 29 lemmata, but one lemma can contain one or more senses marked with a slashed zero, thus the total number of slashed zeros is 40. Interestingly, five lemmata (*of, out, 's, the* and *up*) contain senses marked with both signs.

Part of speech	No. of lemmata with #	No. of lemmata with Ø
Verb	12	9
Noun	11	3
Adjective	7	0
Preposition	5	2
Adverb	2	7
Pronoun	1	4
Article	0	3
Others	3	1
TOTAL	41	29

Table 1: Lemmata with hash signs and slashed zeros by parts of speech

A hash sign can be found to mark the absence of equivalence in one or exceptionally several senses of the lemma, phraseological unit or — in verbal lemmata — multi-word verb. A slashed zero marks zero equivalence in one or several senses of the lemma only and cannot be found in idioms and phrasal verbs sections. More detailed results will be presented below by parts of speech of the lemmata.

3.1 Verbs

The majority of verbal entries contain a hash sign to indicate lack of dictionary equivalent in one sense of the lemma (9 verbs: *bear, begin, bless, bother, bugger, commit, get, let, may*). In the entry for the verb *come*, three senses (i.e., 8, 9 and 12) are marked by the use of a hash, but senses 8 and 9 additionally include an explanatory phrase or a pattern illustration:

- sense 8: the explanatory phrase *v vprašalnih stavkih za how* 'in questions after how' precedes the hash (e.g., *How do you come to be so late? Kako to, da si tako pozen?*);
- sense 9: *come sth (with sb)* before the hash signifies a pattern (e.g., *Don't come the innocent with me! Ne delaj se nedolžnega!*);
- sense 12: translated examples are included following two patterns, i.e., *come + gerund* (e.g., *come flying prileteti; come sobbing prihlipati*) or *come + prepositional phrase* (e.g., *come into effect stopiti v veljavo, začeti veljati; come into force stopiti v veljavo, začeti veljati*).

In the verbal lemmata *cry* and *go*, a hash is used in the idioms section to mark the absence of a dictionary equivalent for a phraseological unit. The verb *cry* is used in the phraseological unit *cry over spilt (AmE spilled) milk*, whose meaning is illustrated by a translated example (*It's no use crying over spilt milk. Po toči zvoniti je prepozno.*). The verb *go* is used in the phraseological unit *be going to do sth*, defined as 'used to show what somebody intends to do in the future' and 'used to show that something is likely to happen very soon or in the future'. In the ESD, both senses in English are translated into Slovene grammatically by the future tense form (e.g., *She's going to ring us. Poklicala nas bo.*). Apart from that, the phraseological unit is accompanied by the explanatory phrase *za izražanje prihodnosti* 'used to express future'.

The analysis of verbal entries containing a slashed zero shows that our material contains three auxiliary verbs, five modal verbs, one full lexical verb and the informal contraction of *does* i.e., 's. The auxiliary verbs *be, do* and *have* do not have an equivalent in Slovene when they act as operators, i.e., the first verb in a finite verb phrase (cf. Quirk et al. 1985: 120) (e.g., *What does he want? Kaj želi?*).

As opposed to auxiliary verbs, which share an association with the basic grammatical categories of tense, aspect and voice, the modal verbs mainly express modal meanings, such as possibility, obligation and volition (Quirk et al. 1985: 129). In the ESD, five out of nine verbs with at least one sense where zero equivalence can be observed belong to modal verbs, i.e., *can, should, used to, will* and *would*. Only one sense of the modal verb *can* is characterized by zero equivalence, i.e., the use of *can* with verbs of perception and with mental-state verbs: e.g., *I can't hear you. Ne slišim te.* As can be seen from this example, the Slovene translation contains only the full verb (i.e., *hear = slišati*), whereas the modal verb *can* remains untranslated. In the entry for the modal verb *should*,

only one sense cannot be rendered into Slovene. The examples of use included for this particular sense show that the lexicographers gathered examples illustrating the use of different senses of the verb *should* that have no equivalent in Slovene and translated them under this specific sense. In the modal verbs *will* and *would*, only one sense that they both have in common is characterized by zero equivalence, i.e., descriptions of personal habits or characteristic behaviour. As is evident from examples illustrating the use of *will* and *would* in this sense, only the full verbs are translated into Slovene, whereas the modal verbs are left out in the Slovene translations (e.g., *She will spend hours on the telephone.* Cele ure preživi na telefonu.; *Her grandparents would take care of her.* Zanj o skrbeli stari starši.). The modal verb *used to* expresses habitual and state meanings. Slovene, however, does not make use of a modal verb in this function, which means that *used to* is not rendered into Slovene by a corresponding modal verb (e.g., *Did he use to work here?* Ali je delal tukaj?). The translated examples show that Slovene translations contain the imperfective form of the full verb in the past tense, but no modal verb is used in the translation.

Have is the only full lexical verb in our database with one out of fourteen senses characterized by zero equivalence. This is the causative use meaning 'to cause something to be done for you by somebody else' (e.g., *He had his head cut off.* Obglavili so ga.). The translation into Slovene contains the translation of the full lexical verbs in the English example (e.g., *cut off* = *obglaviti*); these verbs are (usually) in the third person plural, which here suggests generic or impersonal use.

Apart from the symbol \emptyset , some senses of the auxiliary verbs as well as of modal verbs are also equipped with short descriptions of the function of the auxiliary verb. For instance, (the verb *be*) s sedanjim deležnikom za tvorbo nedovršnih časov 'with the present participle to form progressive tenses'; (the verb *can*) z glagoli, ki izražajo čute in mišljenje 'with verbs expressing perception and the mind'.

3.2 Nouns

In nominal lemmata, a hash can be found to indicate the absence of dictionary equivalents for one of the senses of the lemma (in eight lemmata) or in the idioms section to denote lack of equivalence for a phraseological unit (in three phraseological units). Lack of dictionary equivalents at sense level is especially interesting in the noun *brainchild*. *Brainchild* is a monosemous noun, which means that a dictionary user is offered no equivalents in the TL and can infer the meaning of the lemma from one translated example only (*The system was his own brainchild.* On je bil duhovni oče tega sistema.).

In the entries for the nouns *accident*, *amount* and *comfort*, the hashes appear in the idioms sections to mark a lack of dictionary equivalents for phraseological units:

- *an accident of birth*: *By accident of birth he is entitled to American citizenship.* Ker se je slučajno rodil v Ameriki, ima pravico do ameriškega državljanstva.
- *too close/near for comfort*: *The bombs fell in the sea, many too close for comfort.* Bombe so padale v morje, mnoge veliko preblizu. The translation into Slovene depends on the English adverb (*close, near*) used in the phraseological unit (*too close/near = preblizu*).
- *no amount of sth will do sth*: *No amount of persuasion could make her change her mind.* Naj so jo še tako prepričevali, mnenja ni spremenila. The pragmatic meaning 'used for saying that something will have no effect' is clearly expressed in the Slovene translation.

Pragmatic meaning can also be observed in the entry for the noun *goodness* defined as 'used to express surprise'. In this sense, the noun lacks dictionary equivalents in Slovene, but the illustrative examples show that the noun is used in more or less fixed expressions, although the lexicographers obviously did not decide to include them in the idioms section. Nevertheless, the translation into Slovene clearly reflects the pragmatic meaning of the English noun (*My goodness!* or *Goodness me!* or *Goodness gracious (me)!* Moj bog!).

Zero equivalence marked with a slashed zero can be found in three nominal lemmata (*piece, round, space*) in the ESD. In the sense 'a single item', *piece* is used in a construction consisting of a partitive (i.e., *piece*) linked by the preposition *of* to another noun (e.g., *research, advice*) and expressing quantity partition. The noun *piece* is not translated into Slovene, since its only function is to express the quantity and thus countability of English uncountable nouns. The nouns *raziskava* 'research' and *nasvet* 'advice' are countable in Slovene, as opposed to English where they are uncountable; therefore, no additional lexical item is needed in Slovene to express plurality. A partitive use can also be observed in one sense of the noun *round*, which remains untranslated in Slovene. In the partitive sense, *round* is used with *applause* and *cheers* to mean 'a short period during which people show their approval of somebody/something by clapping, etc.'. The only way of showing the users that *round* in this sense does not have an equivalent in Slovene is to include translated examples (i.e., *round of applause* aplavz, ploskanje; *round of cheers* vzklikanje, pozdravljanje).

3.3 Adjectives

The dictionary under investigation includes seven adjectival lemmata with a hash sign. The hash is used to mark the absence of dictionary equivalents for one phraseological unit, i.e., *wouldn't be seen/caught dead in/at/with sth* and its variant form *wouldn't be seen/caught dead doing sth*, which is included in the idioms section of the lemma *dead*. In all other cases, the hash is used to indicate lack of equivalence for one specific sense of the lemma and is followed by translated illustrative examples. The adjective *delayed-action* is monosemous, which means that the hash sign indicates a complete lack of dictionary equivalents

and two examples following the same pattern, i.e., *delayed-action* + noun, are provided for the dictionary user to retrieve necessary semantic information. Both examples could be regarded as compounds, or more precisely, as terms, and are also rendered into Slovene as such: *delayed-action bomb* bomba s tempirnim vžigalnikom and *delayed-action mechanism* samosprožilec.

Three adjectival lemmata marked with a hash in the ESD, i.e., *gracious*, *great* and *holy*, have some characteristics in common. Firstly, the sense with no dictionary equivalents in Slovene has a pragmatic definition in English: *gracious* 'used for expressing surprise'; *great* 'used to express shock or surprise'; *holy* 'used to emphasize that you are surprised, afraid, etc.'. Secondly, they all express restrictions and constraints regarding usage, which are reflected in the accompanying labels: *gracious* labelled *becoming old-fashioned*; *great* labelled *spoken old-fashioned*; *holy* labelled *informal*. Thirdly, the examples illustrating the use of the lemma with no equivalents are fixed expressions rather than free combinations:

- *gracious*: *Good(ness) gracious!* or *Gracious me!* Moj bog! (two variant forms of the same expression rendered into Slovene in exactly the same way)
- *great*: *Great heavens!* Za božjo voljo.
- *holy*: *Holy cow/cats/mackerel/shit/smoke!* Za božjo voljo!, Sveta nebesa!

3.4 Prepositions

The senses of the prepositional lemmata marked by a hash sign are mostly characterized by a higher number of examples if compared with the number of examples illustrating the senses of lemmata with zero equivalence belonging to other parts of speech. In all prepositional lemmata, the hash is used to mark the absence of equivalents in one sense only rather than to indicate an absence of equivalents in the idioms section.

A careful analysis of prepositional lemmata shows that the sense marked by a hash sign in the ESD contains examples illustrating different senses of the lemma in English. The absence of dictionary equivalents is a characteristic shared by these English senses. This can best be exemplified with the entry for *by*, in which the following four senses in English are represented as one sense in the ESD:

- 'used to say that something happens in a particular kind of light': e.g., *by day* podnevi; *by moonlight* v mesečini; *by daylight* pri dnevni svetlobi;
- 'used to state the rate at which something happens': e.g., *day by day* iz dneva v dan; *bit by bit* pomalem;
- 'used before particular nouns without *the*, to say that something happens as a result of something': e.g., *by mistake* pomotoma; *by accident* po nesreči;
- 'used to show how something is done': e.g., *by yourself* sam.

In the ESD, two prepositions marked with a slashed zero in at least one sense

are included, i.e., *at* and *of*. According to Quirk et al. (1985: 665), *at* can be used to indicate space, time, goal, target, stimulus, standard and reaction, whereas *of* indicates cause, means, subject matter and material. These various senses are also reflected in translations, but there are contexts in which the two prepositions lack an equivalent in Slovene. This can best be seen if we study the examples in the ESD that illustrate the use of *at* (e.g., *at large* na prostosti; *at least* vsaj; *be at its best* biti najlepši). Similarly, the examples of use found in the entry for the preposition *of* exhibit the same characteristic (e.g., *a friend of mine* moj prijatelj; *some of my friends* nekateri moji prijatelji; *a girl of ten* desetletna deklica; *the role of the teacher* učiteljeva vloga). *Of* is frequently used as a postmodifier in noun phrases in a function similar to that of the genitive, which also corresponds to some of the examples of use. In this function, the noun in the genitive is often used in the Slovene translation, but this is not the only possibility for translating *of* followed by a noun phrase. As can be seen from the above examples, the following two options exist: translation by means of an adjective (e.g., *the role of the teacher* učiteljeva vloga — učiteljev is an adjective); translation by means of a determinative possessive (e.g., *a friend of mine* moj prijatelj — *moj* is a possessive pronoun).

3.5 Adverbs

Two adverbial lemmata in the ESD are marked by the use of a hash sign, i.e., *jolly* and *out*. In the entry for *jolly*, the hash appears in the idioms section to mark the absence of dictionary equivalents for the phraseological unit *jolly well*, which is defined by the pragmatic definition 'used to emphasize a statement when you are annoyed about something'. Its connotative value is marked by the labels *informal*, *old-fashioned*, *British English*. This phraseological unit is illustrated by translated examples (e.g., *I'm going to jolly well tell him what I think of him!* Mu bom že povedal, kaj si mislim o njem.). In the ESD, one out of twenty-four senses of the adverb *out* is characterized by the absence of dictionary equivalents. This sense corresponds to the following senses in English: 'used to show that something/somebody is removed from a place, job, etc.' and 'used to show that somebody is no longer involved in something'.

When studying the senses of the adverbs *along*, *away*, *out*, *there*, *through*, *up* and *very* with the slashed zero, it becomes obvious that the adverb is sometimes part of a multi-word verb. For example, in the ESD, the senses of *along* and *away*, which cannot be translated into Slovene without the accompanying verb, are illustrated by examples such as: *How are things coming along at work?* Kako gre delo?; *The water boiled away.* Voda je povrela. As to the first example, *along* is part of the multi-word verb *come along* meaning 'to improve or develop in the way that you want'. Similarly, *boil away* used in the second example is also a multi-word verb meaning 'if a liquid boils away, or if you boil it away, it disappears and turns to gas after reaching a very high temperature'. The spatial senses of *out* ('a long or a particular distance away from a place or from land'),

up ('to or at an important place, especially a large city') and *through* ('travelling through a place without stopping or without people having to get off one train and onto another') represented in the examples in the ESD remain untranslated in Slovene (e.g., *live out in the country* živeti na deželi; *be up at Oxford* biti na oxfordski univerzi; *go straight through to it* naravnost v). Existential *there* is also marked with the symbol \emptyset indicating zero equivalence and its function can only be made clear to the users of a bilingual English–Slovene dictionary by including examples translated into Slovene (e.g., *There's a cinema around the corner*. Kino je za vogalom.). One sense of the adverb *very*, i.e., 'used to emphasize a superlative adjective or before *own*' is characterized by zero equivalence. *Very* as an intensifier here premodifies the nonperiphrastic superlative or the emphatic determinative *own*. It is clear from the examples illustrating this meaning of *very* in the ESD (*the very best quality* najboljša kakovost; *the very last time* zadnjič; *her very own car* čisto njen avtomobil) that the Slovene translation may also contain the superlative, although the superlative in Slovene is not premodified by any word, e.g., the superlative of the adjective *dober* 'good', i.e., *najboljši* 'the best'. The adverb *zadnjič* is an equivalent of *the last time*, whereas *čisto njen avtomobil* is translated literally as 'an entirely her car', which means that an intensifier is also used in Slovene as a premodifier of the possessive pronoun *njen* 'her'.

3.6 Pronouns

The only pronominal lemma with a hash sign is *one*. The hash is used to indicate the absence of equivalents for one out of six senses, i.e., 'used to avoid repeating a noun, when you are referring to somebody/something that has already been mentioned, or that the person you are speaking to knows about'.

Four pronominal lemmata included in the ESD are marked with a slashed zero. If we analyse these lemmata more carefully, we can see that the indefinite pronoun *any* is not rendered into Slovene in contexts such as *We got home without any difficulty*. Domov smo prišli brez težav.; *I haven't any idea*. Pojma nimam and the personal pronoun *it* in contexts such as *Stop it! Nehaj!* (the Slovene translation contains the imperative of the verb *nehati* 'stop'); *What about/of it? Zakaj sprašuješ?* (the Slovene translation can be glossed as 'Why are you asking?'). In these contexts, *it* is used anaphorically, its meaning being 'used to refer to a fact or situation that is already known or happening'. In the entries for the demonstrative pronouns *this* and *these* one sense characterized by zero equivalence can be found in the ESD. Contexts illustrating the absence of a dictionary equivalent are as follows: *This is Mary speaking*. Mary tukaj.; *these trousers of mine* moje hlače. In the latter example, the demonstrative pronoun functions as a premodifier in a nominal phrase which is postmodified by *of* + independent possessive. In the Slovene translation, the demonstrative pronoun is dropped and the whole structure is translated by means of the possessive pronoun followed by the noun (*moje hlače* 'my trousers'). It should be pointed out

that Slovene does not distinguish between determinative and independent possessives and therefore, the translation of such examples corresponds to the translation of the combination determinative possessive + noun, which is also the reason for zero equivalence in this sense of *this/these* in the ESD.

3.7 Miscellaneous

In the ESD, three lemmata with a hash sign lack the part-of-speech label (see Table 1 under Others), i.e., the lemmata *d'you*, *gonna* and *let's*. All three lemmata are monosemous and represent short or informal forms. The example with *d'you* is translated as a question in the present and past tense forms (*What'd you say? Kaj praviš?, Kaj si rekel?*), *gonna* is translated using the future tense form in Slovene (*This isn't gonna be difficult. To ne bo težko.*) and *let's* is translated using the imperative form of the verb in the first person plural (*Let's go to the cinema. Pojdimo v kino.*). It could be said that all three lemmata are translated grammatically rather than lexically. Apart from that, the use of *gonna* is additionally described by a short explanatory phrase, i.e., *pogovorna oblika za going to, ki izraža prihodnost 'an informal form for going to used in reference to the future'*.

Three dictionary entries with a slashed zero treating either the definite (*the, ye*) or indefinite article (*a*) can be found in the ESD. The English articles can sometimes be rendered into Slovene, but in certain contexts, they cannot possibly be translated. This is illustrated by some typical contexts in the dictionary under investigation: *Do you have a car? Ali imate avto?; the house at the end of the street hiša na koncu ulice; Ye Olde Bull Stari bik*. The Slovene nouns *avto* 'car', *hiša* 'house' and *bik* 'bull' are not premodified by any determiner, and the kind of reference a particular noun phrase has in English is lost in the Slovene translation. It is, therefore, impossible to say on the basis of the Slovene translations whether a certain noun phrase has indefinite reference (indicated by the indefinite article) or definite reference (indicated by the definite article).

4. Discussion

The absence of dictionary equivalents is an issue which should be adequately addressed by bilingual lexicographers, who should base their decisions and solutions on an in-depth lexical and grammatical contrastive analysis. This analysis can give them clear insight into the differences of the lexical meaning of the two lexical units as well as the grammatical functions in the respective languages. Special attention should be paid to the various co-texts and contexts in which the lexical item in question is used in the TL and to its rendering into the SL. If different co-texts and contexts suggest that the provision of a dictionary equivalent is not possible, this does not necessarily mean that the lexical item in the SL is untranslatable or untranslated, let alone omitted in the TL. Just

the opposite, an SL lexical item does not disappear in the TL context; however, its equivalence can only be observed at the level of the entire message rather than at word level. In many cases, the lexicographers are faced with the problem of a complete absence of equivalence when dealing with grammatical words. We cannot but agree with Zgusta (1971: 115), who claims that grammatical words are just a part of the total lexicon of the language and lexicographers have to register them and to indicate their meaning, or more precisely, to indicate when and how they are used and with what grammatical function. These words are language-specific, which means that many of their functions in one language are carried out in another language in a completely different way, and no parallels can be drawn between the languages. The semantic content of grammatical words may not even be mentioned in a bilingual dictionary; instead, lexicographers include a description of the function of the lemma in a particular sense, which is also the case in the ESD. Or, as stated by Atkins and Rundell (2008: 472-473), grammatical information is provided instead of lexical equivalence. If no equivalent can be included, the lexicographers need to employ other means to show how the lemma or one of its senses can be translated in specific types of contexts, which means that they should include translated examples of use (for more comprehensive treatment of examples of use in a bilingual dictionary, see Vrbinc and Vrbinc 2016).

In what follows, entries containing senses with zero equivalents are discussed from the pragmatic, lexico-grammatical, lexical and grammatical points of view and possible solutions are proposed.

4.1 Zero equivalence in pragmatic senses

The results show that there is a notable lack of dictionary equivalence in those senses of the lemmata that express pragmatic meaning in the SL. Lexicographers are often faced with lexical items that encode not only a meaning but also an attitude, emotion or a certain degree of politeness and formality conventions, which can reinforce or contradict the speaker's intended meaning (Hartmann and James 1998: 111). This means that a word or word combination can express not only meaning but also the feeling and opinions of the speakers. This is in line with the findings of our study, since we also observed that in a high number of senses reflecting pragmatic use, the reference is to feelings, emotions and opinions. As Atkins and Rundell (2008: 422) state, corpus data show that the speaker's attitude tends to be conventionally lexicalized only in a fairly limited number of frequently occurring words and phrases. Consequently, the lexicographic treatment of lexical items with pragmatic characteristics differs from the treatment of other lexical items. In some cases, dictionary equivalents can be provided which are immediately insertable into the context, but in many cases, contrastive differences between the SL and TL cause untranslatability at word level; therefore, lexicographers are forced to use context or context to show how an SL lexical item is reflected in the TL, which is

also the case in the ESD.

In the entries for *gracious*, *great*, *holy*, *goodness* and *bless*, the hash is used to indicate the absence of equivalence of a pragmatic sense at word level. Taking a closer look at the examples illustrating these senses, we can see that all of them are idiomatic to a certain extent, which is also reflected in their idiomatic translations into Slovene. Given their idiomatic characteristics, they could also be included in the idioms section of the respective entries. If that were the case, zero equivalence would not be an issue, since the sentential form has a perfect equivalent in the TL, i.e., in Slovene.

4.2 Zero equivalence due to lexico-grammatical differences between SL and TL

The contrastive analysis of the examples illustrating the semantics of the senses marked by the hash sign clearly shows that it is not uncommon to find examples combining lexical and grammatical characteristics. This can be attributed to the distinct structure of languages resulting in the fact that they express different characteristics in different ways. Good examples are the English verbs *come* used in the patterns *come* + gerund and *come* + prepositional phrase and *get* used in the patterns *get something done* and *get* + adjective, which are translated into Slovene using the perfective form of the verbs. Similarly, some senses of modal verbs also express grammatical as well as semantic features. A repetitive action in the past (*would* and *used to*), for instance, has no direct equivalent in Slovene; therefore, the only option left to a lexicographer is to resort to examples of use. By translating English examples, the lexicographer gives the appropriate information needed by a dictionary user without having to provide lengthy and complicated theoretical explanations of a specific function.

The use of the article poses problems for the target users of the ESD, i.e., native speakers of Slovene, because articles are not used in Slovene at all and consequently, no parallels can be drawn between English and Slovene as regards articles. In most cases, the English articles are simply dropped in rendering a translation in Slovene. Apart from that, the use of articles in English is often idiomatic and therefore highly unpredictable for non-native speakers. Therefore, the dictionary users should be made aware of the absence of an equivalent in specific contexts by the symbol \emptyset and translated examples.

4.3 Zero equivalence due to grammatical differences between SL and TL

The primary function of the auxiliary verbs *be*, *do* and *have* is to express grammatical relations. They may have different functions in the verb phrase, but their ability to act as operators should be emphasized, since this function is the most common cause of problems for lexicographers when they try to provide dictionary equivalents in a bilingual dictionary. If we take a look at the tense formation in both languages of the ESD, we can see that tenses in Slovene are

formed in a way completely different from how tenses are formed in English. The primary reason for that is that these two languages belong to different types of languages according to morphological linguistic typology: English is an analytic language, while Slovene is a synthetic language. This means that English conveys grammatical relations without using inflections; Slovene, on the other hand, makes abundant use of inflections and exhibits a high morpheme-per-word ratio. In Slovene, the same content may be expressed by using completely different structures or specific lexical items.

From the perspective of a non-native speaker, the phraseological unit *be going to do sth* deserves special attention. In the ESD as well as in monolingual learner's dictionaries, it can be found in the idioms section, but a non-native speaker of English is unlikely to search the idioms section for this word combination. In the process of learning English as a foreign language, learners are taught that this structure is used to form the future tense which expresses intentions, plans, etc. in the future. Therefore, it would be advisable to avoid including this structure in the idioms section of a mono- or bilingual dictionary as a phraseological unit, but rather to treat it as a separate sense with a fixed pattern and a brief theoretical explanation followed by illustrative examples, which should be translated in a bilingual dictionary. The same solution can be provided in other cases where grammatical differences between the SL and the TL exist. For instance, in nominal lemmata, one case of zero equivalence described in the Results section concerns the partitive constructions used to make English uncountable nouns countable. A contrastive analysis shows that in Slovene, the nouns used as equivalents of English uncountable nouns included in our study are countable.

A theoretical explanation providing grammatical information instead of lexical equivalence is sometimes included to help the user to more fully understand the use of the lemma in question (e.g., *be going to do sth, come, gonna, be, have, can*). It must be stressed that the ESD includes explanations that are short, precise, to the point, and above all, characterized by the use of very simple language with almost no technical terms. The metalanguage used in the ESD is Slovene, which also holds true for the short descriptions, since the dictionary is primarily intended for native speakers of Slovene. The native language of the primary target users seems to be the only sensible choice for the metalanguage (cf. Atkins and Rundell 2008: 234), especially if the dictionary is pedagogically oriented, which is the case with the ESD. Short descriptions can be regarded as valuable, since they represent a short comment on the specific use dealt with in a specific sense. In the ESD, they are provided in brackets, are typographically distinct from dictionary equivalents and precede the hash or slashed zero.

4.4 Zero equivalence due to lexical differences between SL and TL

The issue of zero equivalence in adverbial lemmata is often connected with the use of adverbs in multi-word verbs. It should be stressed that in lexicography,

no distinction is made between phrasal verbs, prepositional verbs or phrasal-prepositional verbs. In practice, all multi-word verbs are included and treated under verbal lemmata in a special section (the phrasal verbs section), but in adverbial lemmata we can also come across examples of multi-word verbs. The reason is that lexicographers try to define (in monolingual dictionaries) or translate (in bilingual dictionaries) the meaning of a particle that is a constituent element of a multi-word verb, since the combination of a full lexical verb and a particle or two particles poses problems for non-native speakers of English, who need to learn the meaning of the whole combination. It should be pointed out that multi-word verbs are a peculiarity of English with no exact parallel in Slovene, which means they deserve careful and appropriate treatment. If multi-word verbs were not included in the ESD under adverbial lemmata, the problem of zero equivalence would be resolved, since the comprehension of the adverb does not contribute to the comprehension of the multi-word verb. It is therefore advisable to include multi-word verbs under verbal lemmata only and to avoid listing them under adverbial lemmata.

As far as pronominal lemmata with zero equivalence are concerned, the pronoun *any* used as a non-assertive pronoun in negative sentences lacks an appropriate equivalent in Slovene. Thus, the treatment as provided by the ESD can be assessed as appropriate, since zero equivalence is presented in the form of translated examples of use. The same observation can be made in connection with the personal pronoun *it*, which according to Quirk et al. (1985: 348-349) is the most neutral and semantically unmarked of the personal pronouns, and this is also reflected in Slovene translations of the examples in the ESD.

If we take a close look at the prepositional lemmata, we can see that in the entry for *at*, the examples included in the ESD are idiomatic expressions which should be learnt as whole chunks of lexical semantic-syntagmatic entities (Alexander 1989: 16), since they are semantically mostly not reducible to their immediate constituents. Consequently, *at* is not translated into Slovene, and its absence in the Slovene translation is indicated to make the user aware of zero equivalence. Here, a parallel can be drawn between these idiomatic expressions and multi-word verbs in that it is much more appropriate to include idiomatic expressions under the first full lexical word in the expression rather than under the preposition.

4.5 Other cases of zero equivalence

Our research clearly shows that, in many cases, several senses of a lemma in the SL are characterized by a lack of dictionary equivalence in the TL. As stated by Zgusta (1971: 315), a monolingual dictionary of the SL can be regarded only as a first orientation when gathering information about the meaning of an SL lexical item, but the comparison of the SL and TL lexical items should be made by the contextual method to identify the potential differences between languages. As a rule, the lexicographers compiling the ESD decided to include all

senses of the lemma in English with zero equivalence in Slovene under one sense, the only exception to this rule being the lemma *come*, where three senses have a hash sign. The reason for this exception is that two senses are characterized by a specific structure (i.e., sense 8, 'in questions after how'; sense 9, *come sth (with sb)*). The reduction in the number of senses with a hash seems a sensible decision and can be regarded as a way of simplifying the dictionary entry structure, which is a welcome feature in complex polysemous entries where dictionary users struggle to find the right sense.

5. Conclusion

In bilingual dictionaries, contrastive differences between the SL and TL, as well as features typical of either the SL or the TL, result in different types of equivalence, which are treated by lexicographers in a wide variety of ways. The focus of our study is on how bilingual lexicographers tackle the problem of zero equivalence, since they should not refrain from attempting to fill lexical gaps existing between the SL and the TL. One of the most important conclusions is that, if equivalence cannot be achieved by providing a dictionary equivalent, the problem can be resolved by including examples of use which should be selected very carefully by the lexicographers to enable the users to become familiar with different contexts in which the lemma or one of its senses is used. The most important function of examples in a bilingual dictionary is to show contrastive differences between two languages and to illustrate the differences between the dictionary equivalent(s) offered in the dictionary and the translation of the lemma in context. It should, however, be stressed that it is absolutely necessary to provide a translation of illustrative examples into the TL (Vrbinc and Vrbinc 2016: 308). If the examples remain untranslated, the user is not made aware of the problem of non-equivalence; it could therefore be claimed that such a bilingual dictionary does not fulfil one of its main tasks, i.e., to clearly show how two different languages function in everyday use. We cannot but agree with Gouws (2002: 208), who points out that "[c]ontext and context can play an important role to ensure an optimal retrieval of information from bilingual dictionaries", since lexical items always occur with their collocates; consequently, one can expect them to appear in a similar co-text or context, thus influencing the way they are used in speech and writing. To compensate for the absence of dictionary equivalents, the lexicographers in the ESD resorted to translated examples of use, which can be considered appropriate if the dictionary is to function as an effective communication tool.

To conclude, the main objective of the compilers of any bilingual dictionary is to offer the users dictionary equivalents that can fit as many contexts as possible and can be immediately insertable when translating from the SL into the TL. In cases where no equivalents exist, the users should be offered workable solutions which enable them to deal successfully with the absence of dictionary equivalents. The approaches and solutions used by the lexicographers

compiling the ESD can help dictionary users to make more effective use of the dictionary even in cases where no parallels can be drawn between their mother tongue and a foreign language, thus achieving a more successful and thorough retrieval of information with every look-up.

References

- Adamska-Sałaciak, Arleta.** 2010. Examining Equivalence. *International Journal of Lexicography* 23(4): 387-409.
- Adamska-Sałaciak, Arleta.** 2011. Between *designer drugs* and *afterburners*: A Lexicographic-Semantic Study of Equivalence. *Lexikos* 21: 1-22.
- Alexander, Richard John.** 1989. Fixed Expressions, Idioms and Collocations Revisited. Meara, Paul (Ed.). 1989. *Beyond Words*: 15-24. London: Centre for Information on Language Teaching and Research/British Association for Applied Linguistics.
- Atkins, Beryl T. Sue.** 1992/1993. Theoretical Lexicography and its Relation to Dictionary-Making. *Dictionaries* 14: 4-43.
- Atkins, Beryl T. Sue and Michael Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Gouws, Rufus H.** 2002. Equivalent Relations, Context and Cotext in Bilingual Dictionaries. *Hermes, Journal of Linguistics* 28: 195-209.
- Hartmann, Reinhard Rudolf Karl and Gregory James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hu, Shuqin.** 2010. Context of Situation in Translation. *Journal of Language Teaching and Research* 1(3): 324-326.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik.** 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.
- Svensén, Bo.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Šipka, Danko.** 2015. *Lexical Conflict: Theory and Practice*. Cambridge: Cambridge University Press.
- Vrbinc, Alenka and Marjeta Vrbinc.** 2016. Illustrative Examples in a Bilingual Decoding Dictionary: An (Un)necessary Component? *Lexikos* 26: 296-310.
- Wiegand, Herbert Ernst.** 2002. Equivalence in Bilingual Lexicography: Criticism and Suggestions. *Lexikos* 12: 239-255.
- Wiegand, Herbert Ernst.** 2005. Äquivalenz, Äquivalentdifferenzierung und Äquivalentpräsentation in zweisprachigen Wörterbüchern: Eine neue einheitliche Konzeption. Gottlieb, Henrik, Jens Erik Mogensen and Arne Zettersten (Eds.). 2005. *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*: 17-57. Tübingen: Max Niemeyer.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague/Paris: Mouton.

Colors in French, American and British Dictionaries

Krista Williams, *Department of Foreign Languages and Cultures,
University of Evansville, IN, USA (kw271@evansville.edu)*

Abstract: Colors have senses specific to particular fields such as physics and printing, in addition to senses used in everyday life. This article examines the specialized information found in color definitions in French, American and British dictionaries. We explore whether specialized and non-specialized definitions are lumped or split, how much scientific information is included, if the information included varies diachronically and/or geographically and if dictionaries are consistent in their labeling of colors as members of a group (primary, spectral, etc.). We found that specialized and non-specialized senses of colors are typically lumped, rather than split. This is contrary to the treatment of other words with both specialized and non-specialized senses in the same dictionaries, suggesting that the line between these senses is not clear for colors. We also found that more specialized information has been included over time in French, American and British dictionaries, but that American dictionaries still include the most. Additionally, American dictionaries are more consistent than dictionaries of the other two nationalities in their labeling of colors as members of a group, both in the labels themselves and in labeling all members of a group. Based on our findings, we make suggestions for defining colors in general use dictionaries.

Keywords: COLOR DEFINING, LUMPING AND SPLITTING, SCIENTIFIC INFORMATION IN DICTIONARIES, FRENCH, FRENCH DICTIONARIES, ENGLISH, AMERICAN DICTIONARIES, BRITISH DICTIONARIES, USER EXPECTATIONS

Opsomming: Kleure in Franse, Amerikaanse en Britse woordeboeke. Buiten die betekenis wat kleure in die alledaagse lewe het, beskik kleure ook oor betekenis wat spesifiek is aan bepaalde velde soos die fisika en die drukkerswese. In hierdie artikel word die gespesialiseerde inligting wat in kleurdefinisies in Franse, Amerikaanse en Britse woordeboeke gevind word, ondersoek. Daar word gekyk of gespesialiseerde en nie-gespesialiseerde definisies saamgevoeg of geskei word, hoeveel wetenskaplike inligting ingesluit word, of die inligting diachronies en/of geografies varieer en of woordeboeke konsekwent is in hul etikettering van kleure as deel van 'n groep (primêr, spektraal, ens.). Ons het bevind dat gespesialiseerde en nie-gespesialiseerde betekenis van kleure eerder saamgevoeg as geskei word. Dit is teenstrydig met die hantering van ander woorde met beide gespesialiseerde en nie-gespesialiseerde betekenis in dieselfde woordeboeke, wat daarop dui dat die skeidslyn tussen betekenis nie duidelik is vir kleure nie. Ons het ook bevind dat meer gespesialiseerde inligting met verloop van tyd in Franse, Amerikaanse en Britse woordeboeke ingesluit is, maar dat Amerikaanse woordeboeke steeds die meeste insluit. Boonop is Amerikaanse woordeboeke ook meer konsekwent as die woordeboeke van die ander twee nasionaliteite betreffende hul etikettering van kleure as deel van 'n groep, sowel in die etikette self as in die etikettering van alle groepslede. Op grond van hierdie bevindings maak ons voorstelle vir die definiering van kleure in algemene gebruikerswoordeboeke.

Sleutelwoorde: KLEURDEFINIËRING, SAMEVOEGING EN SKEIDING, WETENSKAP-LIKE INLIGTING IN WOORDEBOEKE, FRANS, FRANSE WOORDEBOEKE, ENGELS, AMERI-KAANSE WOORDEBOEKE, BRITSE WOORDEBOEKE, GEBRUIKERSVERWAGTING

1. Introduction

Colors are a semantic domain with both specialized and non-specialized senses. For instance, when a child asks for a red train, he/she is not using the word *red* to mean "the hue of the long-wave end of the visible spectrum, evoked in the human observer by radiant energy with wavelengths of approximately 630 to 750 nanometers" nor is he/she likely to be interested in the fact that red is an additive primary color. Both of these pieces of information, however, can be found in some lexicographic definitions of *red*. This illustrates the difference between defining a color as a *word* and defining it as a *term*. Defining a word involves describing its usage in attested contexts, such as the child asking for train. Defining a term, however, is typically based on the advice of some authority on the subject (science, math, art, etc.).

While ignoring specialized senses would both disappoint the user and leave out significant semantic portraits (Béjoint 1988), the language used to describe the scientific meaning must be accessible to the average user. As Atkins and Rundell (2008: 450) state, "A definition that provides the necessary content in technically precise language is of no value if it is unintelligible to the users it is aimed at". Color defining is obviously only one semantic domain for which this challenge presents itself. Colors, are, however, considerably more commonly used by the average person than many other scientific terms (phylum names, atomic particles, etc.). Lexicographers defining colors must therefore decide how to treat both specialized and non-specialized senses. This article examines the division of said senses.

Additionally, as dictionaries of different nationalities include varying amounts of specialized information, is there a difference in their treatment of colors? Has the amount of such information changed over time? Before presenting our research to answer these questions, it will first be useful to discuss the inclusion of specialized information in French, American and British dictionaries, the specialized terminology used to label colors as members of a group and how lexicographers treat words with multiple senses.

2. Background

2.1 Sense division

When a word has multiple related senses, lexicographers must decide how many senses a word has that merit their own definitions. These senses "collectively explain how that word contributes to the meaning of all the individual

events" (Atkins and Rundell 2008: 272). These senses may be divided ("split") or grouped together ("lumped"):

The "lumpers" like to lump meanings together and leave the user to extract the nuance of meaning that corresponds to a particular content, whereas the "splitters" prefer to enumerate differences of meaning in more detail; the distinction corresponds to that between summarizing and analysing. (Allen 1999: 61)

Splitting and lumping is an issue for lexicographers defining words with both technical and non-technical senses:

A related problem occurs when words — or senses of words — have a specialist meaning for one group of people as well as a general meaning. To an astrophysicist *space* is a continuum, and the nearest synonym I can think of is *everything*; to most people, *space* is what lies outside the earth's atmosphere. To the former, this planet is part of space: to the latter, it is not. Two senses, or two aspects of the same sense? (Moon 1987: 175-176)

Some scholars have suggested that lexicographers often divide senses too finely (cf. Béjoint 1990 for over-splitting in American dictionaries in particular). Moon (1987: 174-175), for example, states that it is not necessary to split senses so thinly because the context in which a word appears provides the reader with much of the semantic content needed to interpret the word. The meanings of scientific terms, however, are largely independent of the context in which they appear (Béjoint 1988: 358).

2.2 Specialized information in dictionaries

The definitions of scientific and technical terms, both their source and their content, differ from the definitions of most other words, but they make up a large percentage of the wordlist of a dictionary. According to Landau (1974: 241), a conservative estimate would be that forty percent of terms in an unabridged dictionary are scientific or technical. In a smaller dictionary, the percentage may fall to twenty-five to thirty-five percent. In either case, this is a sizeable portion of the dictionary. Although no delineation of the category "scientific and technical terms" has been widely agreed upon, Béjoint (1988: 23) offers his "intuitive" definition, which we will adopt for the purposes of this article: "words that are not used by non-specialists of the science or the technique". Below we refer to scientific and technical terms collectively as "specialized".

Scientific terms are unique in that while most words are defined in modern dictionaries based on their usage in a corpus or citation file, scientific terms are defined based on expert advice. To use Landau's terminology, scientific terms are "imposed," not "extracted" (1974: 242). Landau notes that the concern of these experts is "maintaining the internal coherence of their discipline rather

than faithfully recording how terms are used". This therefore causes a dilemma for lexicographers, whose goal is to represent real usage.

The way in which lexicographers define specialized terms, or whether they include them at all, depends on the amount of encyclopedic information included in their dictionaries. According to Landau (2001: 167), "modern dictionary users expect scientifically precise, somewhat encyclopedic definitions". Many words, including some colors, are used both in scientific terminology and in common usage. Lexicographers encounter a conflict between being true to science or true to common usage. To illustrate this phenomenon, Rey and Delesalle (1979: 24) give the following two sentences:

- (1) le soleil est une étoile
'the sun is a star'¹
- (2) le soleil empêche de voir les étoiles
'the sun prevents the stars from being seen'

These two sentences cannot both be true if the same definition of *soleil* is used. Rey and Delesalle therefore propose that a dictionary should define two senses of *soleil*, one corresponding to "an established cultural description," and the other corresponding to "a scientific notion" (1979: 24). The former represents an extracted definition and the latter an imposed definition. This word must therefore have (at least) two separate definitions to satisfy the modern dictionary user.

2.3 User expectations

The amount of specialized information in a dictionary depends heavily on its intended user. This target audience has varied diachronically, the general trend being from an exclusively upper class, well-educated audience to one with various levels of education and socioeconomic status. Consider the intended audience of three French dictionaries, as described in Béjoint (1994: 109):

Émile Littré's Dictionnaire de la langue française (1872): 'erudites, philologists, physicians, and intellectuals in general'.

Grand Larousse de la langue française (1960): "users 'whose education is above average, students, teachers, well-read individuals.'"

Trésor de la langue française (1971): 'the educated person who is part of what was until recently called the *élite*, but is now preferably referred to as the upper middle or middle middle class, that is to say the most active elements of the main sectors of modern life, and all the people who work with them, not excluding — in fact even giving some special consideration to — all those who write for work or pleasure, those who hold the pens of our culture, or the teachers of all levels who transform all their writings into didactic language so that they can be transmitted to an ever broader section of the masses'.

(Béjoint's translations)

While the target audience of French dictionaries has become more inclusive, the intended user still has an above average social position and level of education.

The target users for British and American dictionaries likewise tend to be part of "the middle classes and to be fairly well educated" (Béjoint 1994: 110). However, despite the similar socioeconomic status and education level of the target user across France, Great Britain and the United States, the amount of encyclopedic information in dictionaries does vary from country to country, often based on user expectations, societal values and the intended use of the dictionary.² Béjoint (1994: 112) summarizes this discrepancy in American, British and French dictionaries. Whereas the British public prefers short definitions with no pictures, French people favor tools related to classical literature and history, such as the Latin translations in the pink pages of the *Petit Larousse*. Americans, on the other hand, want both linguistic and non-linguistic information: encyclopedic facts, usage notes and information such as word-division and etymologies.³ However, these desires of all three nationalities are being superseded by an interest in science and technology:

Everywhere, the typical user has changed because the culture has changed. The interest in classical history and literature ... is being replaced by an interest in science and technology. Dictionaries reflect the changes in society. (Béjoint 1994: 112)

One notable example of these changes is the amount of technical terminology in subsequent editions of the *Dictionnaire de l'académie française*. Zgusta (2006: 146) notes that the 1694 edition contained virtually no technical terms, the 1762 edition some terms, the 1877 edition over 2,000 terms and the 1932 edition "was delayed because of the sharply increasing number of technical terms ..."

More recently, Béjoint (2010: 128) notes the "relative poverty in the representation of scientific and technical terms" in British dictionaries. Their focus is on words rather than on the things they represent. However, in a recent study, Urbinc and Urbinc 2013 found a significant rise in the number of words with subject-field labels in three versions of the *Oxford Advanced Learner's Dictionary*, OALD3 from 1974, OALD4 from 1989, and OALD8 from 2010.

These societal changes are also apparent when one examines the color definitions of dictionaries written in different times and places. The results below include evidence of these societal changes.

2.4 Terminology used to label groups of colors

When colors are defined, they are often labeled as belonging to a group of colors, such as spectral colors, primary colors, etc. In this section, we discuss some of these labels commonly found in French and English dictionaries.

To English speakers, both American and British, the solar spectrum, or rainbow, is conventionally divided into seven colors. From longest to shortest

in wavelength, these spectral colors are: red, orange, yellow, green, blue, indigo and violet. Together, light of all the colors of the spectrum composes white light.⁴ French speakers divide the spectrum into this same seven-color array: *rouge, orange/orangé,*⁵ *jaune, vert, bleu, indigo* and *violet*. These seven colors are referred to as *spectral colors* in this article.

A light source of almost any color can be matched by combining red, green and blue light. These three colors are thus referred to as *additive primaries*,⁶ meaning they are added together to produce other colors. Each of the additive primary colors, when added to its complementary color, produces white light.

Mixing colors of light produces different results than mixing colors of substances such as paints, dyes or inks:

The eye gathers and interprets a mix of wavelengths. Add a wavelength and you give the eye an extra morsel of information to work with ... On a painter's palette, something different happens. The more pigments are added, the *less* colour information the mix contains. Pigments swallow most wavelengths of light; it is the wavelengths they do not absorb which bounce into our eyes. (Ings 2008: 226)

Physical objects absorb light rather than producing it. This absorption is the reason the primary colors of pigment are called *subtractive primaries*. They are combined in order to subtract colors from being reflected, and the colors that are not absorbed are reflected. As with additive primaries, the choice of subtractive primary colors depends upon the medium, but the three most typically chosen, those used for printing, for instance, are magenta (midway between red and violet), cyan (blue-green) and yellow. These pigments are the complements of the three additive primaries of light. Yellow is complementary to blue, magenta to green and cyan to red. Each pigment absorbs one of the primary lights and reflects the other two (Sobel 1987: 65). In other words, pigments are mixed in order to produce reflected light of the additive primary colors red, green and blue (Sobel 1987: 65). This is why unlike the mixture of red and green light, which produces yellow light, mixing red and green pigments results in a dark color.

French speakers also refer to additive and subtractive primary colors:

Additive color mixing applies to light whereas subtractive color mixing applies to matter (Bagard 2009: 40).

Bagard uses the terms "*primaires (ou fondamentales)*" for primary colors. As we will demonstrate below, there are various groups of colors that are labeled *fondamentales*. This label may in fact refer to the spectral colors, the additive primary colors (*rouge, vert, bleu*) or the subtractive primary colors (*cyan, magenta, jaune*).

The last group of primary colors we will discuss is called *Hering psychological primaries*. Contrary to the three-color systems described above, Ewald

Hering, a 19th century German physiologist, proposed that at a psychological level there are in fact four primary colors (excluding white and black). He claimed that the brain considers some colors to be pure. As Sobel (1987: 66) explains: "It is true that all hues are generated by appropriate mixtures of red, green, and blue, but there is a lack of symmetry in our psychological response to these mixtures". For instance, when we mix blue and green light we get a bluish green, but when we mix red and green light, we do not get a reddish green. We instead get yellow: "We feel yellow as an entirely different sensation, psychologically independent of red, green, and blue (unlike orange, say, which clearly retains a feeling of yellow and red)" (Sobel 1987: 66). Bluish-green and yellow, although both are results of mixtures of light, do not have the same status in our minds. The four primary hues thus include red, green, and blue, but also yellow.

All of these groups of colors are labeled in some dictionaries, but as we will demonstrate below, no two dictionaries treat this labeling in the same way.

3. Methodology

We consulted the definitions of the seven spectral colors — red/*rouge*, orange/*orange/orangé*, yellow/*jaune*, green/*vert*, blue/*bleu*, indigo/*indigo*, and violet/*violet* — in order to determine the treatment of specialized and non-specialized senses in color definitions. Because of their status as subtractive primary colors, the definitions of *cyan* and *magenta* (spelled the same in French and English) were also included. We examined the splitting or lumping of specialized and non-specialized senses, the difference in the inclusion of specialized information (across time and nationalities), and the labeling of colors as members of a group. The dictionaries we consulted to answer these questions are found in Figures 1, 2 and 3 below:

Dictionary	Edition	Year(s)	Abbreviation
<i>Le Dictionnaire de l'Académie française</i>	8th	1932–5	DAF8
<i>Le Dictionnaire de l'Académie française</i>	9th	1992–in progress	DAF9
<i>Le Grand Robert de la langue française</i>	online	Continually updated	GR
<i>Le Trésor de la langue française informatisé</i>	online	2002	TLFi

Figure 1: French dictionaries consulted

Dictionary	Edition	Year(s)	Abbreviation
<i>Webster's Third New International Dictionary</i>		1961	W3
<i>American Heritage Dictionary</i>	2nd ⁷	1982	AHD2
<i>American Heritage Dictionary</i>	5th	2011	AHD5
<i>Random House Dictionary of the English Language</i>	2nd	1987	RHD

Figure 2: American dictionaries consulted

Dictionary	Edition	Year(s)	Abbreviation
<i>Concise Oxford English Dictionary</i>	1st	1911	COD1
<i>Concise Oxford English Dictionary</i>	12th	2011	COD12
<i>Collins English Dictionary</i>	online ⁸	2017	CED
<i>Oxford English Dictionary</i>	2nd ⁹	1989	OED2

Figure 3: British dictionaries consulted

4. Results

4.1 Splitting or lumping?

Rey and Delesalle's (1979) suggestion discussed above, i.e., that scientific terms should have two separate definitions, could be applied to color terms, which have both an everyday usage and a precise scientific usage. When dividing the specialized and non-specialized senses of colors, one might expect a treatment similar to that of the noun *tree* in the following entry from the *Collins English Dictionary* (CED):

1. any large woody perennial plant with a distinct trunk giving rise to branches or leaves at some distance from the ground *related adjective* arboreal
2. any plant that resembles this but has a trunk not made of wood, such as a palm tree
3. a wooden post, bar, etc.

4. See family tree, shoetree, saddletree
5. (chemistry) a treelike crystal growth; dendrite
6.
 - a. a branching diagrammatic representation of something, such as the grammatical structure of a sentence
 - b. (*as modifier*) ⇒ a tree diagram
7. an archaic word for gallows
8. (archaic) the cross on which Christ was crucified
9. See at the top of the tree
10. See up a tree

This split in CED (senses 1 and 2) represents a division between what a specialist would consider a tree and what a layperson would consider a tree, or what Moon (1987: 175-176) calls a "specialist" versus a "general" meaning. For a layperson, a tree does not have to be made of wood. For a specialist, it does.

Likewise, consider the definitions of some of the senses of the word *star* from the *Oxford English Dictionary, 2nd Edition* (OED2):

1. a. Any one of the many celestial bodies appearing as luminous points in the night sky. Now usually restricted (in scientific and to some extent in popular language), to the *fixed stars* as distinguished from planets
2. In extended sense, any one of the heavenly bodies, including the sun and moon; sometimes in pl. as a vague designation for the abode of departed spirits; so occas. this star, the earth regarded along with other 'stars' as a place of habitation.

This entry shows a distinct separation between the specialized usage (sense 1) and common usage (sense 2) of the word *star*.

While a similar separation of senses to those found in CED or OED2 is possible, even plausible for colors, it is not what is found in practice. For instance, consider the following definitions from OED2, *Le Trésor de la langue française informatisé* (TLFi) and the *American Heritage Dictionary, 5th edition* (AHD5):¹⁰

red a. Designating **the colour of blood, a ruby, a ripe tomato**, etc., and appearing in various shades at the **longer-wavelength end of the visible spectrum**, next to orange and opposite to violet; of or having this colour. In early use also designating shades of purple, pink, and orange, which are now distinguished by these distinct colour terms. (OED2)

rouge I. —*Adj.* D'une couleur qui parmi les couleurs fondamentales se situe à **l'extrémité du spectre**, et rappelle notamment **la couleur du coquelicot, du rubis, du sang**. (TLFi)

[Of the primary color **at the end of the spectrum**, resembling in particular the color of the **poppy, ruby, and blood**.]

red *n*.1. a. The hue of the **long-wave end of the visible spectrum, evoked in the human observer by radiant energy with wavelengths of approximately 630 to 750 nanometers**; any of a group of colors that may vary in lightness and saturation and whose hue resembles that of **blood**; one of the additive or light primaries; one of the psychological primary hues. (AHD5)

In these three definitions, both spectral information and references to everyday objects are included within the same definition, separated only by commas or semi-colons. Specialized and non-specialized senses are lumped, not split by separate numbers. Even in the AHD5 definition, which is almost exclusively scientific, only a semi-colon divides specialized and common usages of red. A clear distinction is not made between color words and color terms. This mixing is indicative of both the nature of color words/terms and of modern society. Society expects a scientific definition. In the minds of lexicographers, at least, there is not a large separation between a color's specialized and non-specialized senses.

4.2 Inclusion of scientific information in color definitions

Today's society considers specialized information to be very important, even to the non-specialist. Dictionaries of all three nationalities show increased scientific information over time.

Consider, for example, the definitions of *bleu* in *Le Dictionnaire de l'Académie française, 8th edition* (DAF8) and *Le Dictionnaire de l'Académie française, 9th edition* (DAF9) and the definitions of *blue* in the *American Heritage Dictionary, 2nd edition* (AHD2) and AHD5:

BLEU, UE. adj. Qui est de la couleur du ciel quand il est pur. *Satin bleu. Robe bleue. Avoir les yeux bleus.* (DAF8 — 1932–1935)
[That which is of the color of the clear sky. Blue satin. Blue dress. To have blue eyes.]

bleu ★ I. Adj.

★ 1. **Qui, dans le spectre des couleurs, se situe entre le vert et l'indigo** ; qui est de la couleur du ciel quand il est pur, de l'azur, du saphir. (DAF9 — 1992–in progress)

[Situating in the color spectrum between green and indigo; that which is of the color of the clear sky, of azure, of sapphires.]

blue *n*.

1. Any of a group of colors that may vary in lightness and saturation, whose hue is that of a clear sky; **the hue of that portion of the spectrum lying between green and violet; one of the additive or light primaries;**

one of the psychological primary hues, evoked in the normal observer by radiant energy of wavelength approximately 475 nanometers.
(AHD2 — 1982)

blue *n.*

1. The hue of that portion of the visible spectrum lying between green and indigo, evoked in the human observer by radiant energy with wavelengths of approximately 420 to 490 nanometers; any of a group of colors that may vary in lightness and saturation, whose hue is that of a clear daytime sky; one of the additive or light primaries; one of the psychological primary hues.
(AHD5 — 2011)

These successive editions of American and French dictionaries show that specialized information has increased in importance in these dictionaries. AHD5 differs from AHD4 in that it adds wavelengths as well as the fact that the color is evoked by radiant energy. DAF9 mentions spectral position, whereas DAF8 does not. Not only that, but DAF9 puts this spectral information *first*. However, of the four spectral colors currently included in DAF9, only three give their spectral positions: *bleu*, *jaune* and *orange*, but not *indigo*.¹¹

Similarly, the British dictionaries we examined give spectral positions for some, but not all spectral colors. In OED2, the definitions of blue, cyan-blue,¹² green, indigo, red, orange and yellow, but not violet, include their spectral position. CED gives the spectral position of all spectral colors except indigo. There is also a diachronic difference in the versions of the *Concise Oxford English Dictionary* (COD) we examined. In COD1, green, red, violet and yellow are defined in relation to their spectral position. In COD12, the definitions of blue¹³ and indigo also include spectral information. Dictionaries of all three nationalities, thus, have included more specialized information through time.

As discussed above, American dictionaries include more specialized information than French or British ones. Some American dictionaries include information on the physics of how humans perceive color, information that is lacking in the British and French dictionaries in our study, with the exception of CED. It may be surprising to see a British dictionary that includes wavelengths. However, Béjoint (1994: 74-75) notes that CED is the first example of a British college dictionary, a type imported from the US. He describes a college dictionary as: "an encyclopedic dictionary in one volume ... with particular emphasis on the present language and on non-classical culture — particularly scientific and technological". This difference in the amount of scientific information in French and American dictionaries was also confirmed by the findings that no French dictionary in this study included wavelengths for colors other than *vert*. Only *Le Grand Robert de la langue française* (GR) and TLFi included this information, but we have found no reason this color in particular was treated differently. Since both GR and TLFi agree, however, on the wavelength for *vert* (0,52 μ), and, as we will see below, dictionaries generally do not

agree on wavelength figures, it seems likely that they got this figure from the same source (or each other). Annie Mollard-Desfour, the color specialist for the TLFi, explains why she chose not to include wavelengths in her definitions:

As this is a language dictionary, and not a specialized work, [or] an encyclopedia, I purposefully chose not to define colors by their wavelengths. This holds no interest for a language dictionary. We are interested in words and culture, which is often contrary to definitions used by scientists, physicists, etc. (Personal communication, 29 September 2012)¹⁴

This confirms Béjoint 1994's claim discussed above that French lexicographers include less encyclopedic content than their American counterparts.

Curiously, the English dictionaries that do include wavelengths do not agree on wavelength measurements. As these dictionaries give no source of their figures, it is difficult to know why this discrepancy occurs. Consider the measurements listed in AHD5, the *Random House Dictionary of the English Language* (RHD) and CED:

Color	Wavelengths in nanometers		
	AHD5	RHD	CED
Red	630–750	610–780	620–740
Orange	590–630	590–610	585–620
Yellow	570–590	570–590	575–585
Green	490–570	500–570	500–575
Blue	420–490	450–500	445–490
Indigo	420–450	(Wavelength not listed)	(Wavelength not listed)
Violet	380–420	400–450	390–445

Figure 4: Wavelengths in AHD5, RHD and CED

Neither the AHD nor the CED editorial staffs were able to determine the source of these figures. Although these numbers are not far from one another, there is a notable discrepancy at the short-wave end of the spectrum. There is no wavelength given for indigo in CED or RHD, and between the wavelengths of violet and blue, there is no gap in the spectrum which indigo could occupy. If indigo were to be part of the spectrum, it would have to occupy part of the same space as blue and violet. However, the definition of indigo in CED specifies that it is a spectral color. AHD5 does list a wavelength for indigo, but it is part of the range given for blue. These discrepancies likely stem from the fact that many modern scientists only divide the spectrum into six parts, leaving out indigo (Waldman 2002: 193). Thus, there is a mismatch between what the specialist and everyday person consider to be spectral colors.

With this conflict between scientific and common usage, as well as the lumping of specialized and non-specialized senses, the definitions of colors do not clearly distinguish between scientific and non-scientific uses. Although

definitions have become more scientific, this scientific information has been incorporated into the non-scientific definitions. Contrary to a case like *star* above, the division between specialized and non-specialized senses of colors is not so clear. The boundary between color *term* and color *word*, then, is less distinct.

4.3 Treatment of groups of colors

As discussed above, the terminology used to define groups of colors varies from dictionary to dictionary. The French dictionaries that included such terminology, i.e., TLFi, DAF8 and GR are not consistent in labeling all colors of a given group. For instance, in TLFi, the five colors *rouge*, *jaune*, *bleu*, *indigo* and *violet* are each defined as one of the "*sept* ('seven') *couleurs fondamentales*". *Orangé* is defined as one of the "*sept couleurs primitives*". One must assume that *vert* is the intended seventh color, but its definition includes neither of these group names, being defined curiously by wavelength, as mentioned above. TLFi includes the following subentry in the entry for *couleur*:

Couleur fondamentale (ou primaire). Chacune des **trois** couleurs à partir desquelles l'on peut reproduire toutes les autres par des mélanges en proportion convenable.
[Fundamental (or primary) color. Each of the **three** colors which, when mixed in appropriate amounts, can form all other colors.]

According to this definition, only *rouge*, *bleu* and *vert* should be labeled as being members of the group called *fondamentale*, which is not the case. There is no definition of *couleur primitive* in this dictionary.

DAF8 is less consistent still, defining only two colors as members of a group, and assigning two different names to the group itself. *Indigo* is defined as one of the *sept* ('seven') *couleurs primitives* and *orangé* is defined as one of the *sept couleurs fondamentales*. Neither of these groups is defined in the dictionary itself. It is therefore not clear whether the *sept couleurs primitives* and the *sept couleurs fondamentales* refer to the same group of color terms.

In GR, *bleu* and *jaune* are defined as one of the *sept couleurs fondamentales* of the solar spectrum. *Indigo* is called one of the *couleurs fondamentales du spectre* ('spectrum'), with no mention of the number of colors in the set. The following sentence is found in the entry for *rouge*: "*Le rouge (orangé) est une des sept couleurs fondamentales*" [(Orangish) red is one of the seven fundamental colors]. There is no reference to the spectrum here. If all four colors defined as *fondamentale* in some manner (*bleu*, *jaune*, *indigo* and *rouge*) are included, this means that four out of the stated seven are marked as such. *Vert*, *violet* and *orange/orangé* are not defined as belonging to any of these groups.

Of the three subtractive primaries (*cyan*, *magenta* and *jaune*), only *cyan* and *magenta* are marked in GR as belonging to a group: "une des trois couleurs monochromatiques fondamentales utilisées dans la reproduction des images

polychromes" [one of the three monochromatic fundamental colors used in the reproduction of polychromatic images].

The following note, marked Sc for *scientifique* ('scientific'), is part of the entry of the word *couleur*:

♦ Sc. La sensation de couleur est fonction des propriétés physiques de la lumière (longueur d'onde) et de sa diffusion. La lumière blanche (solaire) est décomposée par le prisme en couleurs dites spectrales (→ Violet, indigo, bleu, vert, jaune, orangé, rouge). Les couleurs du spectre, du prisme, de l'arc-en-ciel (→ Arc-en-ciel, cit. 3 et 5). Couleurs simples, primitives. **Couleurs fondamentales: le jaune, le rouge et le bleu**, couleurs à partir desquelles on peut produire les autres couleurs. Couleurs composées. Couleur complémentaire (d'une couleur primaire), celle qui résulte du mélange des deux autres couleurs primaires. Le vert, couleur complémentaire du rouge (l'orangé, du bleu; le violet, du jaune). Le mélange optique d'une couleur et de sa couleur complémentaire donne le blanc. (GR)

[♦ Sc. The sensation of color is the result of the physical properties of light (wavelength) and its diffusion. White (solar) light is divided into colors called spectral colors by a prism. (→ Violet, indigo, blue, green, yellow, orange, red). The colors of the spectrum, prism, rainbow (→ Rainbow, cit 3 and 5). Basic, primitive colors. **Fundamental colors: yellow, red, and blue**, colors from which one can produce other colors. Blended colors. Complementary color (of a primary color), that which is the result of the mixture of two other primary colors. Green is complementary to red (orange to blue, violet to yellow). The mixture of light of a color and its complementary color create white.]

From this paragraph, *les couleurs fondamentales* should include *jaune, rouge* and *bleu*. However, *vert, violet* and *orange/orangé* are in fact labeled as *fondamentale* in their definitions in this dictionary. There is no reference to the solar spectrum in the explanation of *fondamentale* above, which leads one to believe that there are (at least) three sets of *couleurs fondamentales* considered in this dictionary: *les couleurs fondamentales, les couleurs fondamentales du spectre (solaire)* and *les couleurs monochromatiques fondamentales utilisées dans la reproduction des images polychromes*. Of these three sets, none labels all of its members.

The British dictionaries OED2 and COD12 give information about primary colors. COD12 labels *cyan, magenta* and *yellow* as primary subtractive colors and lists their complementary colors. There is no mention, however, of these complementary colors forming their own group of primary colors. OED2 labels three colors as primary in some manner:

yellow

1. a. ... constituting one (the most luminous) of the primary colours ...

orange

5. (More fully orange-colour)... one of the so-called seven colours of the spectrum, occupying the region between red and yellow ...¹⁵

indigo

3. The colour yielded by indigo, reckoned by Newton as one of the seven prismatic or primary colours ...

These are the only members of these three groups ("primary colours", "so-called seven colours of the spectrum" and "prismatic or primary colours") to be labeled as such. Furthermore, although it is not clear from the definition of yellow if "primary" refers to spectral colors or subtractive primary colors, it appears that "primary" means subtractive primary for *yellow* but spectral for *indigo*.

CED is another British dictionary that labels primary colors:

Red is the complementary colour of cyan and forms a set of primary colours with blue and green
Yellow is the complementary colour of blue and with cyan and magenta forms a set of primary colours
Green is the complementary colour of magenta and with red and blue forms a set of primary colours
Blue is the complementary colour of yellow and with red and green forms a set of primary colours
Cyan 1. a highly saturated green-blue that is the complementary colour of red and forms, with magenta and yellow, a set of primary colours
Magenta 1. a deep purplish red that is the complementary colour of green and, with yellow and cyan, forms a set of primary colours

Figure 5: Groups of colors in CED

As shown in figure 5 above, this dictionary acknowledges the existence of two sets of primary colors — additive and subtractive, but does not use these terms. It does, however, mention the complementary color, the member of the other set of primaries that either absorbs it or is absorbed by it. For instance, cyan absorbs red, so cyan and red are complementary colors. Pigments of the subtractive primaries (cyan, magenta and yellow) absorb the additive primaries (red, green and blue).

The treatment of color groupings in the dictionaries consulted is another example of the difference in importance of scientific and technical information in French, British and American dictionaries. The American dictionaries consulted and CED were more consistent with their labels than the French diction-

aries or the other British dictionaries. *Webster's Third New International Dictionary* (W3) and AHD5 were the American dictionaries consulted that marked colors as primary. The following charts show the terms used for groups of colors, and the colors that are defined as belonging to this group.

Label	Colors
(four) psychologically primary hues	red, green, yellow, blue
(six) psychologically primary object colors	red, green, yellow, blue, white, black
subtractive primaries	yellow, magenta, cyan

Figure 6: Groups of colors labeled in W3

Label	Colors
additive or light primaries	red, green, blue
psychological primary hues	red, yellow, green, blue
subtractive primary	yellow, magenta, cyan

Figure 7: Groups of colors labeled in AHD5¹⁶

Both of these dictionaries label all colors in a group. There are, however, differences in the categories chosen and the labels used to identify their members. The subtractive primary colors and the psychological(ly) primary hues are the same in W3 and AHD5. W3, however, does not label additive primaries, but does include a category called "psychologically primary object colors". No group with this name is found in any of the dictionaries or literature consulted in this study. The group of colors, however, matches the Hering psychologically primary colors discussed in 2.4 above. One must suppose that it is this concept of psychological primacy being referred to, but the origin of the label "psychologically primary object colors" is unclear.

Dictionary	Group	Color(s)
GR	(trois) couleurs monochromatiques fondamentales utilisées dans la reproduction des images polychromes	magenta, cyan
GR	(sept) couleurs fondamentales	rouge
GR	(sept) couleurs fondamentales du spectre solaire	bleu, jaune
GR	couleurs fondamentales du spectre	indigo
TLFi	(sept) couleurs primitives	orangé

TLFi	(sept) couleurs fondamentales	rouge, bleu, jaune, violet, indigo
DAF8	(sept) couleurs primitives	indigo
DAF8	(sept) couleurs fondamentales	orangé
W3	(four) psychologically primary hues	red, green, blue, yellow
W3	(six) psychologically primary object colors	red, green, blue, yellow, white, black
AHD5	additive or light primaries	red, blue, green
AHD5	psychological primary hues	red, green, blue, yellow
AHD5	subtractive primary	magenta, yellow, cyan
RHD	primary color	red, blue
RHD	secondary color	orange
CED	a set of primary colors	red, blue, green
CED	a set of primary colors	magenta, yellow, cyan
COD12	one of the primary subtractive colours	magenta, yellow, cyan
OED2	primary colours	yellow
OED2	one of the so-called seven colours of the spectrum	orange
OED2	one of the seven prismatic or primary colours	indigo

Figure 8: The color terms defined as members of a group in our data

The chart above shows that while none of the French dictionaries (GR, TLFi and DAF8) defined all members of *any* group, the American dictionaries W3 and AHD5 were consistent in the terms used and in defining all colors of a group. RHD was the only American dictionary not to do so. While the British dictionary CED was consistent in its group labeling, COD12 included only one group (subtractive primary colors) and OED2 labeled three colors each with their own unique label. While it is clear that American dictionaries label groups of colors more often and more consistently within one dictionary, there does not seem to be a consensus on which groups of colors to label or what terminology to use.

5. Conclusion

Lexicographers face the difficulty of deciding how to treat specialized terms, particularly those that are polysemous with a non-specialized word. We found that specialized and non-specialized senses are not split in color definitions in the French, British and American dictionaries consulted. Specialized and non-specialized senses of colors were shown to be overlapping. This suggests that the distinction between color words and color terms is not as distinct as it is for

words like *star* and *tree*.

The type of scientific information included in definitions in our data was primarily concerned with the spectrum of light. Specifically, definitions referenced the wavelength with which the color is associated and/or the color's position on the spectrum. The dictionaries to include wavelengths were the American dictionaries AHD1, AHD5 and RHD, as well as the British dictionary CED. The French dictionaries GR and TLFi included the wavelength only for *vert*. Overall, American dictionaries were found to include more scientific detail than French and most British dictionaries, validating claims made in Béjoint 1988 and 1994 that it is particularly important in American dictionaries. Our study confirmed that specialized information has, however, increased in importance in the dictionaries of all three nationalities. This confirms Béjoint 1988 and Landau 1974 that scientific information has risen in importance over time.

Another type of specialized information in our data was the defining of colors as members of a group, such as *primary* or *fondamentale*. With regards to the grouping of color terms, with the exception of CED and COD12, British and French dictionaries were found to be inconsistent both in defining all members of a group as such and in the naming of groups themselves. This is consistent with the observation that American dictionaries place a higher value on specialized information.

Based on our findings, we make the following suggestions for defining colors in a general use dictionary:

- a. If one color in a certain group (primary, spectral etc.) is labeled as belonging to that group, all members should be. This would solve the inconsistencies such as those seen in GR, TLFi, DAF8, OED2 and RHD.
- b. Both the choice of groups to define and the labels chosen for them should be considered carefully. Selection should be based on the probability of the groups and labels being meaningful to the reader, likely because they are working in a color-related field such as physics or art. W3's choice of defining the Hering primaries as "psychologically primary object colors" is questionable, as we have not found this term in any other source as of April 2017.
- c. The terminology used to label the group should be defined in the dictionary. Cases like OED2, GR and TLFi show that this is necessary. Labels are not helpful if it is unclear what they refer to.

Ideally, in order to follow these suggestions, colors should be either assigned to one editor or defined according to a set of guidelines. In addition to correcting the confusion caused by the cases above, this would eliminate instances such as those found in GR and TLFi, where only one color is defined by the wavelength at which it is observed, as well as OED2, COD1 and COD12 where some spectral colors are defined based on their position on the spectrum and others are not.

Endnotes

1. Unless otherwise noted, English translations in this article are ours.
2. For descriptions of the methodology lexicographers use to determine dictionary user needs and expectations, see Schierholz 2015 and Béjoint 1994: 140-168.
3. As an anonymous reviewer noted, it would be interesting to see the results of recent studies on whether these preferences hold true in the age of the internet and online dictionaries. As we will see below, dictionaries still reflect these differences, but this could be due more to tradition than to current user expectations.
4. In reality, there is no clear division among these seven colors, or any scientific reason for the rainbow to be divided into seven segments. This seven-part division is traced back to Isaac Newton, who divided the spectrum into seven colors to reflect the seven notes in a diatonic musical scale, and because he considered the number seven to be lucky. He had earlier divided the spectrum into five colors, but then added orange and indigo to bring the number to seven (Shapiro 1994: 619).
5. The original adjectival form of *orange* was *orangé*. Some dictionaries, the *Grand Robert de la Langue Française*, for example, still include *orangé* in their list of spectral colors instead of *orange*.
6. Other sets of additive primaries are possible, based on the purpose of color mixing. The additive primaries used for television sets, for example, are a somewhat orange red, a slightly unsaturated blue and a yellowish green. These deviations from red, blue and green are made in order to compensate for the loss of brightness caused by the phosphors at the long and short ends of the spectrum not being very intense. Tiny dots of these colors placed close together produce the illusion that we are looking at mixtures of these colors. This process is similar to that used in impressionism (Sobel 1987: 62).
7. AHD2 is a collegiate edition.
8. This is a digitized version of the *Collins Concise English Dictionary*, 12th edition.
9. Because the OED's color definitions have been revised decades apart, if at all, we chose the latest print version, OED2 (1989).
10. Here and elsewhere in the text, we have bolded parts of definitions to which we want to draw special attention.
11. DAF9 is being edited alphabetically: thus *rouge*, *vert* and *violet* are not yet included.
12. Defining cyan this way is very unexpected, as it is not one of the traditional seven spectral colors.
13. Blue is defined in COD12 as "of a colour intermediate between green and violet ...", which does not overtly mention the spectrum. The spectrum is, however, implied, as blue does fall between these two colors on the spectrum.
14. It is unclear why she did not follow this philosophy for the term *vert* in TLFi.
15. The definition of orange was updated in 2004 and no longer uses the term "spectral colour". It does still give the color's spectral position.
16. All three of these groupings are defined within this dictionary.

Acknowledgements

This article is based on a section of the doctoral dissertation, *The Lexicographic*

Treatment of Color Terms, defended in May 2014 at Indiana University, Bloomington. Thanks go to Kevin Rottet and Michael Adams for their guidance at that stage and anonymous reviewers for their assistance with this version. Thanks also to Rachel Hurley for her editorial assistance.

References

- The American Heritage Dictionary*. 1991. Second college edition. Boston, MA: Houghton Mifflin.
- The American Heritage Dictionary of the English Language*. 2011. Fifth edition. Boston, MA: Houghton Mifflin Harcourt.
- Allen, R. 1999. Lumping and Splitting. *English Today* 60: 61-62.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bagard, G. 2009. *La coloration: Méthode globale*. Dompierre sur Mer: GBvéga éditions.
- Béjoint, H. 1988. Scientific and Technical Words in General Dictionaries. *International Journal of Lexicography* 1(4): 354-368.
- Béjoint, H. 1990. Monosemy and the Dictionary. Magay, T. and J. Zigány (Eds.). 1990. *BudaLEX '88 Proceedings: Papers from the 3rd International EURALEX Congress, Budapest, 4-9 September 1988*: 13-26. Budapest: Akadémiai Kiadó.
- Béjoint, H. 1994. *Tradition and Innovation in Modern English Dictionaries*. Oxford: Clarendon Press.
- Béjoint, H. 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Catach, L., P. Robert, A. Rey, D. Morvan and L. Laporte. Continually updated resource. *Le Grand Robert de la langue française*. Paris: Dictionnaires Le Robert.
- Collins English Dictionary*. 2017. 12th edition [online]. Glasgow: HarperCollins. Available at: <http://www.collinsdictionary.com>.
- Dictionnaire de l'Académie française*. 1932-5. 8th edition. Chicago: ARTFL Project Department of Romance Languages and Literatures Division of the Humanities University of Chicago, 2010.
- Dictionnaire de l'Académie française*. 1992-ongoing. 9th edition. Paris: ATILF. <http://atilf.atilf.fr/academie9.htm>.
- Flexner, S.B. and L.C. Hauck (Eds.). 1987. *The Random House Dictionary of the English Language*. Second edition. New York: Random House.
- Fowler, H.W., F.G. Fowler and D. Crystal. 2011. *The Concise Oxford Dictionary of Current English: 1911 First Edition*. Oxford: Oxford University Press.
- Gove, P.B. 1961. *Webster's Third New International Dictionary of the English Language. Unabridged: A Merriam-Webster*. Springfield, MA: G. & C. Merriam
- Ings, S. 2008. *A Natural History of Seeing: The Art and Science of Vision*. New York: W.W. Norton.
- Landau, S.I. 1974. Scientific and Technical Entries in American Dictionaries. *American Speech* 49: 241-244.
- Landau, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. New York/Cambridge: Cambridge University Press.
- Mollard-Desfour, A. 2012. RE: *Trans.: Les couleurs dans le TLFi*. Email message to K. Williams, 29 September 2012.
- Moon, R. 1987. Monosemous Words and the Dictionary. Cowie, A.P. (Ed.) 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985*. Lexicographica Series Maior: 173-182. Tübingen: Niemeyer.

- Rey, A. and S. Delesalle.** 1979. Problèmes et conflits lexicographiques. *Langue Française* 43(1): 4-26.
- Schierholz, S.J.** 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25: 323-352.
- Shapiro, A.E.** 1994. Artists' Colors and Newton's Colors. *Isis* 85(4): 600-630.
- Simpson, J.A. and E.S.C. Weiner (Eds.).** 1989. *The Oxford English Dictionary*. Second edition. Oxford: Clarendon Press.
- Sobel, M.I.** 1987. *Light*. Chicago: University of Chicago Press.
- Trésor de la langue française informatisé.* 2000. Paris: ATILF. <http://atilf.atilf.fr/>.
- Stevenson, A. and M. Waite.** 2011. *Concise Oxford English Dictionary*. 12th edition. Oxford: Oxford University Press.
- Vrbinc, M. and A. Vrbinc.** 2013. Differences in the Inclusion and Treatment of Terminology in OALD3, OALD4 and OALD8. *Lexikos* 23: 440-455.
- Waldman, G.** 2002. *Introduction to Light: The Physics of Light, Vision, and Color*. New York: Dover.
- Zgusta, L.** 2006. *Lexicography Then and Now: Selected Essays*. Ed. by Dolezal, F.S.F. and T.B.I. Creamer. Lexicographica. Series Maior 129. Tübingen: Max Niemeyer.

Revisiting Lemma Lists in Swahili Dictionaries

Beata Wójtowicz, *Department of African Languages
and Cultures, University of Warsaw, Warsaw,
Poland (b.wojtowicz@uw.edu.pl)*

Abstract: When compiling a dictionary, a lexicographer has a set of decisions to make — starting with drawing up a lemma list to such issues as formatting a dictionary entry. Relying on corpus data while designing a lemma list and describing entries is standard in present lexicography, but there are still decisions — like the choice of a lemma or how to treat derivatives — that are often intuition-based. This article aims to investigate whether decisions put forward in Swahili dictionaries comply with users' expectations. We analyse log files from the new Swahili–Polish dictionary to investigate why looking up words goes wrong, and evaluate the choice of a lemma and the treatment of derivatives in Swahili dictionaries. Based on such data we intend to expand or modify the existing electronic dictionary to adapt to users' level of grammar and dictionary structure knowledge. During this research we identified a list of lemma lacuna that cause the majority of unsuccessful Swahili searches. The study shows that users know and understand the lemmatisation strategy of the dictionary but also reveals which word forms cause the most problems and how the lemma list of Swahili dictionaries could be expanded.

Keywords: DICTIONARY USER RESEARCH, LOG FILES ANALYSIS, SWAHILI–POLISH DICTIONARY, LEMMA LIST, DERIVATIVES

Opsomming: Die herbesoek van lemmalyste in Swahili-woordeboeke. Wanneer 'n woordeboek saamgestel word, moet 'n leksikograaf 'n reeks besluite neem — van die opstel van 'n lemmalys, tot kwessies soos die formatering van 'n woordeboekinskrywing. Om staat te maak op korpusdata wanneer 'n lemmalys opgestel en inskrywings beskryf word, is standaardpraktyk in die huidige leksikografie, maar daar is steeds dikwels besluite — soos die keuse van 'n lemma of hoe om afleidings te hanteer — wat op intuïsie gebaseer is. Hierdie artikel beoog om te ondersoek of besluite wat in Swahili-woordeboeke geneem is, voldoen aan gebruikers se verwagtings. Ons analiseer loglêers van die nuwe Swahili–Poolse woordeboek om te ondersoek waarom die opsoek van woorde skeefloop, en evalueer die lemmakeuse en die hantering van afleidings in Swahili-woordeboeke. Ons beoog om die bestaande elektroniese woordeboek op grond van hierdie data uit te brei of te wysig om aan te pas by gebruikers se vlak van kennis ten opsigte van grammatika en woordeboekstruktuur. Tydens hierdie navorsing het ons 'n lys van leemtes ten opsigte van lemmas geïdentifiseer wat die meerderheid van onsuksesvolle Swahili-soektogte veroorsaak. Die navorsing toon dat gebruikers die woordeboek se lemmatiseringstrategie ken en verstaan, maar openbaar ook watter woordvorme die meeste probleme veroorsaak en hoe die lemmalys van Swahili-woordeboeke uitgebrei kan word.

Sleutelwoorde: WOORDEBOEKGEBRUIKERSNAVORSING, LOGLÊERONTLEDING, SWAHILI-POOLSE WOORDEBOEK, LEMMALYS, AFLEIDINGS

1. Introduction

When a new dictionary is compiled a reference to a corpus is a standard procedure (cf. De Schryver et al. 2006). We use a corpus in different ways in dictionary production, either for updating an already existing dictionary or for compiling a brand new dictionary from scratch (cf. Atkins and Rundell 2008: 97). The vexing question has always been what to include in a dictionary. We may rely on a frequency list derived from a corpus during the headword selection, but already at this stage we have to make decisions as to the choice of lemma or how to treat various items, like derivatives or multiword expressions. Texts are made of word forms whereas in dictionaries we expect lemmas, the so-called dictionary headwords or citation-forms. The choice of a citation-form has direct impact on the ways the user conducts look-ups of certain items and how much knowledge of grammar is needed to successfully consult the dictionary.

The dictionary data should be presented in such a way that a user can easily access it. The success of finding the data depends on the access structure (cf. Gouws and Prinsloo 2005), that has undeniably changed in the era of electronic lexicography. In an electronic dictionary it is often the lemma itself that provides access to the article, and therefore the choice of lemma is a crucial decision the lexicographer has to make. This is especially true in case of Bantu languages where lemma is not intuitive and in some cases not identical to any word forms. The user has to learn basic grammar and know the structure of a dictionary to be able to use it. Thus far it was not known how the users manage but now, in the era of electronic lexicography, we can observe, to some extent, users' behaviour and choices. It gives us possibility to modify the shape of a lemma and a content of the lemma list to respond to users' needs.

No matter how good the theories and methods used in compiling a dictionary are, it is the user who ultimately evaluates the usefulness, efficiency and user-friendliness of a particular lexicographic work. Therefore all lexicographic decisions have to be taken with the user in mind and especially the users' skills must be taken into account (cf. Atkins and Rundell 2008, Prinsloo and De Schryver 1999). Building on this assumption, a new Swahili-Polish dictionary¹ was created and posted online as a student resource. The dictionary has been created as an electronic resource, but its printed counterpart has also been published (Wójtowicz 2013). It contains over 6 000 Swahili entries and over 7 000 entries in a searchable Polish index in the electronic version. The dictionary's lemma list is mainly based on a Helsinki Corpus of Swahili-derived frequency list (HCS 2004) of over ten thousand lemmatised entries. It is targeted at learners of Swahili, who already have some knowledge of language grammar and the structure of a Swahili dictionary. This paper aims to address the question of how well the users meet these constraints by investigating the

log files of the dictionary. We intend to research whether users know how to access Swahili dictionary articles — that is their lemma choices as compared to dictionary lemma list. Various decisions concerning its macro- and microstructure were made in accordance with the Swahili lexicographic tradition. Therefore, even though we are investigating the log files of this particular dictionary, we intend to relate our findings to other dictionaries of Swahili as well.

To meet these goals, in subsequent sections we will report on the choice of citation-forms and the treatment of derivatives in various Swahili dictionaries. To make an evaluation we analyse the log files of a Swahili–Polish dictionary that was made available online four years ago. We aim to examine strings that were used during look-ups and then compare them with what the dictionaries offer as a look-up strategy. We also intend to investigate the reasons why looking up words does not always goes well.

2. Log files as a tool of dictionary user research

In his seminal work, Samuel Johnson recognises users as an integral part of the lexicographic process (Atkins and Rundell 2008: 5) and the logical expectation by now would be that research into dictionary use and user profiles is already well established (cf. Lew 2011b, Lew and De Schryver 2014, Töpel 2014). However, analysis of log files — which alongside questionnaires, protocols or observation, is one of the approaches to the study of dictionary use (Lew 2011a) — is not exploited as often as one could expect. Töpel (2014) reports on the unsatisfactory situation noting that the dictionary users and their actions are still unknown and that more research is still needed.

According to Bergenholtz and Johnsen (2005), analyses of log files may be used as a tool for improving Internet dictionaries. They consider log files to be a useful supplement to corpus-based lemma selection since they reveal lemma lacuna, frequent misspellings, frequency of searches for multiword units, etc. De Schryver and Joffe (2004: 188) opine "that an automated analysis of the log files will enable the dictionary to tailor itself to each and every particular user".

Even though such possibilities seem tempting, the limitations of this source have to be taken into consideration as well (cf. Lew 2011a). Lew (2011a) remarks that log files don't give answers on the context of dictionary use or on the user himself. With or without these limitations in mind researchers investigate log files most often to reveal which lemmas have been successfully retrieved; which lemmas have been requested but were not found; which lemmas have been looked up and how often; and which words have been used in a search field. Based on this information, one can modify the content of a dictionary to meet the users' needs.

The log files analysis was reported already in several studies that studied different aspects of dictionary use. In the study of Laufer and Hill (2000), who investigated what information is looked up and how unknown vocabulary is retained, a log file analysis was combined with a vocabulary test to check for

vocabulary retention. In their study, Laufer and Hill (2000) and Nesi (2000) exploit log file analysis to compare electronic and printed dictionaries. A different approach was adopted by Lemnitzer (2001), who was interested in the reasons why looking up words goes wrong. He tested his assumptions on the log files of four bilingual electronic dictionaries. The study of De Schryver and Joffe (2004) also concerns files obtained from the normal use of an electronic dictionary rather than of a specially designed context. The authors for the first time ask a question later investigated more deeply in further studies (cf. De Schryver et al. 2006, Verlinde and Binon 2010, Koplenig et al. 2014, Trap-Jensen et al. 2014, Müller-Spitzer et al. 2015), about whether dictionary users actually look up frequent words. The research leads to an important conclusion: it seems that there is a relationship between the corpus frequency and the frequency of look-ups. Müller-Spitzer et al. (2015) further claim that "frequency does matter — even in lower frequency bands", which matters enormously in the era of corpus-lexicography. This is because dictionary compilers are provided with evidence about the value of corpus-based lexicography as the users frequently look up frequent words even beyond the first few thousands. The frequency list thus remains the main source of data especially when compiling a small dictionary lemma list. Another study of Bergenholtz and Johnsen (2005) shows how to use log file analysis method as a tool for improving electronic dictionaries. They report on the sources of unsuccessful searches and user search behaviour.

The studies show that analysing log files successfully leads to filling gaps in dictionary lemma lists and discovering other problems related to unsuccessful look-ups. However this kind of research does not address all issues concerning dictionary user studies. In order to get a full picture of the dictionary consultation process we need wider studies that combine all of the different methods. None of the methods answers all of the questions related to dictionary use.

3. Citation-forms in Swahili dictionaries

De Schryver et al. (2006: 68) notice that "not all primary speakers of Swahili can look up 'words' in their own language (as this implies being able to cut off pre- and suffixes), and even trained learners and scholars often need more than one look-up round before they find what they are looking for (as sound changes between formatives are not always predictable)".

Kosch (2013: 202) further elaborates that how well users cope with looking up words in a Bantu language dictionary and to what extent their expectations are met, is largely dependent on such factors, like consultation skills — that is, their previous exposure to dictionary pedagogy, their knowledge of the structure of a Bantu language, and the dictionary design itself. Bantu language learners often mimic their dictionary habits from non-Bantu language dictionaries and are not aware of a dictionary design owing to the agglutinative

structure of the language, which calls for a specialised approach to lemmatisation.

The most important part of a consultation process concerns the choice of a citation-form. A citation-form serves different functions, but first of all, it is a form by which the user will use a dictionary in order to find other phonological, morphological, syntactic, semantic, and etymological information associated with it (cf. Kiango 2005).

In the case of languages with long lexicographical traditions, the citation-forms have long been determined and lexicographers easily avoid the problem of discrepancies between dictionaries. But the question remains of how the users manage with the lexicographers' decisions. It is they who have to lemmatise word forms found in texts. It is assumed that the user knows how to search in a dictionary. As an aid there are guidelines in the introductory part. But still, it is expected that users learn the rules before conducting a search. We can only guess how often they comply with this demand.

The methodology of lemma selection, which has been well researched for European languages and is above all based on Latin, cannot always be without proper scrutiny applied to languages for which the grammatical structures significantly differ from those of Indo-European languages (cf. Knowles and Mohd Don 2004). According to lexicographical recommendations, all forms which naturally come to mind to users when searching a dictionary should function as headwords. But due to the complex morphological structure of Bantu words, the choice of the citation form is not always obvious (cf. Kiango 2000). Prinsloo and De Schryver (1999) give a comprehensive introduction to the lemmatisation strategies in Bantu languages. There are two lexical traditions applied to the Bantu languages. These are word traditions with lemmas based on complete written words, and stem tradition with lemmas based on the stems of written words without their prefixes.

Swahili is a Bantu language and as such, it is characterised by agglutinative morphology and a noun class system, whose reflexes are manifested both lexically, on the noun, and syntactically, via agreement. This is demonstrated in the six sentences below, where the only crucial lexemic variable is the initial noun, with which all other elements have to agree (CL stands for a class marker).

- (1) a. M-tu huyu m-zuri m-moja a-li-ye-anguka
 CL1.man this.CL1 CL1-nice CL1-one CL1-PAST-CL1.REL-fall
 'this one nice man who fell down'
- b. Wa-tu hawa wa-zuri wa-tatu wa-li-o-anguka
 CL2.people these.CL2 CL2-nice CL2-three CL2-PAST-CL2.REL-fall
 'these three nice people who fell down'
- c. M-fuko huu m-zuri m-moja u-li-o-anguka
 CL3.bag this.CL3 CL3-nice CL3-one CL3-PAST-CL3.REL-fall
 'this one nice bag which fell down'

- | | | | | | |
|----|--|------------|-----------|------------|-------------------------|
| d. | Ma-chungwa | haya | ma-zuri | ma-tatu | ya-li-yo-anguka |
| | CL6.oranges | these.CL6 | CL6-nice | CL6-three | CL6-PAST-CL6.REL-fall |
| | 'these three nice oranges which fell down' | | | | |
| e. | Ki-tu | hiki | ki-zuri | ki-moja | ki-li-cho-anguka |
| | CL7.thing | this.CL7 | CL7-nice | CL7-one | CL7-PAST-CL7.REL-fall |
| | 'this one nice thing which fell down' | | | | |
| f. | Kalamu | hizi | n-zuri | tatu | zi-li-zo-anguka |
| | CL10.pens | these.CL10 | CL10-nice | CL10.three | CL10-PAST-CL10.REL-fall |
| | 'these three nice pens which fell down' | | | | |

Each nominal lexeme is classified as belonging to one of fifteen classes, and each of these classes has a default mapping onto the agreement pattern that holds most of the other elements of the sentence. Note that both noun-class membership and the resulting agreement are in most cases manifested by prefixal means (in the above examples only the demonstrative pronoun breaks this rule). This is also visible in (2) below, in the present-tense paradigm of the verb *kuanguka* 'to fall down'.

- (2)
- | | | |
|----|------------------------------------|----------------------------------|
| a. | ku-anguka 'to fall down' | |
| | INF-fall.down | |
| b. | ni-na-anguka 'I fall down' | e. |
| | 1SG-PRES-fall.down | tu-na-anguka 'we fall down' |
| c. | u-na-anguka 'you fall down' | f. |
| | | m-na-anguka 'you (PL) fall down' |
| d. | a-na-anguka 'he/she/it falls down' | g. |
| | | wa-na-anguka 'they fall down' |

All infinitives in Swahili begin with *ku-* (or *kw-*), and practically all verbal stems could be substituted for the stem *anguka* in the paradigm above. Prefixes change depending on the grammatical context, and the only element that remains unchanged is the stem.

In an established Swahili lexicographic tradition on paper dictionaries, noun headwords are introduced in singular form together with the nominal prefix, for example *mtu* 'man'. On the other hand, verbs use sequences identical to the infinitive form, but without the infinitive prefix *ku-*, for example *penda* 'love' (instead of *kupenda*). If the traditional lexicographic practice used e.g. for European languages was followed here in citation form selection, and if verbs were cited in their infinitival forms, most of the space in Swahili dictionaries (and that is also true of many other Bantu languages in which infinitives are morphologically assigned to a noun class 15) would be taken by the letter "k". In fact, all verbs would end up there, which is hardly user-friendly or sensible. A similar problem concerns adjectives: listing their fully inflected forms would result in distributing most of them under the letters of alphabet that their agreement prefixes begin with, which is predictable and regular. As in the case of inflected verbs (cf. the paradigm in example 2), using the full, inflected forms would result in massive redundancy. Therefore, adjectives, numerals, and pro-

nouns are represented by non-prefixal forms, for example *zuri* 'good'. To inform the user that the given word necessitates the addition of a prefix in order to take on a proper form, the headwords are in some dictionaries preceded by a hyphen, for example *-zuri* 'good' (e.g. in Abdulla et al. 2002). Occasionally pronouns, especially demonstrative and possessive, are included in dictionaries in their full form, that is, with a class prefix of a noun they modify, for example *changu* 'mine, class 7' (Baba Malaika 1994). The possible ways we may treat these issues change when we switch to electronic lexicography.

Despite the Swahili lexicographic tradition, the question of the structure of citation-forms in Swahili is still vital. Kiango (2000: 25) recognizes the infinitive as the basic natural form of Bantu verbs. The only reason why we should not put them in such shape in a dictionary is the alphabetical order discussed above. As for the nouns, Kiango (2000: 31) shows that certain nominal stems are unnatural forms, and therefore a noun in a singular form with a prefix is entered as a citation-form in most dictionaries of Bantu languages. However, this method is not convenient for beginners of the language, who are not able to easily identify singular and plural forms of a noun. It could be solved by applying a method of entering both forms into a dictionary, but this violates the principle that "citation-forms should be stems from which other inflected forms could be produced" (Kiango 2000: 31-32). Both forms represent one lexeme, and therefore they cannot be entered as two separate entries. We could violate this principle when handling irregular forms, but still it is against the principle of economy. Again, it can be easily solved in an electronic version of a dictionary, where we can, for example, allow searches on plural forms, which are provided within an entry.

Indeed, the electronic form introduced novelty into accepted solutions. When the *TshwaneDJe Swahili-English Dictionary*² joined the market, it was the first and still remains the only corpus-driven electronic dictionary of Swahili with a new approach to the lemmatisation of headwords (De Schryver et al. 2006). The content of the dictionary is based on web-based corpus data and it includes over 16,000 entries. The most interesting innovation is that the headwords include orthographic forms in addition to stems chosen on the basis of a frequency count. The dictionary does not include morphological analysis as such, but the user may search for the frequently used word forms.

3.1 The treatment of derivatives in Swahili dictionaries

The most important issue with respect to the macrostructure in dictionaries of Bantu languages regards the handling of derivatives, sometimes referred to as the "lumping vs. splitting" debate (cf. Bański and Wójtowicz 2011). Kosch (2013) discusses the issue with reference to the demands that lexicographers place on the users. The dictionary design is then motivated by various expectations, ranging from low-level, where basic look-up skills according to the letters of the alphabet are assumed to have been mastered already; to medium-level

expectations that assume the user is able to look up words in a stem-based dictionary; to high-level demands, where intuitive dictionary skills no longer suffice.

The discussion arises from the fact that Bantu derivational word-families can be extremely numerous, especially those based on verbal roots — for example, De Schryver and Prinsloo (2001: 225ff) count over 140 regular derivatives of the root *reka* 'buy, purchase' in Sepedi. The problem that we see deals with presenting the derivatives to the user, and the basic question is whether to lump them all into the entry of the root, or whether to distribute them across the dictionary giving each derivative the status of a main entry.

Radical lumping means cramming all the information into a single place in a dictionary and often denying independent status to the most commonly used words only because they happen to be derivatives; additionally, the user is required to know word-formation mechanisms (which are always less transparent than the inflectional system) in order to identify the base form of the given derivative, which means searching for e.g. *utumishi* 'civil service' in the entry for *-tuma* 'assign/give work to sb', which e.g. in Johnson's (1939/1985) dictionary takes up half a page. De Schryver and Prinsloo (2001: 224) report on a similar case in a Sepedi–Afrikaans–English dictionary, where the entry for *reka* and its derivatives takes up an entire page of dense print. There is also the possibility of partial lumping when some derivatives are kept with their roots and some are given the status of a main entry. Kiango (2000) recommends lumping only regular forms, but this does not help if it is the regular forms that happen to have high frequency and should therefore be presented to the user in separate entries.

The choice of the splitting strategy may be considered primarily practical — as Zgusta (1971: 16) notes, "[...] we must not forget that the lexicographer is doing scientific work, but that he publishes it for users whose pursuits are always more practical, at least as regarded from his own point of view." Given this, a user-friendly dictionary of Swahili should list the broadest possible range of derivatives as separate entries. It is not without reason, as Herms (1999) notes, that students of Swahili praise the "friendly" dictionary of Baba Malaika (1994), who adopts this kind of distributed approach to derivatives, while De Schryver and Prinsloo (2001) note the lack of popularity of dictionaries that group words on etymological grounds and/or under verbal or nominal roots.

The lumping approach has one undeniable virtue — it keeps word families together. In agglutinative languages, it can even show how the particular members are derived (cf. Bosch et al. 2007). Scattering derivatives across the entire dictionary means severing the lexical and semantic ties between closely related lexemes and practically hiding some of them from users looking up the root form.

Clearly, both approaches have their virtues and their disadvantages, and both of them have been suggested in the literature and used in practice. Bosch et al. (2007) argue for lumping (in South Bantu languages), and that is what

Johnson (1939/1985) did for Swahili. De Schryver and Prinsloo (2001) argue for splitting and this is what Sacleux (1939) and TUKI (2001) have implemented to varying degrees of success. Madan (1903/1992) and Abdulla et al. (2002) represent mixed approaches, whereby derivatives representing the same part of speech as the derivational base are placed inside the entry of that base form, while derivatives representing different parts of speech are listed separately.

Due to above problems it is really important that the users of a dictionary understand the lemmatisation approach used in a certain dictionary (cf. De Schryver et al. 2006).

The issue can now be, at least partially, addressed in an electronic form as showed in a Swahili–Polish dictionary where derivational families are presented to the user in the form of a searchable graph. The lemma-oriented approach is no longer necessary in electronic dictionaries as it is possible that each segment of the dictionary may be viewed in different ways by different users, thus eliminating the need for any kind of macrostructure.

4. Log files of a Swahili–Polish dictionary

The Swahili–Polish dictionary user queries have been recorded and saved since January 2013. They are noted in four log-files, namely: Swahili found entries, Swahili not-found, Polish found and Polish not-found. The users may search for Swahili and Polish headwords, but they have to choose the language before they conduct a search. The files note a string the user has typed in the search box and the number of searches for that certain string. Neither IP nor any other identification is taken into account. We assume that many of the users are university students who use the university network. This has been also attested in the data itself, as indicated below.

The data is saved in .csv files and the analysis was carried out by the author manually and with the use of regular expressions only. In the Swahili found file, additionally to the string and the number of searches, information on POS and ID of an entry that was returned to the user is provided. Additional analysis may be conducted on data from the Google Analytics that has been launched as well.

The structure of the dictionary allows a search for Swahili headwords and plural forms of nouns, which are included in the dictionary in their full forms. Derivatives like pronouns with class prefixes and irregular verbal forms are treated as separate entries and users can also search for them. The most difficult operation is the decomposition of the verbal complex. The user has to cut off all prefixal morphemes and search for the verbal root or extended root instead.

Since we are interested in revealing whether users know the search strategies of Swahili dictionaries implemented in this dictionary as well, and whether they know how to choose an appropriate lemma to find what they are looking for, we will investigate mainly the Swahili not-found log file in com-

parison to the Swahili found file. As the log files only record what the users have entered into the search box, it should be stressed that we can never be sure who the user was or the user's actual intentions (cf. Lew 2011a). We are also aware of the limitations of this study as we investigate a very narrow group of dictionary users — that is, mostly Polish students of Swahili. Many of them were therefore trained how to use dictionaries of Swahili; they know the grammar of the language and the structure of the dictionary under investigation. Nevertheless many unsuccessful searches were still recorded and it was interesting to reveal their sources.

Over a four-year period, up to 15 February 2017, 53,592 queries were made. This makes for an average of 36 look-ups per day, with an increase observed in searches over time — from 25 in 2013/14 to 46.5 since 2015. The number falls during holidays, especially the long summer vacations, to an average of 15 look-ups per day, which corresponds to our assumption that the dictionary is mostly used by students of the language.

The number of look-ups noted in each file is 25,466 (48%) in the Swahili found, 14,708 (27%) in the Swahili not-found, 8,052 (15%) in the Polish found, and 5,366 (10%) in the Polish not-found. Thus, the majority of the look-ups (75%) were in the Swahili–Polish direction. When we compare only the searches for Swahili entries, 63% of them are noted as found and 37% as not-found.

As for the number of unique strings searched for, 4,430 strings were noted in the Swahili found, 8,912 in the Swahili not-found, 2,366 in the Polish found, and 3,364 in the Polish not-found. The numbers are much higher in the not-found sections, but the majority of these look-ups, like 73% of the strings in the Swahili not-found, were looked up only once. In comparison, in the Swahili found file only 29% of the strings were looked up once.

4.1 The analysis of the searches

The problems identified while analysing log files may guide us towards a decision on how to improve the dictionary. By identifying missing lemmas we may improve the coverage of the dictionary, while other issues may influence our approach towards the lemmatisation strategy or the search method. For example, in their study based on log files analysis, Bergenholtz and Johnsen (2005) argue for including additional verbal forms, like imperative and passive in a Danish dictionary, or expanding it by adding some of the items identified as lemma lacuna. On the other hand, they did not find the problem of misspellings to be of an important nature.

A study of the log files from the Swahili–Polish dictionary revealed a number of specific problems encountered by the users while consulting the dictionary. They fall into several categories. To identify these categories, the top 500 strings of the Swahili not-found file were analysed one by one and annotated as a Polish word, an orthographic word, a wrongly lemmatised form, a

spelling mistake, a proper name, a multiword expression and a lemma lacuna. The number of strings identified in each category is presented in figure 1.

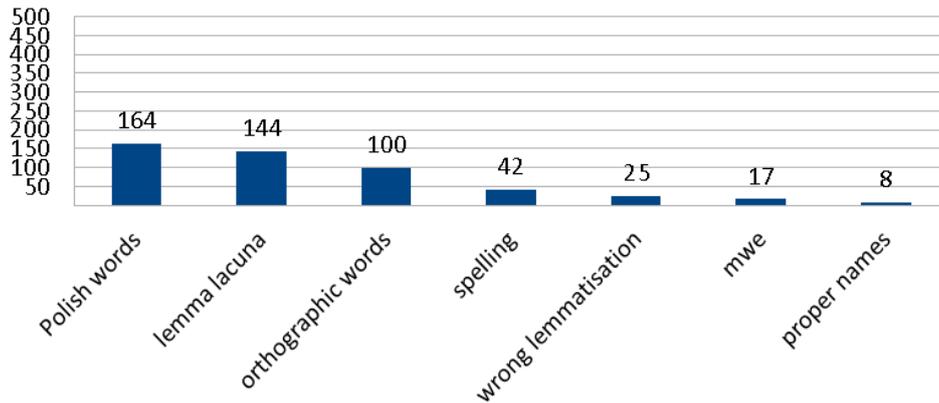


Figure 1: Number of strings per category among the top 500

The top 500 strings were searched for 3,661 times. That constitutes 26% of all searches. The remaining 74% of the searches were for the other 8,400 strings.

The first problem we notice concerns the choice of language. Since the user has to switch modes in order to look for Polish words, unfortunately as many as 11% of the searched for strings that are noted in the Swahili not-found file were identified as Polish words. They stand for 25% of the strings searched for at least twice. Among the top 100 searches in a Swahili not-found file 39 were identified as Polish words and 164 were identified among the top 500. This is also the case with the Polish not-found searches, where 25 missed words among the top 100, and 166 among the top 500, are Swahili. This shows clearly that an option of changing a language does not work well as users forget to switch the languages. Also, the search method should take into consideration all lemmas, Swahili and Polish, during every lookup process.

Out of 500, 144 (28%) strings were identified as lemma lacuna — possible new lemma candidates. Among them, 92% represent lemmas that are also present on the frequency list used to build a dictionary — mostly low-frequency words, like *burudika* 'be appeased' or *manukato* 'perfume'. Only 12 words are not on the frequency list, like *pepa* 'sway', *kauri* 'cowrie shell'. Based on this data, and an analysis of the found entries, we came to a conclusion that this dictionary should be expanded further based on the frequency-list lemmas (Wójciewicz 2017). The list was derived from a Helsinki Corpus of Swahili but only several thousand of the most frequent items were included in the dictionary so far.

100 strings (20%) were identified as orthographic words. These are the forms that the users come across in texts and were looked up in their full form, as for example *anataka* 'she/he wants', *wakicheza* 'as they play'. Among them, 27 are

inflected verbs, 22 — infinitives, like *kuja* 'to come', *kwenda* 'to go', and 51 — adjectives and pronouns with class prefixes, like *mzuri* 'good, CL.1', nouns with a locative suffix *-ni*, as *usoni* 'on the face', or nouns combined with a possessive pronoun, like *mwanangu* 'my child'. All these prefixes and suffixes should be deleted in order to perform the look-up.

There are 42 misspellings, and other unidentified words in this log. Only a few of them are spelling mistakes affected by pronunciation, as when users are spelling the word as it is pronounced, like *mojo* instead of *moyo* 'heart'; by omitting an apostrophe, like *ngombe* instead of *ng'ombe* 'cow'; or inserting a space. There are also some strings that could not be identified as Swahili or Polish words, like *nala*.

During the analysis we have identified only 4 strings out of the top 100 and 25 out of the top 500 as wrongly lemmatised Swahili words. These are the forms that the user probably knew needed to be modified somehow, and tried to lemmatise them but failed, as in *likuwa* (may be from *alikuwa*), *pokuwa* (possibly from *ijapokuwa*), *engi* (probably from *wengi*), *sanya* (part of a verb *kusanya* that looks similar to infinitive but has the morpheme *ku* as a part of a stem).

The remaining 17 strings are multiword expressions, like *baba yetu* 'our father', and 8 strings are proper nouns, like *Timon* and *Kilimanjaro*. It was also interesting to reveal in the Polish not-found file, that users looked up Polish personal names, like *Ania*, *Tomasz*, or names of cities, like *Warszawa*.

A closer examination of the top ten Swahili not-found searches reveals that seven items represent orthographic forms of entries that are present in the dictionary. These are adjectives with a class prefix *mzuri* and *nzuri* 'nice', *njema* 'good', the infinitives *kuja* 'to come', *kwenda* 'to go', *kufa* 'to die' and the inflected verbal form *ninakupenda* 'I love you'. There is also a multiword expression and the name *Mufasa* known from the movie "The Lion King". Only one item, *patia* 'to get for', may be identified as a lemma lacuna.

De Schryver et al. (2006) mention that users often seem to greet the dictionary on arrival. This was also recalled in an interview with Barak Turovsky from Google Translate (Orliński 2017). According to his data the most frequently translated expressions in all languages of the world are 'how are you' and 'I love you'. He finds 'I love you' to be among the top three searches in every language. In our data *kocham cię* or *kocham* 'I love you' in Polish, in its not lemmatised forms, is the first and the third most popular search in the Polish not-found file and the third most popular search in the Swahili not-found. Among the Polish found searches, *miłość* 'love' and *kochać* 'to love', are in the top 25 searches. *Ninakupenda* 'I love you' in Swahili is the eighth most frequently searched expression in the Swahili not-found log, and its lemmatised form *penda* is only among the top 170 in the Swahili found searches. However, since the users of the dictionary are mostly Polish, this explains why they want to express themselves in the Polish–Swahili direction. The users are also interested in greetings, and *cześć* 'hello' is the second most often looked up Polish word.

In their study on reporting look-ups of frequent words in Sepedi and English, De Schryver and Joffe (2004: 190) mention a high number of searches for the offensive and sexual sphere words: "genuine frequent words are looked up on the one hand, and then those words that only mother-tongue speakers know but, as they are taboo, *never* pronounce in public". This was also reported in a study by Bergenholtz and Johnsen (2005). In our data only eight Polish searches among the top 100 are concerned with the sexual sphere — none among Swahili words, as probably users, the learners of the language, are not familiar with these words yet.

One interesting issue to discover was that Polish users quite often, (approximately 10% of the not-found Polish searches), are looking for not lemmatised Polish words, like *lubię cię* 'I like you', *pozdrawiam* 'I greet you'. This also corresponds to the findings of De Schryver et al. (2006) who report on users treating a dictionary like a Web search engine. Our users also put longer strings in the search box, like multiword expressions or even whole sentences, sometimes in other languages. These are, for example, in Swahili *mzuri na wewe?* 'I'm fine and how are you?' with a question mark as a part of several searches, or, in Polish, *jak się masz* 'how are you', or *sto pięćdziesiąt dwa* 'hundred fifty two'.

Of course the question remains how such findings should influence our decisions regarding lemma list and dictionary structure. We may easily add the missing lemmas the users are often looking up and this is what we have already done with this dictionary. We should definitely change the search method by eliminating the need for a language choice. But should we change the lemmatisation strategy based on the analysis? When we compare the number of searches in the Swahili found and Swahili not-found files we see that the majority of not-found searches were carried only once, so they may not be significant.

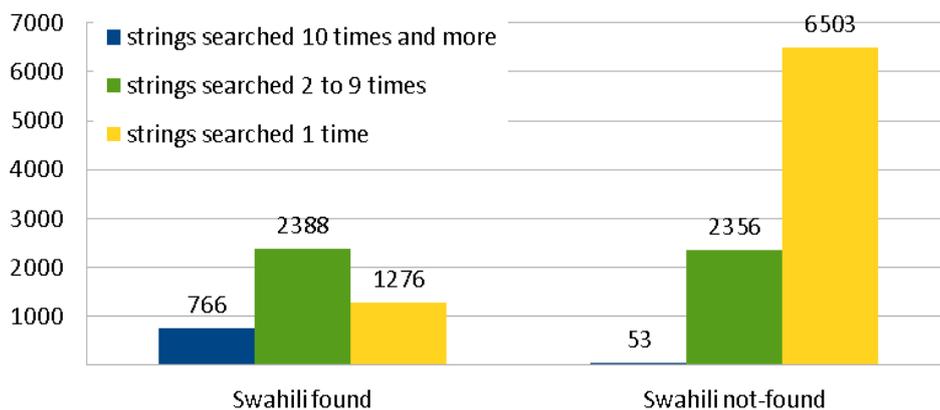


Figure 2: Frequency of searches per string

The top 500 strings in the Swahili found file were carried 11,300 times, while in the Swahili not-found only 3,600 times. The overall number of searches is probably too small to make reliable assumptions. When we analyse the first fifty missed strings that were looked up at least 10 times in the Swahili not-found file, we see that Polish words provide the biggest group, and we find orthographic words next in line in volume. Again, as in the bigger sample, the majority of forms are those with class prefixes attached, the next group are infinitives and then finally we see inflected verbs in only two instances. So based on this data, the next step of expanding the dictionary could be adding adjectives with class prefixes and then infinitival forms of verbs.

When we aim at evaluating the strategy of giving the derivatives a status of main entries, we must stress that the great majority of identified verbal lemma lacuna are verbs with an extension suffix. On the other hand among the top 100 Swahili found searches we have 50 verbs, but only 9 with an extension. When we enlarge the number of analysed strings to 500, among 168 verbs, 49 (30%) have extensions. Therefore, verbal derivatives are not among the most frequently looked up items, but they are searched for quite often, and including them on a headword list proves to be a good, user friendly strategy. Besides frequency, there is also an argument that such a strategy does not leave users in doubt as to the conclusions regarding meaning of derivatives (cf. Gouws and Prinsloo 2005).

As for the nouns, out of over 200 nouns identified among the top 500 most frequently searched items, only 13 are for plural forms. A dozen or so searches are for full forms of pronouns, like *yangu* 'mine', *vile* 'those', *pale* 'there'. But there are only two such searches among the top 100: these are *hivyo* 'that' and *nyingi* 'many'. None of the top searches are for the stem of a pronoun. When we expand the analysis to all of the strings, it appears that there were only a few searches for chosen stems, like *ako* 'your', and *le* 'that'.

Overall, evaluation should take into consideration the number of searches in each file. The majority of the look-ups were successfully retrieved Swahili headwords. The users seem to understand and successfully apply the lemmatisation procedure to find translations of the word forms found in texts. However, they look up full forms of pronouns rather than their stems — these are the forms that are the most difficult to lemmatise, especially the demonstrative pronouns, which are often used and lexicalised. Based on the not-found searches we shall consider expanding the dictionary with full forms of adjectives, which seem to be the most searched for from among orthographic words. The other group represents infinitival forms of verbs. It is possible that users use this form, as this is the lemma for Polish verbs and they mimic their habits from Polish dictionaries. As the last addition we should consider full forms of verbs.

5. Conclusion

Log file analysis helps us to reveal how Internet dictionaries are used. While

monitoring user queries, we may keep track of which lemmas are looked up how often, and investigate the reasons why looking up words goes wrong. In our study of log files from the Swahili–Polish dictionary we revealed some problems caused by the searching method that requires users to choose the language of their search and to lemmatise word forms. As dictionaries of Swahili demand some knowledge of the grammar in order to conduct a successful search, we were interested to uncover how well the users manage their queries.

We have identified a list of lemma lacuna that cause the majority of unsuccessful Swahili searches. The study shows that the users seem to understand the lemmatisation strategy and successfully apply it to the word forms they want to find translations for. They often search for extended verbs and full forms of pronouns, so these need to be treated as headwords. If we were to expand the dictionary based on this data, we should consider adding full forms of adjectives, and infinitival forms of verbs as these two groups, besides lemma lacuna, compose the most missed searches. This came as a surprise to us, as we assumed that the most difficult operation is the decomposition of the verbal complex, when the user has to cut off all prefixal morphemes and search for the verbal root or extended root instead. Searches for full forms of verbs also caused some unsuccessful searches, but they were not as many as we expected.

Endnotes

1. <http://kamusi.pl/> [accessed 20.04.2017]
2. <http://africanlanguages.com/kiswahili/> [accessed 20.04.2017]

References

A. Dictionaries and corpora

- Abdulla, A., R. Halme, L. Harjula and M. Pesari-Pajunen (Eds.).** 2002. *Swahili–Suomi–Swahili-sanakirja*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Baba Malaika.** 1994. *Modern Swahili Modern English*. Arusha: Danish Training Center for Development Co-operation.
- (HCS) Helsinki Corpus of Swahili.** 2004. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC — Scientific Computing Ltd.
- Johnson, F. (Ed.).** 1985 [1939]. *A Standard Swahili–English Dictionary (founded on Madan’s Swahili–English Dictionary)*. Oxford: Oxford University Press.
- Madan, A.C.** 1992 [1903]. *Swahili–English Dictionary*. New Delhi: Asian Educational Services.
- Sacleux, Ch. (Ed.).** 1939. *Dictionnaire Swahili–Français*. Paris: Institut d’Ethnologie.
- (TUKI) Taasisi ya Uchunguzi wa Kiswahili.** 2001. *Kamusi ya Kiswahili–Kiingereza. Swahili–English Dictionary*. Dar es Salaam: Chuo Kikuu cha Dar es Salaam.
- Wójtowicz, B.** 2013. *Słownik suahili–polski*. [Swahili–Polish Dictionary]. Warszawa: Elipsa.

B. Other literature

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- 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.
- Bergenholtz, H. and M. Johnsen. 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes, Journal of Linguistics* 34: 117-141.
- Bosch, S.E., L. Pretorius and J. Jones. 2007. Towards Machine-Readable Lexicons for South African Bantu Languages. *Nordic Journal of African Studies* 16(2): 131-145.
- De Schryver, G.-M. and D.J. Prinsloo. 2001. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of *Frequency-based Tail Slots* — With Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001. *Makala ya kongamano la kimataifa Kiswahili 2000: Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.
- De Schryver, G.-M. and D. Joffe. 2004. On How Electronic Dictionaries are Really Used. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 187-196. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- 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.
- Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Herns, I. 1999. Swahili Lexicography: The Swahili-German Dictionary. Wolff, H.E. (Ed.). 1999. *Contributions to Bantu Lexicography. Languages and Literatures* 10: 1-8. Leipzig: University of Leipzig Papers on Africa.
- 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.
- Kiango, J.G. 2005. Problems of Citation Forms in Dictionaries of Bantu Languages. *Nordic Journal of African Studies* 14(3): 255-273.
- Knowles, G. and Z. Mohd Don. 2004. The Notion of a "Lemma". Headwords, Roots and Lexical Sets. *International Journal of Corpus Linguistics* 9(1): 69-81.
- Koplenig, A., P. Meyer and C. Müller-Spitzer. 2014. Dictionary Users Do Look up Frequent Words. A Log File Analysis. Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*: 229-249. Berlin: Walter de Gruyter. (Lexicographica Series Maior 145.)
- Kosch, I. 2013. Expectation Levels in Dictionary Consultation and Compilation. *Lexikos* 23: 201-208.
- Laufer, B. and M. Hill. 2000. What Lexical Information Do L2 Learners Select in a Call Dictionary and How Does It Affect Word Retention? *Language Learning & Technology* 3(2): 58-76.
- Lemnitzer, L. 2001. Das Internet als Medium für die Wörterbuchbenutzungsforschung. Lemberg, I., B. Schröder and A. Storrer (Eds.). 2001. *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*: 247-254. Tübingen: Max Niemeyer Verlag.

- Lew, R.** 2011a. User Studies: Opportunities and Limitations. Akasu, K. and S. Uchida. (Eds). 2011. *ASIALEX2011 Proceedings. Lexicography: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto, Japan, August 22–24, 2011*: 7-16. Kyoto: Asian Association for Lexicography.
- Lew, R.** 2011b. Studies in Dictionary Use: Recent Developments. *International Journal of Lexicography* 24(1): 1-4.
- Lew, R. and G.-M. de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
- Müller-Spitzer, C., S. Wolfer and A. Koplenig.** 2015. Observing Online Dictionary Users: Studies Using Wiktionary Log Files. *International Journal of Lexicography* 28(1): 1-26.
- Nesi, H.** 2000. On Screen or in Print? Students' Use of a Learner's Dictionary on CD-ROM and in Book Form. Howarth, P. and R. Herington (Eds.). 2000. *EAP Learning Technologies*: 106-114. Leeds: Leeds University Press.
- Orliński, W.** 2017, March 17. Sztuczna Inteligencja jest Glupsza od Trzylatka. *Gazeta Wyborcza*. <http://wyborcza.pl/magazyn/7,124059,21512640,sztuczna-inteligencja-jest-glupsza-od-trzylatka-orlinski.html?disableRedirects=true>. [Accessed 27.04.2017.]
- Prinsloo, D.J. and G.-M. de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Töpel, A.** 2014. Review of Research into the Use of Electronic Dictionaries. Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*: 13-54. Berlin/New York: De Gruyter.
- Trap-Jensen, L., H. Lorentzen and N.H. Sørensen.** 2014. An Odd Couple — Corpus Frequency and Look-up Frequency: What Relationship? *Slovenščina* 2.0, 2(2): 94-113. <https://dsl.dk/medarbejdere/medarbejdere-publikationer-m-m/ltj/an-odd-couple>. [Accessed 05.04.2017.]
- Verlinde, S. and J. Binon.** 2010. Monitoring Dictionary Use in the Electronic Age. Dykstra, A. and T. Schoonheim (Eds). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010*: 1144-1151. Ljouwert: Fryske Akademy — Afûk.
- Wójtowicz, B.** 2017. Evaluating a 12-Million-Word Corpus as a Source of Dictionary Data. *International Journal of Lexicography* 2017: 1-15. doi: 10.1093/ijl/ecx011.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia.

Representing the Cultural Dimension of Meaning in Learner's Dictionaries — From the Perspective of Chinese EFL Learners in L2 Reception

Mei Xue, *School of Law & Humanities, China University of Mining and Technology (Beijing), Beijing, P.R. China (meix99@yahoo.com)*

Abstract: Meaning in dictionaries ranks as the most frequent data type consulted by language learners. An extensive range of devices have been developed to represent the meaning of lexical items in learner's dictionaries. Starting from interpreting foreign language learners' lexicographic needs in specific user situations, this paper aims to examine the implications of customizing definitions from the cultural perspective with specific reference to Chinese learners of English. The purpose of this paper is three-fold: i) to understand the status quo of the lexicographic practice of English learner's dictionaries in terms of representing the cultural dimension of a lexical item; ii) to interpret the impacts of foreign language learners' language and cultural background on their lexicographic needs in terms of cultural aspects in L2 reception; iii) to address the identified problems with present lexicographic practices through model articles. It is expected that the findings in this paper can shed some light on interpreting other ethnic groups of learners' lexicographic needs in different situations and contribute to improving the services of learner's dictionaries in the future as well.

Keywords: CHINESE EFL LEARNERS, CULTURAL DIMENSION, L2 RECEPTION, LEARNER'S DICTIONARIES

Opsomming: Die weergee van die kulturele dimensie van betekenis in aanleerderswoordeboeke — uit die oogpunt van Chinese EVT-leerders in L2-resepsie. Betekenis in woordeboeke word beskou as die mees frekwente datasoort wat deur taal-aanleerders opgesoek word. 'n Omvattende reeks hulpmiddels is ontwikkel om die betekenis van leksikale items in aanleerderswoordeboeke weer te gee. Hierdie artikel begin by die interpretasie van vreemdetaalleerders se leksikografiese behoeftes in spesifieke gebruikersituasies en beoog om die implikasies van die pasmaking van definisies uit 'n kulturele perspektief te ondersoek met spesifieke verwysing na Chinese aanleerders van Engels. Die doel van hierdie artikel is drievoudig: i) om die status quo van die leksikografiese praktyk van Engelse aanleerderswoordeboeke te verstaan ten opsigte van die weergee van die kulturele dimensie van 'n leksikale item; ii) om die uitwerking van vreemdetaalleerders se taalkundige en kulturele agtergrond op hul leksikografiese behoeftes te interpreteer ten opsigte van kulturele aspekte by L2-resepsie; iii) om die geïdentifiseerde probleme met die huidige leksikografiese praktyke deur modelartikels aan te spreek. Daar

word vermag dat die bevindinge in hierdie artikel lig sou kon werp op die interpretasie van aanleerders uit ander etniese groepe se leksikografiese behoeftes in verskillende situasies en ook 'n bydrae sou kon lewer tot die verbetering van die bruikbaarheid van toekomstige aanleerderswoordeboeke.

Sleutelwoorde: CHINESE EVT-LEERders, KULTURELE DIMENSIE, L2-RESEPSIE, AANLEERdersWOORDEBOEKE

1. Introduction

Meaning in dictionaries ranks as the most frequent data type consulted by language learners. Numerous empirical studies on meaning determination resort to a specific learning task, namely text reception, production, and vocabulary learning to examine the effectiveness and efficiency of the dictionaries in different formats. Empirical findings generally show that it is still very challenging for foreign language learners to select the correct sense for a given context (Chan 2012: 135, 2014: 42; Ding 2015: 25-26; Nesi and Tan 2011: 90-91; Ptasznik and Lew 2014: 251). Undoubtedly, these empirical studies shed some light on understanding the problems with the present lexicographic practices, which is certainly helpful for devising models for future lexicographic practice. It is of paramount significance to make theoretic explorations on the critical areas like "the data to be included in these dictionaries, the structures to present and accommodate the data, the functions of these dictionaries and the way they should respond to the needs of their target users" (Gouws 2014: 157).

The emergence of e-lexicography has been refreshing the concept of dictionaries as well as the presentation of various data types in dictionaries (Tarp 2011). The digital medium offers golden opportunities to approximate the longstanding issue of customization in lexicographic practice (Verlinde and Peeters 2012: 148). However, the "current e-dictionaries too often do not live up to the expectations of their users", as "too many lexicographic e-products were developed without any influence from innovative theoretic suggestions" (Gouws 2014: 156-157). Quite a number of empirical studies show that the electronic dictionaries are not necessarily advantageous over their paper counterparts in terms of the provision of meaning to serve different learners (Chen 2010; Dzie-mianko 2012). Hence, it is a propitious moment to identify the target users' needs for meaning information in a specific user situation when discussing the provision of the customized services for users. Attention should also be paid to the profile of the target users, as the "users in general never need information in general" (Tarp 2009: 46).

This paper explores the implications of customizing definitions from the cultural perspective with specific reference to Chinese learners of English. Given the distance in L1 (Chinese) and L2 (English), Chinese learners may have some specific information needs for meaning in cultural aspects when they are engaged in different English learning activities. Starting from interpreting Chi-

nese learners' particular needs for cultural information during L2 reception, this paper examines the status quo of the present lexicographic practices with English learner's dictionaries. Proposals are made for a more salient representation of the cultural dimension of the lexical items from Chinese learners' perspective. It is expected that the findings in this paper can shed some light on interpreting other ethnic groups of learners' lexicographic needs in different situations and contribute to the improved lexicographic practices in the future.

2. Lexicographic Function and the Representation of Meaning in Learner's Dictionaries

Dictionaries above all are utility tools conceived and produced to solve the practical problems arising in concrete situations (Tarp 2008), as the "lexicographic needs are not abstract needs, but are always related to the specific type of users who find themselves in a specific type of social situation" (Tarp 2009: 25).

2.1 The Function of a Learner's Dictionary

Different types of users may have different lexicographic needs, and the same users may have different needs in different situations. In this respect, the Function Theory of Lexicography distinguishes between four fundamental extrallexicographical situations: communicative, cognitive, operative and interpretive situations (Tarp 2008, 2011). In the process of foreign language learning, foreign learners are frequently situated in communicative and cognitive situations. The communicative situations can be further specified into text reception in L2, text production in L2, translation from L1 to L2 or vice versa. There are fundamental differences between the communicative and cognitive situations in terms of foreign learners' information needs. In cognitive situations, "it is very difficult to obtain a clear idea of the exact information that the users need, and as a result, of the data required to provide this information" (Tarp 2011: 65), as learners' objective is to gain knowledge. On the other hand, in communicative situations, learners' needs are concrete. For example, learners in reception situations mainly consult dictionaries for meaning information to facilitate comprehension, whereas in production they may need grammatic and pragmatic information to help them properly use a lexical item.

The genuine function of a learner's dictionary is to satisfy foreign learners' needs in a specific language learning situation, but it is to be noted that there are always unpredictable information needs in additional situations, although such cases are relatively rare (Tarp 2011). In functional terms, the optimal dictionary is mono-functional (Bergenholtz 2011), but "this ideal is hardly ever commercially viable in printed dictionaries; it can be approximated in electronic dictionaries" (Heid 2014: 48). The electronic medium makes it possible to

set up a repository of data for various lexicographic purposes (Heid 2014; Granger and Lefer 2016). The presentation interface of data can cater for the individual learners' needs in different situations. But, the electronic dictionaries may produce an "abundance of unstructured data" leading to the "information overload" and the ultimate "information death" (Gouws and Tarp forthcoming). Hence, efforts have to be taken to minimize the search-related costs (within a decent time) and comprehension-related costs (with low cognitive efforts) (Bergenholtz 2011; Tarp 2011). In other words, to warrant a successful information retrieval during dictionary consultation, the data on the interface of the article should be relevant and comprehensible, and the presentation of the required data should be explicit.

2.2 The Representation of Meaning in Learner's Dictionaries

Meaning is a central notion in lexicography and has been attracting wide discussion in the past thirty years. The digital medium has been extending the repertoire of defining devices available in printed age and may present a multimodal solution to definitions. But, without a coherent framework to guide how to employ the diverse defining devices to meet the target users' needs in concrete situations, the potential offered by technology can be of decorative value as shown in Lew and Doroszevska's (2009) study on animation. Foreign learners' difficulty in sense selection is still apparent as evidenced in numerous studies mentioned above, despite the availability of various defining devices in electronic dictionaries.

Referring to a multitude of devices for definitions, Liu (2015) proposes a concept of multimodal definitions used for e-lexicography. According to Liu (2015), the linguistic function of meaning (i.e. ideational, interpersonal and textual) can be realized by technology in four dimensions (i.e. content, form, space and time). The multimodal definitions are mainly based on functional linguistics. But, it is to be noted that the notion of senses is different in lexicography and linguistics and there is a tenuous relationship between the lexicographic sense and linguistic notion (Lew 2013: 284-285). Functional linguistics is concerned about what language does and how it does so in a given context, whereas various linguistic data in a dictionary are combined to accomplish the function of the dictionary instead of the linguistic function in social contexts. Linguistic developments definitely contribute to the development of lexicographic tools, and the electronic medium may speed up the interactions between linguistics and lexicography. However, both the linguistic developments and technology potentials have to be interpreted from the lexicographic perspective and critically adapted for lexicographic practice. Hence, it tends to be arbitrary to claim that:

E-dictionaries can renew its information more frequently and easily than their print counterparts. This can help guarantee the recency and accuracy of meaning which prove to be the major concerns of users (Liu 2015: 255).

The electronic dictionaries enjoy the advantages of updating the included data regularly, and the renewed data may satisfy some concerns of some users in some situations, but not necessarily the major concerns of users in different situations.

The electronic medium seems to put an end to the struggle over the trade-off between the clarity and precision in the provision of definitions in paper dictionaries. But, there is a need to "find the balance between telling the fullest story and deciding what's useful for the average reader" (Fatsis 2015). In digital age, lexicographers have to decide what information is relevant to the target users' needs and how to present the relevant information explicitly. *MD (Macmillan Dictionary) Online* introduces a two-sentence policy to convey the conventional meaning and the connotative or pragmatic/cultural information in definitions. A similar approach is adopted by *CED (Collins English Dictionary) Online*. The addition of the information about the attitude or a motivation of a speaker to use a word is especially useful for foreign learners in text production. This has been explicitly stated by Stein (2002):

The lexicographic description of some words requires an explicit and precise description of the pragmatic situation in which they are used or in which the activity denoted by them is performed (95).

Pragmatic restrictions are useful to guide foreign learners' proper use of a L2 lexical item. On the other hand, in L2 reception contexts, they may encounter some objects or concepts typical of L2 culture and hence need the relevant cultural elaborations to facilitate comprehension. Learners' needs with regard to cultural aspects during L2 reception are authentic and deserve attention in lexicography. However, there is relatively little discussion in this respect. Therefore, this paper aims to examine the provision of cultural data, with specific reference to Chinese learners' needs arising in L2 reception. The purpose of this paper is three-fold: i) to understand the present lexicographic practice of English learner's dictionaries in terms of representing the cultural dimension of the lexical items; ii) to interpret the impact of the learners' language and cultural backgrounds on their needs with regard to cultural aspects during L2 reception; iii) to address the identified problems with proposals. It is to be noted that the online dictionaries referenced in this paper is for research purposes, not for an assessment of their popularity or quality.

3. The Gap between Users' Needs and the Present Lexicographic Practice

Words and expressions in a language are tinted with cultural dimensions, but there are degrees of culture-specificity. Some items are more culture-bound than others. For instance, the expression *barefoot doctor* is culture-loaded, describing a social phenomenon in China in a specific historical period. In contrast, the referent of a *magpie* is similar in different countries, but its associated symbolic meaning varies with different cultures. The loss of the track of the

cultural allusion underlying a lexical item could lead to communication failures in cross-cultural contexts. Zgusta (1989) has stated:

[...] since language is embedded in culture, cultural data are important to the learner not only for steering his linguistic behaviour but frequently for choosing the correct lexical equivalent. Such cultural information can be understood in a broad way, Undoubtedly a good part of this information is of encyclopaedic character; be this as it may, it belongs to what the learner has to learn (3).

Culture can be universal as well as specific. People around the world share a common concept of *New Year*, but the specifications of *New Year* vary with different cultures in terms of its history, significance and rituals. The overlapping and discrepancies in cultural schema in L1 and L2 may facilitate or block foreign language learning. For example, the average German learners of English could have much more in common with English culture than the average Chinese learners. Hence, Chinese learners could have more needs for information typical of English culture than their German counterparts have in the process of learning English. Therefore, the inclusion and the presentation of the cultural data in a learner's dictionary should be considered with reference to the target users' language and cultural backgrounds.

3.1 Chinese Learners' Difficulty in Comprehending L2 Culture-loaded Words

Meaning data in learner's dictionaries help dictionary users smooth away the blocks to comprehension in L2 reception. But, when there is a wide discrepancy in the conception of an object or an idea in L1 and L2, a factual definition of the referent in L2 may be insufficient for solving the possible receptive problems. In such cases, L1 equivalents alone may not work effectively either. For instance, when foreign learners of English encounter the word *robin* and may have different information needs depending on their prior knowledge, reading skills, and the immediate reading context: i) the word *robin* is completely a new word, and they want to know what it refers to; ii) they may infer that the word *robin* refers to a kind of bird, but are curious to confirm their assumptions or prior knowledge about this type of bird; iii) they may have the inferences about the bird, but the given context demands the suggestive meaning and they cannot figure it out. To sum up, foreign learners of English, like Chinese learners may need the information about the denotative and symbolic meaning of a lexical item in L2 reception. When foreign learners are living in the cultural settings of the addressed language, they may also need to identify the referents for their daily life.

The definition "*a small brown bird with a red chest*" in *CLD (Cambridge Learner's Dictionary) Online* is sufficient for solving users' problems in the first situation above. On the other hand, the provision of a lexical remark in italics like "*robins mostly appear in the winter and are commonly pictured on Christmas*"

cards" in *CD (Cambridge Dictionary) Online* (see Figure 1) seems to fulfill Chinese learners' information needs in the third context above. However, this lexical remark still demands learners' efforts to abstract the implied message that a *robin* is a symbol of Christmas in the western culture. A click on the translation button in both *CLD Online* and *CD Online* will direct learners to the interface of the corresponding bilingual article in which four Chinese equivalents to *robin* are provided. However, if Chinese learners have limited zoology knowledge about the bird *robin*, these equivalent terminologies are no better than a short English explanation.



Figure 1: The article *robin* from *CD Online*

In contrast, the cultural note in *OALD (Oxford Advanced Learner's Dictionary) Online* as shown below is quite explicit and spares learners' efforts to retrieve the overtones carried by *robin*. *MD Online* goes a step further by supplementing a pictorial illustration to help learners identify the referent in the physical world, which contributes to the cognitive learning of this word as well.

robin ...

-Culture

In Britain the *robin*, sometimes called *robin redbreast*, is strongly associated with Christmas and pictures of robins, especially in snow-covered gardens, are often used on Christmas cards.

Unfortunately, such a way of representing the cultural dimensions of a word is not a common practice in *MD Online*. The criteria for including such cultural elements are implicit and intangible. For example, the word *chard* is illustrated with a relevant picture, whereas no pictorial illustrations are used to explain *artichoke*, *kale*, *Romanesco*, and *arugula*, etc. It is strange that different defining policies are employed to describe these vegetables, given the fact that all these varieties of vegetable are foreign to people whose cultural backgrounds are distanced from European culture. For instance, the definition of *artichoke* "a round green vegetable with thick pointed leaves arranged close together in layers" from *MD Online* almost means nothing to most Chinese learners, as they do not share the similar schema knowledge about this vegetable common to Europeans. Inevi-

tably, the pragmatic information in the second sentence of the definition "*After cooking it, you pull off each leaf and eat the soft bottom part*" may not get across to Chinese learners at all. A picture can be much more effective than a thousand words, when it comes to describing something typical of one culture but foreign to other cultures. Lexicographers should take dictionary users' cultural background into account upon defining the L2 culture-loaded lexical items.

The distance between English and Chinese leads to a wide divergence of the reflective and social meaning of some words, like the words for colors, animals and numbers, etc. A magpie is "*a bird with black and white feathers and a long tail*" (from *CLD Online*) in physical terms, but its symbolic meaning is divergent in British and Chinese culture. In British culture, the magpie is thought to be in connection with bad luck, especially when you only see a magpie alone, whereas the symbolic meaning of a magpie takes on a bright note in the East. In China, the magpie is a symbol of good luck, joy, marital bliss and long lasting fortune. The different conception of a magpie in Chinese as opposed to English culture may elicit confusions in understanding if the given reading context calls for the cultural associations of this word. A learner's dictionary does not only need to pay attention to the lexical items with referents typical of L2 culture, but also those having different referential or connotative meanings in other language and cultures. It might be unrealistic to include various encyclopaedic information in a general English learner's dictionary, but it is feasible to let the dictionary raise users' awareness of divergent interpretations of the concerned lexical items in different cultures. However, none of the learner's dictionaries under examination seriously addresses the issues in this respect.

It goes without saying that the inclusion of cultural data is necessary for L2 reception. Equally important, the content of the cultural data should be relevant and revealing from the perspective of the target users. Take the cultural note on *mockingbird* in *OALD Online* as an example:

mockingbird ...

-Culture

Many people believe it is morally wrong to kill one, and this idea was used in the novel *To Kill a Mockingbird*. It is also mentioned in several traditional songs and is the official state bird of Arkansas, Florida, Mississippi, Tennessee and Texas.

The starting point of this cultural note is to explain the allusion to *mockingbird*, but the content of the culture note may confound a foreign learner who has never read the book *To Kill a Mockingbird*. Such a circumlocutory explanation leaves dictionary users wondering why it is morally wrong to kill a *mockingbird*. Such a practice also occurs in *OD (Oxford Dictionaries) Online* and *CD Online*. For example, the article *mistletoe* states that it is traditional to kiss under the *mistletoe*. Given their limited knowledge about L2 culture, it is highly likely that average users insist on what this tradition means to people kissing under

the mistletoe. Therefore, it is recommendable to make the cultural allusions apparent and revealing to foreign learners. If the explanations tend to be too complicated for such a short note, an external hyperlink to a further elucidation of the concerned topic can be provided as an option.

Suppose Chinese learners turn to bilingual solutions when they encounter the above-mentioned words. The search results from the most popular English–Chinese online dictionaries like *Haici*, *Jinshan Ciba* and *Youdao* are rather depressing. The general practice of these online bilingual dictionaries is to provide one or several equivalents to *robin* (驹鸟, 知更鸟, 鸫), *artichoke* (洋蓟), *curly kale* (羽衣甘蓝), *arugula* (芝麻菜), *Romanesco* (罗马花椰菜), *mockingbird* (嘲鸫) and *mistletoe* (槲寄生). The Chinese equivalents are nomenclatures from zoology and botany. These terms are quite abstract and foreign to average Chinese people and do not help reception at all, not to say become part of their knowledge. The encyclopaedic notes provided at the end of each article in *Youdao* tend to be incomplete and unreliable. Many of the illustrative examples in *Jinshan Ciba* are abstracted from unknown sources and thrive on various errors. The problems with the online English–Chinese dictionaries have been identified in earlier studies (Chen 2010; Ding 2008) and unfortunately have not been attended to much yet. In short, the most popular online English–Chinese dictionaries in China mainly focus on providing lexical equivalents and translation of example sentences, leaving the cultural dimension of English lexical items almost untouched.

3.2 Chinese Learners' Difficulty in Comprehending Chinese Loanwords

Chinese loanwords in English may also pose challenges to Chinese learners during text reception. Due to the varied source languages and different transcribing systems, Chinese learners may experience difficulty in identifying these English words of Chinese origin, for example *bok choy*, *cheongsam*, *chow mein*, *chop suey*, *choy sum*, *fan tan*, *dim sum*, *ho-ho*, *kowtow* and *pekoe*, etc. The word *Chinese* is frequently used as a synonym to Mandarin, but linguistically, Chinese is an umbrella term encompassing a number of regional varieties called *fang yan* (dialects) in China, each with its own sub-varieties: Mandarin, Wu, Gan, Xiang, Hakka, Yue, and Min. Mandarin Chinese/Putonghua. Cantonese and Amoy are the main source languages for Chinese words entering English. It is to be noted that Chinese dialects are not all mutually intelligible. This means that most of the established Chinese loanwords of Cantonese origin maybe foreign to the majority of Chinese speakers in China (Yang 2009: 102).

Xia and Zhai (2016: 13-14) identify four problems with the definitions of Chinese loanwords included in *OALD8* (*Oxford Advanced Learner's Dictionary in its 8th edition*): the lack of indicating regional and referential differences, the missing denotative meaning and the definitions not being included. The referential differences are mainly caused by the inroad of Chinese loanwords into English. For example, *chop suey* is of Cantonese origin, referring to "a Chinese-

style dish of small pieces of meat fried with vegetables and served with rice" (in *OALD Online*), whereas its Chinese equivalent *zá sui* (杂碎) in Mandarin (from bilingual dictionaries like *Haici*, *Jinshan Ciba* and *Youdao*) refers to a dish of cooked entrails of cattle or sheep. The word *dim sum* is also of Cantonese origin specifically referring to Cantonese snacks or pastries in English, whereas its Chinese equivalent *diǎn xīn* (点心) in Mandarin Chinese covers various snacks from different parts of China. As illustrated, the referential meaning of some Chinese loanwords are greatly restricted to their source language. Therefore, it is preferable for lexicographers to mark the regional restrictions in the definition section; otherwise, it may cause conceptual confusion in different receptive contexts. In English–Chinese dictionaries, it is also necessary to supplement the regional restrictions in definitions or explanatory notes to raise Chinese learners' awareness of referential differences in different cultures.

The close examination of the definitions of *kowtow* in *OALD Online*, *MD Online*, *CLD Online* and *CD Online* shows that the objective description of the traditional Chinese custom is omitted, which fails to fully represent the conception embodied in the word *kowtow*. Figure 2 below shows that only the metaphorical meaning of *kowtow* is presented in *CED Online* for learners of English.



Figure 2: Part of the article *kowtow* from *CED Online*

Kowtowing is a solemn rite paying respects to the senior or the superior in ancient China and is still performed in many significant social situations (i.e. the Chinese Spring Festival, Qingming Festival, and wedding ceremony) today. The unanimous omission of the literal explanation of the kowtowing custom in *MD Online*, *OALD Online* and *CLD Online* is inexplicable. *CED Online* designed for general purpose does offer an explanation of "to touch the forehead to the ground as a sign of deference". However, this definition does not explain the cultural contexts in which the ritual takes place. The omission of the literal meaning of *kowtow* and the lack of contextual information on the ritual of kowtowing may not influence Chinese learners or the learners sharing the same cultural schema, but it would impose the misconception of this custom on other ethnic groups of learners in receptive and cognitive situations.

3.3 The Problems with the Present Lexicographic Practice

The detailed analysis in this section shows that the online dictionaries do not actively respond to foreign learners' needs in cultural dimensions. Specifically, the following aspects deserve attention. First, the inclusion of cultural data tends to be incidental and unpredictable. Second, the content of the cultural data is not explicit and demands many efforts in the process of information retrieval. Third, cultural differences in denotative or connotative dimensions should be attended to, to elicit foreign learner's awareness in L2 reception. Fourth, the conceptual meaning of Chinese loanwords in Chinese and English deserve due attention. Last, but not the least important, bilingual solutions should be clearly defined in foreign language learning contexts.

4. Proposals

This section demonstrates some proposals through model articles for future lexicographic practice, highlighting EFL learners' needs with regard to cultural aspects during L2 reception. Chinese learners of English are taken as target dictionary users to illustrate customization of lexicographic assistance demonstrated in some model articles. The proposals are mainly targeted at online dictionaries and can be adapted for dictionaries in other forms.

4.1 Presenting the Relevant Cultural Data Explicitly

When the conception represented by a L2 lexical item does not belong to the cultural universe of foreign learners', it is necessary to make such part of cultural data apparent to them and help them into the cultural dimension associated with the word for better understanding. Take the word *mockingbird* as an example:

mockingbird a small white and gray North American bird that copies the songs of other birds.

-Culture: *mockingbird* is able to mimic almost any sound it hears, symbolizing joyfulness, cleverness, playfulness, and communication. In the book "*To Kill a Mockingbird*", *mockingbird* symbolizes innocence. It is morally wrong to kill a mockingbird, as it is neither harmful nor destructive, but only makes music for people to enjoy.



In the above model article, the definition, pictorial illustration and culture note are combined to enhance comprehension. The pictorial illustration shows the referent and help learners identify this bird in reality, which also contributes to cognitive learning. The culture note explicitly states the universal symbolic meaning of *mockingbird* as well as its metaphorical significance in the well-known book in association. The culture note in the model article is collapsed below the definition with a guide word **Culture**. All the information can be accessed through a click. For learners who rely on their mother tongue for text reception, the L1 equivalents together with the above-mentioned data should be supplemented for them to identify the referent for cognition purposes and future communication in L1, as the provision of only a L1 equivalent is very challenging for learners with limited knowledge about birds.

In the model article *mistletoe*, the cultural dimension of *mistletoe* finds its expression in definitions, pictorial illustrations and the culture note. Pictorial illustrations are not immediately presented on the interface out of concern that learners may be either interested in the plant on the tree or the plant as a Christmas decoration. Learners can access the picture of mistletoes growing on trees or being made as Christmas decorations by clicking the link to pictures. The culture note falls into two parts: the legends associated with *mistletoe* and the explanatory note about the tradition of kissing under the *mistletoe*. The metaphorical meaning of *kissing under the mistletoe* will be explicitly stated to resolve the possible questions about this tradition.

mistletoe a parasitic plant that grows on the branches or trunks of other trees and is often used as decorations at Christmas.

+For Pictures

- Culture

+ Legends

+ kissing under the mistletoe

The model article *artichoke* below is showing the possible treatment of the words referring to the objects or ideas common to L2 culture, but foreign to other cultures. A pictorial illustration of *artichoke* in combination with the pragmatic information in the definition clearly indicates what kind of vegetable *artichoke* is and how it can be cooked in general. The detailed cultural information about this vegetable is folded in the cache with a symbol icon "+" when needed. The second edition of *Longman Dictionary Of Contemporary English* has long ago stressed that "[t]he illustrations and the definitions work together to ensure that important language points and aspects of Western culture are explained in the clearest possible way" (F49 as cited in Stein 1991: 106). For those objects having varieties, like *kale*, the pictorial illustrations of the common curly *kale* can be presented for general understanding. In an online English-Chinese dictionary, it is also necessary to supplement the Chinese equivalents like 洋蓟, 朝鲜蓟, 菊芋 with pictorial illustrations. By knowing the name of the

vegetable in their mother language, foreign learners could feel much confident in future communication on this topic.

artichoke: boil this vegetable in salted water or steam it, then pull off each leaf and eat the soft bottom part.

+More



4.2 Presenting the Information about Cultural Differences

To deal with the cultural dimensions of lexical items, Bergenholtz and Agerbo (2014: 498) propose that the meaning elements be mainly distributed in three fields, namely a main meaning field, a lexical remark and a synonym remark. There is a risk to have some information overlapped in the three fields as shown in their given example of *swastika* (see Bergenholtz and Agerbo 2014: 500). The given example has addressed the cultural implications of *swastika* in German and India, but omitted its significance in Chinese culture. This present paper proposes that the definition part focus on the general meaning of the lexical item, whereas the lexical remark part elaborates the cultural differences. Moreover, the function of the dictionary plays a role in determining the contents of the lexical remarks. For text reception purposes, the lexical remark focuses on describing the diversified dimensions of the word in different cultures. In contrast, for production purposes, lexicographers need to go beyond the mere factual description of the cultural significance embodied in the lexical items and give explicit suggestions to guide learners to use the words properly or behave themselves appropriately in social contexts.

In the case of *magpie*, a contrastive presentation of the cultural allusions in Chinese and British culture would be favorable for L2 reception and cognition. For example:

magpie ...

-Culture: In British culture, magpies are believed in different numbers to foretell death, a funeral, and bad luck. The popular saying about magpies: One for sorrow. Two for mirth. Three for a wedding. Four for death or birth, depending on the origin. In Chinese culture, magpies symbolize good luck, joy, marital bliss and long lasting fortune.

The content of the culture note in the model article is intended to raise the users' awareness of conceptual differences in British and Chinese culture and minimise possible confusion during L2 text reception due to their own cultural backgrounds.

4.3 Addressing the Problems with Chinese Loanwords in English

Depending on the problems Chinese learners have with loanwords during reception, the lexicographic assistance can be diverse. First, when the comprehension difficulty is caused by the old transcribing systems, a pinyin-based equivalent can be a solution. For example, *ti-tzu* is *dí zi* (笛子); *kylin* is *qí lín* (麒麟). As these objects and concepts are typical of Chinese culture, it is recommended that illustrations and explanations should be provided for learners with different cultural backgrounds. If the word has symbolic meaning or is associated with traditional stereotypes in Chinese culture, short explanatory notes are needed. Second, when the referential difference is caused by source languages of Chinese loanwords, an explanatory note or a reminder is necessary to clarify the divergences. Therefore, the Chinese equivalent to *dim sum* is not *diǎn xīn* (点心), but *yuè shì diǎn xīn* (粤式点心). The additional regional restriction *yuè shì* (粤式) demonstrates that *dim sum* is of Cantonese origin. This practice has been adopted in many English–Chinese online dictionaries. Third, when the same objects represented by the Chinese loanwords take different names in different geographic areas in China, the L2–L1 solution should take into account the variant nomenclatures. For instance, *bok choy* in the northern parts of China is called *yóu cài* (油菜), whereas it is called *xiǎo bái cài* (小白菜) in the southern part of China. And *yóu cài* (油菜) in South China normally refers to the rapeseed plant. Last, the cultural dimensions embodied in Chinese loanwords should be interpreted in Chinese social contexts, as illustrated in the case of *kowtow*.

Given the increasing contact between Mandarin Chinese and English, spurred by the fast economic development in China, many pinyin-based words and expressions are used in the western media and some of them have entered English dictionaries, like *fengshui*, *laogai*, *guanxi*, etc. There are trends of replacing the established Chinese loanwords with pinyin-based words. For example, *dumpling* is *jiaozi*, *ang pow* is *red packet*, and *kung fu* is *wushu*. The emblematic traditional dress worn by Chinese women is called *cheongsam* (of Cantonese origin) or *qipao* (of Mandarin origin). Xia and Zhai's (2016) criticism of the mix-up of conceptual differences with regard to *cheongsam* and *qipao* in *OALD8* is partially true, given the historical usage of *cheongsam*. But, the current usage of *cheongsam* has conformed to the conception of *qipao* as shown in the mainstream media in China and the western world, referring to the symbolic traditional dress worn by Chinese women (Yang 2009). A search in Ngram Viewer indicates that the use of *qipao* has been rising steadily and quickly since 1980s and is approximating the usage frequency of *cheongsam*.

Given the above understanding, it is preferable to explicitly state the referential meaning of *cheongsam* in contemporary and historical contexts to help dictionary users cope with possible receptive situations.

The model article below shows a bilingual solution to reception problems, presenting the current usage of *cheongsam* and its original meaning respectively. The first part includes the variant expression, L1 equivalent, definition in L2 and the culture notes. The culture notes fall into two parts: the pictorial illustrations of different types of *cheongsam/qipao* and the pragmatic information about how to dress up properly with regard to *cheongsam/qipao*. The illustrations of various *qipao* are used to help learners identify the referents and raise their awareness that *qipao* comes in diverse styles. And dressing-up constitutes another dimension of the concept of *qipao* in Chinese culture, which is much more relevant to cognitive learning than to reception, and help learners dress up *qipao* properly in relevant social situations. These two categories of information are folded under the guideword **Culture** for learners' further needs and will not appear on the immediate interface. The second part presents the referential meaning of *cheongsam* in old times and indicates the regional label. The usage label *old* is meant to remind learners that this original meaning is old-fashioned and not in current use. An external link is provided for further knowledge of the garment for both females and males. For reception purposes, it is preferable to give both *cheongsam* and *qipao* headword status, whereas cross-references between these two items are required to help with productive purposes. It is a pity that none of the learner's dictionaries mentioned in this study has chosen *qipao* as a lemma, although *OD Online* and *OED online* have granted it main entry status.

cheongsam

1. =qipao, 旗袍:
The curve-accenting, body-fitting Chinese one-piece dress with side slits.
- **Culture**
+ *qipao* types (pictures)
+ dressing up *qipao*: hairstyles, shoes, and accessories
2. (*old*, Cantonese)=cheong sam/chang shan, 长衫:
a long garment for both men and women
+ pictures

The use of a L1 equivalent in this model article is aimed to help learners quickly solve their reception problems or confirm their intended ideas in production situations. Certainly, the L1 parts can be removed or augmented depending on the targeted learners' L2 proficiency level. This model article is demonstrating the possibilities of inserting L1 facilitating props for learners when needed. This paper holds that L2–L1 bilingual solutions should not be reduced to the expedient provision of equivalents. The conception of bilingual dictionary

is far more than furnishing semantic-pragmatic equivalents (Tarp 2008). Some recent explorations have been made to define the concept of bilingual learner's dictionaries (Granger and Lefer 2016; Halpern 2016). Given the focus of this paper, no elaboration is made in this respect.

4.4 Summary

As stated at the beginning, the devices used to represent the meaning of lexical items are diversified, including definitions, examples, equivalents, illustrations, usage notes, usage labels, etc. In many cases, different types of data are required to be combined to help dictionary users comprehend the meaning of the concerned lexical items in different contexts. The model articles in this section demonstrate the possibilities of catering for EFL learners' specific needs with regard to cultural aspects during L2 reception. Highlights are attached to the cultural differences in the conception of the words in L1 and L2 and the concepts typical of L2 culture. Chinese learners' particular problems with English words of Chinese origin are also addressed. The model articles have highlighted the significance of pictorial illustrations when it comes to elucidating the objects or concepts typical of L2 culture. Although technology has extended the concept of illustrations to graphs, diagrams and photographs, the detailed principles of using illustrations presented by Stein in 1991 are still worth careful consideration in terms of meaning representation. Otherwise, the use of illustrations can be superfluous.

5. Conclusion

This paper has described Chinese learners' information needs with regard to cultural aspects during English reception and reveals the deficiency in the present lexicographic practice in this respect. During L2 reception, foreign learners are mainly exposed to the cultural importance of the concepts reflected in L2 lexical units. On the other hand, they are also restrained by their own world knowledge acquired in their mother tongue/L1. Learners' needs for lexicographic assistance with regard to cultural aspects are authentic and imminent in particular contexts. Admittedly, foreign learners can resort to specific reference works for a profound understanding about the culture of the addressed language. But, for the cultural dimensions of a specific lexical item required in an immediate situation, average learners would turn to dictionaries for a quick solution. Therefore, it is essential to include the relevant cultural data in English learner's dictionaries to assist foreign learners of English in different situations.

This paper shows that the inclusion of cultural data, as well as the content of the cultural data should not only rest with the schema of the users within the cultural scope of the addressed language, but also those outside the cultural

circle of the addressed language. It holds true that the digital media has provided great potential for reference works which will serve users' needs more effectively than their print-bound predecessors (Gouws and Tarp forthcoming). However, to serve users' needs effectively calls for a more user-oriented lexicography (Verlinde and Peeters 2012: 162). The medium of a dictionary may alter data quantity, storage and presentation, but never change the quality of the data required to realize the genuine function of the dictionary. This paper has explored the possibilities of customizing lexicographic assistance with regard to cultural aspects during L2 reception, with reference to Chinese learners and advocates further discussion on other ethnic groups of learners in a wider scope of user situations.

Acknowledgements

I would like to thank Professor Sven Tarp for his insightful comments. Special gratitude also goes to Dr. Hugues Steve Ndinga-Koumba-Binza and two anonymous reviewers for their constructive suggestions for improvement. This research is funded by China Scholarship Council (Grant No. 201606435015).

References

A. Dictionaries

- Cambridge Dictionary Online*: <http://dictionary.cambridge.org/>
Cambridge Learner's Dictionary Online: <http://dictionary.cambridge.org/dictionary/learner/>
Collins English Dictionary Online: <https://www.collinsdictionary.com/>
Haici: dict.cn
Jinshan Ciba: <http://www.iciba.com/>
Macmillan Dictionary Online: <http://www.macmillandictionary.com/>
Oxford Advanced Learner's Dictionary Online: <http://www.oxfordlearnersdictionaries.com/>
Oxford Dictionaries of English Online: <https://en.oxforddictionaries.com/>
Oxford English Dictionary Online: <http://www.oed.com/>
Youdao: <http://dict.youdao.com/>

B. Other literature

- Bergenholtz, H.** 2011. A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 187-207. London/New York: Continuum.
- Bergenholtz, H. and H. Agerbo.** 2014. Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries. *Hermes: Journal of Language and Communication in Business* 52: 125-139.
- Chan, A.Y.W.** 2012. The Use of a Monolingual Dictionary for Meaning Determination by Advanced Cantonese ESL Learners in Hong Kong. *Applied linguistics* 33(2): 115-140.

- Chan, A.Y.W.** 2014. Using LDOCE5 and COBUILD6 for Meaning Determination and Sentence Construction: What Do Learners Prefer? *International Journal of Lexicography* 27(1): 25-53.
- Chen, Y.Z.** 2010. Dictionary Use and EFL Learning: A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.
- Ding, J.** 2008. Bottom-up Editing and More: The e-Forum of the English-Chinese Dictionary. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, 15-19 July 2008*: 339-343. Barcelona: Universitat Pompeu Fabra.
- Ding, J.** 2015. A Study of English Majors in a Chinese University as Dictionary Users. *Lexicography ASIALEX* 2: 5-34.
- Dziemianko, A.** 2012. On the Use(fulness) of Paper and Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 319-341. Oxford: Oxford University Press.
- Fatsis, S.** 2015. The Definition of a Dictionary. *Slate Magazine*. Accessed on 15 November 2016. <http://www.slate.com>.
- Gouws, R.H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Gouws, R.H. and S. Tarp.** Forthcoming. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography*. ecw030, <https://doi.org/10.1093/ijl/ecw030>.
- Granger, S. and Marie-Aude Lefer.** 2016. From General to Learners' Bilingual Dictionaries: Towards a More Effective Fulfilment of Advanced Learners' Phraseological Needs. *International Journal of Lexicography* 29(3): 279-295.
- Halpern, J.** 2016. Compilation Techniques for Pedagogically Effective Bilingual Learners' Dictionaries. *International Journal of Lexicography* 29(3): 323-338.
- Heid, U.** 2014. Natural Language Processing Techniques for Improved User-friendliness of Electronic Dictionaries. Abel, A., C. Vettori and N. Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014*: 47-61. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
- Lew, R.** 2013. Identifying, Ordering and Defining Senses. Howard, J. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 284-302. London: Bloomsbury Publishing.
- Lew, R. and J. Doroszewska.** 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.
- Liu, X.Q.** 2015. Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos* 25: 210-232.
- Nesi, H. and K.H. Tan.** 2011. The Effect of Menus and Signposting on the Speed and Accuracy of Sense Selection. *International Journal of Lexicography* 24(1): 79-96.
- Ptasznik, B. and R. Lew.** 2014. Do Menus Provide Added Value to Signposts in Print Monolingual Dictionary Entries? An Application of Linear Mixed-Effects Modelling in Dictionary User Research. *International Journal of Lexicography* 27(3): 241-258.
- Stein, G.** 1991. Illustrations in Dictionaries. *International Journal of Lexicography* 4(2): 99-127.
- Stein, G.** 2002. *Better words: Evaluating EFL Dictionaries*. Exeter: University of Exeter Press.
- 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.** 2009. Reflections on Data Access in Lexicographic Works. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century: In Honor of Henning Bergenholtz*: 43-62. Amsterdam/Philadelphia: John Benjamins.
- Tarp, S.** 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011.

e-Lexicography. The Internet, Digital Initiative and Lexicography: 54-70. London/New York: Continuum.

Verlinde, S. and G. Peeters. 2012. Data Access Revisited: The Interactive Language Tool-box. Granger, S. and M. Paquot (Eds). 2012. *Electronic Lexicography*: 147-162. Oxford: Oxford University Press.

Xia, L.X. and L.W. 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.

Yang, J. 2009. Chinese borrowings in English. *World Englishes* 28(1): 90-106.

Zgusta, L. 1989. Idle Thoughts of an Idle Fellow; or, Vaticinations on the Learners' Dictionary. Tickoo, M.L. (Ed.). 1989. *Learners' Dictionaries: State of the Art*: 1-9. Singapore: SEAMEO Regional Language Centre.

Etymology in the *Polish Academy of Sciences Great Dictionary of Polish**

Przemysław Dębowski, *Institute of the Polish Language of the Polish Academy of Sciences, Poland* (przemyslaw.debowski@ijp.pan.pl)

Bogumił Ostrowski, *Institute of the Polish Language of the Polish Academy of Sciences, Poland* (bogumil.ostrowski@ijp.pan.pl)
and

Jadwiga Waniakowa, *Institute of the Polish Language of the Polish Academy of Sciences, Poland* (jadwiga.waniakowa@ijp.pan.pl)

Abstract: The article offers an insight into etymological information provided in the *Polish Academy of Sciences Great Dictionary of Polish* (Pol. *Wielki słownik języka polskiego PAN*, WSJP PAN). The dictionary and the rules of producing the entries are briefly presented. These rules influence the way of working on etymology within the project. The main part of the article is devoted to the presentation of the field relating to origin and the etymological information which is given in one-word entries. The principles of elaborating upon this information in the dictionary are presented in detail and illustrated with examples. Moreover, some problems, like the question of borrowings, calques and semantic motivation, are discussed in the paper as well.

Keywords: ETYMOLOGY, POLISH ACADEMY OF SCIENCES GREAT DICTIONARY OF POLISH (WSJP PAN), LEXICOGRAPHY, POLISH

Résumé: *Étymologie dans le Grand dictionnaire de la langue polonaise de l'Académie polonaise des sciences.* L'article porte sur les informations étymologiques fournies dans le *Grand dictionnaire de la langue polonaise de l'Académie polonaise des sciences* (pol. *Wielki słownik języka polskiego PAN*, WSJP PAN). Il présente le dictionnaire même, ainsi que les règles de l'élaboration des articles qui influencent la façon de travailler sur l'étymologie dans le cadre du projet. La partie principale de l'article est consacrée à la présentation du champ relatif à l'origine et aux informations étymologiques qui sont données dans les articles n'ayant qu'un seul mot comme entrée. Les principes de l'élaboration de ces informations dans le dictionnaire sont présentés en détail et illustrés avec des exemples. De plus, l'article offre également une discussion d'autres problèmes, à savoir, la question des emprunts, les calques et la motivation sémantique.

Mots-clés: ETYMOLOGIE, GRAND DICTIONNAIRE DE LA LANGUE POLONAISE DE L'ACADÉMIE POLONAISE DES SCIENCES (WSJP PAN), LEXICOGRAPHIE, POLONAIS

* This is a slightly modified English version of the paper "Etymologia w *Wielkim słowniku języka polskiego PAN*" which will be published in the journal *Prace Filologiczne* (2017).

1. Introduction

The aim of the article is to present the principles of elaborating upon etymological information in the most important lexicographic project being carried on in Poland in the 21st century, namely the *Polish Academy of Sciences Great Dictionary of Polish* (Pol. *Wielki słownik języka polskiego PAN*, WSJP PAN). First of all, the dictionary and the rules of producing the entries are briefly presented. These overriding rules influence and condition the way of working on etymology within the project. In this situation, the task of etymologists is to find a balance between the traditional principles of offering etymological information and the requirements of an electronic dictionary of a purely synchronic character, available to a wide audience from all over the world.

Therefore, the main part of the article is devoted to the presentation of the field relating to origin and the etymological information given in one-word entries, which are the most important ones for the etymologists working in the WSJP PAN. The rules of elaborating upon this information in the dictionary are presented in detail. They embrace various types of words from the point of view of their origin and are illustrated with examples. Finally, some other problems are discussed in the paper as well, including among others: borrowings, calques, semantic motivation, problematic etymologies — in short, the questions that etymologists inevitably encounter in their everyday work.

2. About the *Polish Academy of Sciences Great Dictionary of Polish* (WSJP PAN)

As one can learn from the introduction to the dictionary, authored by its editor, Piotr Żmigrodzki¹, the *Polish Academy of Sciences Great Dictionary of Polish* is a documentary dictionary, i.e. one which is based on authentic data.² The basic source of the dictionary is in the first place the *National Corpus of the Polish Language* (Pol. *Narodowy Korpus Języka Polskiego*, NKJP). Another important source of data is a supplementary corpus created at the Institute of the Polish Language of the Polish Academy of Sciences especially for the sake of the dictionary, which is comprised of texts that are not a part of the NKJP. Yet another different source is the Polish Internet, besides which, the authors' own choice of entries is possible too. To a limited extent the authors of the entries also use other lexicographic resources developed at the Institute of the Polish Language of the Polish Academy of Sciences. The WSJP PAN includes only those lexical units (words, phraseological units, proverbs) which have been attested to in those resources.

The WSJP PAN is a dictionary of Contemporary Polish in a broad sense of the word, as it covers the time beginning with the year 1945, yet one needs to emphasise that due to the character of the sources the majority of the material comes from the last decades of the 20th century and the beginning of the 21st century. The WSJP PAN is at the same time a descriptive dictionary, and not a

normative one. It needs to be stressed that the authors do not remove from the description those language forms which are considered to be incorrect. They only limit their actions to informing the user about the normative unacceptability of the noted facts, following the *Great Dictionary of Correct Polish* (Pol. *Wielki słownik poprawnej polszczyzny PWN*, see Markowski WSPP) and the stylistic qualification of substandard units.

The WSJP PAN is at the same time an academic dictionary due to the fact that the authors strive to apply the achievements of the 20th century Polish linguistics as much as possible, especially with regard to the semantic, inflectional, and syntactic description of the lexical units. However, the description provided by the dictionary is at the same time user-friendly, because it is addressed to the broadest possible group of Polish speakers.

The WSJP PAN is finally an electronic dictionary *sensu stricto*, i.e. it has been planned as such from the very beginning, and not as a transfer of the printed dictionary into an electronic form. Due to this it is possible to apply an entry structure and a form of presentation that breaks away from the limitations typical of paper dictionaries.³ One can find the following types of entries in the dictionary: single words (one-segment units) representing various parts of speech, phraseological units, most frequently used proverbs, abbreviations, acronyms and selected proper names.⁴

The dictionary can be accessed on the Internet at the following address: <http://wsjp.pl>.

3. Etymological information in the WSJP PAN

3.1 The task of etymology

As is commonly known, the task of etymology is to establish the origin of words. It is worth emphasising that an etymological investigation aims for the reconstruction of a previous, even original form of the word, its root, and of an earlier, possibly original meaning, i.e. its etymological meaning. It is then important to reconstruct the initial motivation for the word and explain its form, i.e. define its morphological structure. Thus, etymology deals with both the formal side of vocabulary and revealing its semantic motivation and development. None of the aspects of the reconstruction, neither the formal nor the semantic one should dominate and none of them can be ignored.⁵ A rule is applied here which may be called the "law of the form and content agreement", an investigation concerning the formal development must always go hand in hand with the studies of the development of the meaning.⁶ Etymological studies rely on the methodology of historical-comparative grammar, which allows for identification of historically related words, i.e. those which are equivalent to each other, after the application of adequate language laws. One may talk about credible etymology when all the structural modifications from a given form going back to the original form may be explained by means of the influ-

ence of specific language laws, while semantic changes are rational and have semantic parallels.

It is obvious that the issue concerning etymological investigation is extremely diverse in its character, as the form of the word may not always be explained as inherited or borrowed. It may also be a result of morphological, word-formation (as e.g. the Pol. KAWIARNIA 'a place for drinking coffee') or semantic processes (semantic extension, narrowing or shift, cf. e.g. the Old Polish BIEGUN 'runner, messenger' and the Contemporary Polish BIEGUN in a number of scientific meanings: 'pole', 'rocker', etc.).

3.2 Selection of entries in the WSJP PAN and etymology

An important issue for the team of etymologists employed by the WSJP PAN to consider are the principles of entry selection in the dictionary which condition, indeed, define the steps taken when drafting the information concerning the origin of words. In the WSJP PAN the criteria for selecting entries are the graphic form of the entry word and the grammatical interpretation of the unit. Thus, separate entries describe units that differ in terms of spelling (capital/lower case letter), e.g. BARAN (an animal – 'sheep'; skin – 'sheepskin'; about a man – 'dunderhead') – BARAN ('Aries': constellation; a Zodiac sign; a person born under this Zodiac sign) as well as units of a different grammatical categorisation at the level of so called parts of speech. The consecutive entries of the same shape are indexed with Roman numerals, e.g.:

- PIEC I (verb – 'to bake'), PIEC II (noun – 'furnace, stove');
- ABSOLUTNIE I (adverb – 'absolutely'), ABSOLUTNIE II (functional word used by the speaker to emphasize that he has no doubts about what he is saying);
- PIĘĆ I (numeral – 'five'), PIĘĆ II (noun – 'a very good mark at school');
- ZIELONY I (adjective – 'green'), ZIELONY II (noun – 'the colour green').⁷

As a result, the entry articles are not divided on the basis of their etymology, as — according to the dictionary authors — this would mean a confusion of the criteria of synchronic and diachronic character. Besides, a consistent adherence to etymology in connection with the use of only the direct source of borrowing in the entries would have to lead to a significantly larger number of "double" entries (as for a number of meanings of lexemes, such as e.g. MYSZ 'a device for navigating the computer', SEZON 'a set of episodes of a TV series', which are semantic borrowings from English *mouse* and *season*, one would have to postulate separate entries). These types of rules concerning the entry selection in the WSJP PAN have resulted in the information about the origin being given separately for each meaning-subentry.

Naturally, in many entries the origin is common for all the subentries.⁸ However, entries where the origin of particular subentries is different are rela-

tively frequent. An example of such a situation may be a polysemous entry ŁATA, which has as many as 11 meanings-subentries in the WSJP PAN:

- (1) 'a piece of material sewn on a hole in clothing in order to cover it or in some other worn area in order to reinforce it';
- (2) 'a piece of material of a certain colour, shape or with a certain inscription, worn on clothes by people who are in this way stigmatised';
- (3) 'what fills in or covers a hole formed in something';
- (4) *colloquial* 'a correction or an update of an earlier version of a computer programme in order to remove defects, add new functions to it or enlarge its productivity';
- (5) 'a spot of a contrasting colour against its background';
- (6) 'a part of beef carcass cut out between the round, the plate and the loin';
- (7) 'a large, wooden board square or rectangular in shape, fastened with nails to elements of a wooden construction, mainly a roof';
- (8) 'a smooth board used for checking the evenness of the layers of plaster, floor and other surfaces and for their levelling';
- (9) 'an instrument in the shape of a wooden strip with a scale, used for measuring length and differences in height';
- (10) *technical* 'a device in the shape of a post with a scale on it driven into the river bed, used for measuring the water level in the river';
- (11) *technical* 'in the wood industry, lumber with a roughly square cross section and 32 mm to 10 cm thick'.

The first three senses as well as the fifth and sixth one have the Proto-Slavic **lata* 'flap, petal; insert, patch (on clothing, shoes)' as a base, the fourth meaning is a calque from the English word *patch*, the seventh and eighth senses are derived from the Middle High German *latte, late* 'roof patch, prop, beam', while the ninth, tenth and eleventh come from the German *Latte* 'slat'.⁹

3.3 Principles of providing etymological information in the WSJP PAN

The principles of providing information concerning the origin of particular words have changed in relation to what was binding at the outset of the work on the WSJP PAN and what was applied in practice in entries completed before 2012 (cf. the above footnote 2). It was assumed then that etymological information would be given only in entries describing foreign words. Beginning with 2013 etymological information is provided in all the entries in preparation which have a one-word entry form belonging to the following types: regular entries (nouns, adjectives, adverbs, verbs, pronouns, and numerals) and function entries (e.g. prepositions, particles) as well as in some phraseological entries. In other words, etymology is given for each one-word entry, whereas multiple-word entries usually do not contain etymological information. In the

one-word entries which were drafted earlier (before 2013), the origin of words is also gradually being supplemented at present.¹⁰

In the case of native words etymological information is usually limited to quoting a reconstructed Late Proto-Slavic form which retains a suitable notation following the data from etymological dictionaries of the Polish language and detailed etymological studies. Here are some examples:

- GŁOWA 'head': <Proto-Slavic **golva*>;
- SPAC 'to sleep': <Proto-Slavic **sъpati*>;
- PIĘKNY 'beautiful': <West Proto-Slavic **pěknъ* 'motley, coloured, colourful, decorative'; the nasal vowel *-ę-* in *piękny* is a secondary development>.

When the etymology of a given word is not obvious (for instance when the views of etymologists diverge), and we quote it from a specific source, we use the reference in the form of an acronym of this work. For example the abbreviation "Bor" at the end of the information providing the origin of the word in the WSJP PAN indicates that it was based on the interpretation of facts and linguistic premises postulated by Wiesław Boryś in his *Etymological Dictionary of Polish* (SEJPBor).

In the case of derivatives explained synchronically only their derivational base is indicated, i.e. the entry where one can find full information concerning its origin, e.g.

- KWIATEK 'flower': <see *kwiat*> ('flower');
- POGODNY 'sunny; cheerful': <see *pogoda*> ('weather');
- GRZANKA 'toast': <see *grzać*> ('to heat').

As regards the method of referring to other entries in the case of native words and early borrowings, we rely on the system of cross-referents used in SEJPBor. This approach ensures cohesion and referential uniformity of etymological information in all the WSJP PAN entries.

In the case of a derivative which is not explained synchronically, full information concerning its origin is provided, just as in the case of remaining native words, cf.:

- PRZYCHÓD 'income': <verbal noun from the prefixed verb *przychodzić*; see *chodzić*>.

In some entries the meaning of the Proto-Slavic etymon differs from the sense (senses) of the Polish continuation to such an extent that its semantic development needs to be explained and presented. Thus, in entries in which it is necessary, we devote some attention to historical semantics, especially if the explanations refer to some not so broadly known aspects of Slavic culture, cf. e.g.

- PRZYSIĄC 'to swear': <Proto-Slavic **prisęgti* 'to reach, touch'; the sense of 'swearing an oath, taking a pledge' is secondary, it probably came into

being in a phrase **prisęgti zemjō* 'to touch, reach the ground, the soil', hence 'to take a pledge, swear an oath (on the earth)', in connection with the ritual of taking a solemn vow with ancient Slavs during which the person taking a vow touched the ground or ate earth, kissed it or held it in the hand (Bor);¹¹ see *sięgać*>.

The origin of foreign words, fully or partly borrowed from another language, i.e. borrowings and calques, is established on the basis of data from dictionaries of foreign words, information from etymological dictionaries, monolingual and bilingual dictionaries and from detailed etymological studies. An example of a fully imported word is:

— KAMUFLAŻ 'camouflage': <French *camouflage*>.

A foreign word that is the base of the borrowing is quoted in its original spelling (with the maintenance of diacritics) when a given language is written with the help of Latin script, cf. e.g.:

— BANKRUT 'bankrupt' <French *banqueroute*>.

The graphic form of a foreign word that is the base of a borrowing is always provided when it differs from its graphic form in Polish. In the case of a word coming from a language which uses a different alphabet than the Latin script, the word that constitutes the base of the borrowing is given in transliteration, with the retention of suitable diacritics, e.g.:

— BORYKAĆ SIĘ 'to struggle, cope': <Ukrainian *borykătysja*>.

The length of a sound (in the form of a horizontal line above the letter) is marked only in the case of long vowels, cf.:

— LUMBAGO: <Latin *lumbāgō*>.

The meaning of the foreign word in the source language is always given when it differs from the sense of the borrowed word. Listed below are some examples of detailed situations connected with borrowings:

- (1) a given word has been borrowed from another language, yet its spelling and sense have not been modified in comparison to the source word in that, e.g.:
 - BARMAN: <English>;
- (2) a given word has been borrowed from another language, but its graphic form has been modified in relation to the source word in the language from which it has been borrowed, while the meaning of it in both languages is the same, e.g.:
 - MAKIJAŻ 'makeup': <French *maquillage*>;

- (3) a given word has been borrowed from another language, its graphic form has not been changed in relation to the source word in that language, however, the meaning of the word in the two languages differs, e.g.:
- FILIA 'a branch of an institution, company (...)': <Latin 'daughter'>;
- (4) a given word has been borrowed from another language, its graphic form has been changed in relation to the source word in that language, and the meanings in the two languages are different, e.g.:
- ADWENT 'in the catholic liturgy the period covering the four weeks before Christmas, advent': <Latin *adventus* 'arrival, coming'>.

Among the WSJP PAN entries one may also come across borrowed words-quotations. The same principles of etymological description apply to them as to the other borrowings, cf.:

- AD REM 'relevant, pertinent': <Latin>;
- AB OVO 'from the beginning': <Latin 'from the egg'>.

Some borrowings found in the WSJP PAN also belong to the category of internationalisms. If a word that is the base of a borrowing appears in more than one language and there is no certainty as to which one it has been taken from, the qualifier "internac." (< Pol. *internacjonalizm*) is added as well as symbols of these languages along with forms of the word to be found in them. Moreover, internationalisms are quite often composed of morphemes which have their origins in Latin and/or Greek. In such a case the etymological description comprises not only forms in the three foreign languages, but also the origin of those forms, including Greek and/or Latin morphemes. Here is how it looks:

- AGRESJA: <international: Eng. *aggression*, Fr. *agression*, Germ. *Aggression*, from Latin *aggressiō*>;
- LEUKOCYT: <international: Eng. *leucocyte*, Fr. *leucocyte*, Germ. *Leukozyt*, from Neo-Latin *leucocytus*>; <Gr. *leukós* 'white' + Gr. *kýtos* 'jug, vessel; body'>.

In the last example one can see that if borrowings are composed of foreign morphemes, these morphemes are always presented together with their meaning, similarly as in e.g.:

- MAGNETOTERAPIA: <international: Eng. *magnet therapy*, Fr. *magnétothérapie*, Germ. *Magnettherapie*>; <Late-Greek *mágnēs, -ētos* 'magnet' + Greek *therapeía* 'care, respect, treatment'>.

In the case of borrowings we also very often deal with a situation in which some language mediates in the process of borrowing. Then as the first language provided is the one from which the word was directly borrowed by Polish, and further sources are listed (if they are evident, commonly known or important in some other way), e.g.:

- POLITYK 'politician': <Germ. *Politiker*, from Latin *politicus*>.

If some word is a derivative of a foreign word, the dictionary uses a cross-reference to the derivational base. In the indicated entry, full information concerning the origin is provided, e.g.:

- BANKRUTOWAĆ 'to go bankrupt': <from *bankrut*>;
- BARMANKA 'barmaid': <from *barman*>.

The field of origin in the WSJP PAN offers not only strictly etymological information. This is usually a concise commentary which may be helpful in the understanding of the meaning of the unit, and it is not a part of its definition. It most often concerns eponyms, winged words, proverbs, phraseological units with components referring to characters, geographical names or historical events (biblical, mythological) as well as terms with a proper name as an element. We can quote the following entries as examples:

- KURONIÓWKA *colloquial* 'unemployment benefit' <from the name: Jacek Kuroń (1934–2004), social and political activist, minister of labour and social policy in the years 1989–90 and 1992–93)>;
- KOŚCI ZOSTAŁY RZUCONE 'the die is cast; something has happened whose results cannot be reverted': <words which were allegedly uttered by Julius Caesar when crossing the Rubicon river and this way starting a civil war in Rome in 49 BC>.

3.4 Problematic etymologies

As mentioned above, the origin of words is not always obvious. We often need to decide which of the numerous etymological conceptions offered in various studies is more reliable. Then we apply strictly defined principles which allow us to choose a better reconstruction of an earlier form and meaning, as well as a more probable semantic motivation.¹² We can use the word MANDARYNKA 'mandarin orange' as an example. Tokarski's *Dictionary of Foreign Words* s.v. (Pol. *Słownik wyrazów obcych*, see Tokarski SWO) suggests that it may derive from the noun *mandaryn* 'a high public official, dignitary and scholar in ancient China' due to the colour of the robes of those Chinese officials, or from the island of Mandara = Mauritius. Sobol's *Dictionary of Foreign Words* s.v. (Pol. *Słownik wyrazów obcych*, see Sobol SWO) presents only one option, namely that *mandarynka* is a borrowing from the French *mandarine*, which in turn comes from the Spanish (*naranja*) *mandarina* 'an orange from the Island of Mandara' (the Island of Mauritius in the Indian Ocean).¹³ The only way out in this case was to check the origin of the expression in Spanish. It allowed for a final conclusion that the French (*orange*) *mandarine* comes from the Spanish (*naranja*) *mandarina*, in a literal sense 'the Mandarin (orange)', and the semantic motivation of the *mandarine* is the colour of the robes of mandarins, and not the name of the Island of Mauritius.¹⁴

When compiling the information concerning the origin of words for the

WSJP PAN we must often correct erroneous etymologies present in etymological dictionaries and dictionaries of foreign words. It may be illustrated by the example of the entry WOLONTARIUSZ: <Latin *voluntarius* 'volunteer'>. In Sobol SWO s.v. an incorrect origin of the word from the French *volontaire* is given, while both the form of the word (the ending *-usz*) and the historical attestations (with *-u-*) as well as their dating (cf. WSJP PAN s.v.) point to the borrowing from Latin.

When working on etymology we also have to deal with a situation when the etymology of a given word is missing. The problem usually concerns colloquial, older, specialist or dialectal vocabulary, mainly that which originates from foreign languages.¹⁵ Its origin has not been investigated in Polish so far. An example may be the entry PACYNKA, which has four subentries, and almost each one has a different origin. Especially the meaning 'a puppet fitted over the hand which moves thanks to the movement of fingers' has caused us numerous problems. After a long investigation we opted for the following solution: <possibly: Germ. *Patsche* 'hand, little paw', possibly: Pol. **palcynka* 'a puppet manipulated by the movement of fingers'; unclear etymology; the word propagated by Jan Sztaudynger, an expert on puppet theatres, who — according to his own words — borrowed it from his grandfather> (cf. WSJP PAN s.v.).¹⁶ Such mysterious words which do not have an established etymology are numerous. Listing them, however, would take too much space. The task of etymologists is to investigate what their origin is likely to be.

A considerable problem in the study of the origin of words in the WSJP PAN are calques. Indeed, although many of them have long been identified, numerous words (and still new ones) lead to the conclusion that they may reflect the semantic structure of elements from another language. The problem of calques is most complex for us despite a relatively rich literature on the subject.¹⁷ In general terms, we can suspect the existence of calques in contemporary Polish vocabulary especially when we deal with the newest lexis and terminology, particularly in the field of economy, politics and computer technology. Of certain help in this respect, although to be approached with a large dose of caution, is the ESJPBańk.¹⁸

4. To wind up

To sum up the discussion, one needs to clearly emphasise that the field of origin in the WSJP PAN, as of now, does not provide etymology *par excellence*, but only a concise piece of etymological information which allows the reader to gain some orientation as to where a given word comes from. As a result, derivatives are cross-referenced to their derivational bases, and only the entries which describe basic words provide some broader knowledge concerning their origin. Despite offering such limited information concerning etymology, we fully apply all the methodological principles to be adhered to in the investiga-

tion of the origin of words. We meticulously analyse the form and semantics of each word, at the same time paying the utmost attention to the chronological aspect, which is particularly important as regards borrowings, since much can be deduced from the knowledge of when a given word was attested to the first time.

Notes

1. Cf. http://wsjp.pl/pobieranie/Zasady_opracowania_WSJP.pdf, p. 7.
2. The first version of the rules of the dictionary preparation was drafted in 2012 and can be found at the following address: <http://rcin.org.pl/dlibra/docmetadata?from=rss&id=53406>.
3. More broadly about the WSJP PAN, its history and prospects cf. Żmigrodzki 2015.
4. Cf. http://wsjp.pl/pobieranie/Zasady_opracowania_WSJP.pdf, p. 8.
5. The literature on this subject is very extensive, cf. e.g. Ślawski 1958 and, recently, Jakubowicz 2010: 35–44.
6. Language laws, methodological principles followed in the reconstruction of the form and meaning, and semantic changes may be a subject of separate discussion, yet there is no space for analysing these problems here.
7. Cf. http://wsjp.pl/pobieranie/Zasady_opracowania_WSJP.pdf, p. 8.
8. Cf. http://wsjp.pl/pobieranie/Zasady_opracowania_WSJP.pdf, p. 9.
9. Obviously, the German *Latte* 'slat' continues the Middle High German *latte*, *late* 'roof patch, prop, beam'.
10. Cf. http://wsjp.pl/pobieranie/Zasady_opracowania_WSJP.pdf, p. 20.
11. As the (Bor) abbreviation indicates, it is Boryś' own explanation excerpted from SEJPBor.
12. Quoting the rules of etymological methodology would exceed the limitations of this paper.
13. Such an etymology can also be found in Andrzej Bańkowski's *Etymological Dictionary of Polish* s.v. (Pol. *Etymologiczny słownik języka polskiego*, see ESJPBańk). It is worth mentioning that Bańkowski also developed its etymology in Sobol SWO.
14. It appears that the Island of Mauritius has never had the name of Mandara. What source the conception presented in both dictionaries is based on is unknown, at the same time, it may also be seen that dictionaries repeat a number of claims one after the other without any further in-depth verification.
15. At this point it is worth adding that etymology has been provided for a certain portion of the Old Polish vocabulary listed in the glossary of Old Polish edited by Decyk-Zięba and Dubisz (2008), while an etymological dictionary of Polish dialects is currently in preparation (cf. Waniakowa 2013).
16. A paper by Jadwiga Waniakowa on the origin of the word *pacynka* (in different meanings) is currently in preparation.
17. For instance, the recently published guide to English loanwords in Polish (Witalisz 2016), which particularly devotes much attention to various types of calques from the English language, has proved very helpful in this respect.
18. Bańkowski sees calques almost everywhere he turns. He spotted them even where they are quite unlikely, and certainly not evident. In fact, he never considered the possibility of their independent, parallel semantic development in the Polish language.

References

- Decyk-Zięba, Wanda and Stanisław Dubisz (Eds.).** 2008. *Glosariusz staropolski. Dydaktyczny słownik etymologiczny*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- ESJPBańk = Bańkowski, Andrzej.** 2000. *Etymologiczny słownik języka polskiego*. T. 1: A–K. T. 2: L–P. Warszawa: Wydawnictwo Naukowe PWN. The next volume published as: Bańkowski, Andrzej. 2014. *Etymologiczny słownik mowy polskiej*. T. 3: R. Częstochowa: Linguard.
- Jakubowicz, Mariola.** 2010. *Drogi słów na przestrzeni wieków. Zarys słownika motywacji semantycznych na materiale przymiotników słowiańskich odziedziczonych z prasłowiańszczyzny*. Warszawa: Sławistyczny Ośrodek Wydawniczy (SOW) Instytut Sławistyki Polskiej Akademii Nauk.
- Markowski WSPP = Markowski, Andrzej (Ed.).** 2004. *Wielki słownik poprawnej polszczyzny*. Warszawa: Wydawnictwo Naukowe PWN.
- NKJP = Narodowy Korpus Języka Polskiego:** <http://nkjp.pl/>.
- SEJPBor = Boryś, Wiesław.** 2005. *Słownik etymologiczny języka polskiego*. Kraków: Wydawnictwo Literackie.
- Sławski, Franciszek.** 1958. Uwagi o badaniach etymologicznych nad słownictwem słowiańskim. *Z Polskich Studiów Sławistycznych, Językoznawstwo. Prace na IV Międzynarodowy Kongres Sławistów w Moskwie*: 99-107. Warszawa: Komitet Słowianoznawstwa i Komitet Językoznawstwa Polskiej Akademii Nauk.
- Sobol SWO = Sobol, Elżbieta (Ed.).** 1999. *Słownik wyrazów obcych*, wyd. I, dodruk X. Warszawa: Wydawnictwo Naukowe PWN.
- Tokarski SWO = Tokarski, Jan (Ed.).** 1980. *Słownik wyrazów obcych*. Warszawa: Wydawnictwo Naukowe PWN.
- Waniakowa, Jadwiga.** 2013. *Etymologiczny słownik gwar polskich — nowe zadanie w badaniach historyczno-porównawczych*. *Polonica XXXIII*: 317-325.
- Witalisz, Alicja.** 2016. *Przewodnik po anglicyzmach w języku polskim*. Biblioteczka Towarzystwa Miłośników Języka Polskiego nr 27. Kraków: Wydawnictwo JAK.
- WSJP PAN = Żmigrodzki, Piotr (Ed.).** 2006–. *Wielki słownik języka polskiego PAN*: <http://wsjp.pl>.
- Żmigrodzki, Piotr.** 2015. *Wielki słownik języka polskiego PAN — historia, stan obecny i perspektywy rozwoju po 2018 roku*. *Biuletyn Polskiego Towarzystwa Językoznawczego* 71: 177-187.

Philip Durkin. *The Oxford Handbook of Lexicography*. 2016, xxiii + 698 pp. ISBN: 978-0-19-969163-0. Oxford: Oxford University Press. Price £95.00.

The Oxford Handbook of Lexicography is a collection of papers in 37 chapters by various authors segmented in four parts with respective themes — synchronic, diachronic, specialist lexicography and specific topics related to lexicography. Contributors to this handbook with either established lexicographical methods or lexicographic and linguistic background turn their experiential compiling practices or research into their articles. The authors have included not only theories, quotations and lexicographic examples, but also references, illustrations of entries and electronic displays, which simplify the text and facilitate understanding.

Part I deals with the three types of synchronic dictionary and main micro-structure features: the dictionary for general users (DGU), the monolingual dictionary intended for L2 learners (MLD), the bilingual dictionary and definition (including definition in general, explanation in MLD and equivalence in the bilingual dictionary in chapters 7, 8 and 9 respectively). The description is in each case uniformly initiated by the original motivation, development, current issues, conclusions and future prospects. Representing a different approach from these more customised ones, such as MLD, DGU faces a somewhat declining tendency which Bejoint believes is still "a quasi-universal need" (Durkin 2016: 24). The enviable supremacy of MLD is evidenced by the creative and dynamic way of explaining meaning to learners. Among the current issues in bilingual lexicography, the "definition" in bilingual lexicography is the foremost component, the four major equivalence strategies proposed in chapter 9, i.e. cognitively, translationally, explanatorily and functionally equivalent, though with no sharp boundaries, together with a supplementary meaning-elucidating strategy, can ensure successful explanation of source language meanings. In chapters 5 and 6 corpus construction and interrogation, an indispensable methodology in dictionary compilation and research, is introduced.

Part II discusses historical dictionaries, reflecting their distinctive methodologies and approaches that are not well-known to outsiders. The structural and presentational challenges posed by the diachronic perspective in historical dictionaries are manifested by the uncertain exposition of the history of a word or a loanword and the documentation of the history of a word in an entry which reveals the duplicate effect of coincidental individual usage by negligence, subsequent gradual secularization and popularization and possibly grafting from other languages. In the electronic era, online corpora, though not specifically for the historical lexicographer, provide sufficient and well-balanced quotations for the attestation of definition on the one hand; on the other, they help ease the strain of identifying the etymology or the homonymy.

Part III presents some of the most important types of specialist dictionaries and their pertinent existence, histories and methodologies. Contributors are required to show not only the highlights and peculiarities but also commonalities connecting the universal lexicographic works. These dictionaries cover the

theoretical and practical issues of specialist lexicography in ten different domains (i.e. place names and person's names, pronunciation, spelling, slang, etymology, dead languages, thesauri, regional dialects and science and technology), roughly beginning with the history and status quo of a specialist dictionary, definition of relevant terminologies and boundaries of inclusion and exclusion, and then the basic arrangement and structure, reciprocal issues concerning specialised lexicography and academic research of other domains, and prospects and challenges in the digital age. Specialist lexicography features varied selection source and authoritative and subjective expert item selection as chapter 24 points out that "the use of corpora has limited applicability" (Durkin 2016: 398), while the consultation of experts from the respective fields withstands interrogations, because the hierarchical distinction of scientific and technical terms and general words is essential. The same applies to the DUDEN German spelling dictionaries that insist on their own principles regardless of official orthographic use. After the meticulous and systematic research of synchronic and diachronic thesauri, contributors Kay and Alexander hold that the thematic layout "becomes the most attractive hybrid" (Durkin 2016: 380) in the digital age without causing browsing problems. This part also reveals the close relationships between specialist lexicography and relevant topics that still present an instance of "the road less taken" and require in-depth research. The contributors agree with Fuertes-Olivera and Tarp, who note that "specialised lexicography is an interdisciplinary activity" (2014: 192).

Part IV deals with other lexicographic problems, most of which are yet to be solved with significant implications, as well as the discussions of the practicalities and theories in dictionary compilation. The proposed two-way traffic between cognitive linguistic theory and lexicography in chapters 26 and 27 (echoing chapter 7 in Part I) remains one of possible avenues for further lexicographic research. Complexities arise when handling multiword units and meaning and sense relations (polysemy, homonymy, hyponymy, metonymy, synonymy and antonymy) in dictionaries. Many researchers (of chapters 25, 26, 27, 28, 33, 35, etc.) recommend the combination of corpus and human methods, i.e. the interrogation of corpus together with the application of lexicographers' instinct and experience. In chapter 31 even "more versatile possibilities of electronic platforms" (Durkin 2016: 513) are called for in producing more target-oriented and problem-specific lexical tools, while in chapter 36 the web presence of the end product may help maintain the publishers' competitive edge. The chapters in this part taking up the clues left in previous parts make the whole book a unified whole. The final two chapters are innovative in topic. Nesi presents the modern digital world and numerous demanding users have influenced the way dictionaries are used and produced, while in the last chapter the correlation of political, historical and social factors with the editing and publication of dictionaries of national varieties of international languages is elaborated.

This book can be recommended for four main reasons.

First of all, it covers a broad spectrum of research topics and brims with

insights useful for the practicalities and theories of lexicography of all types. The weight placed on the treatment of diachronic lexicography and specialist lexicography makes up for the deficiency in lexicographic handbooks of the same sort, which is to "draw into focus the very many similarities in both the challenges and opportunities each faces in the contemporary environment" (Durkin 2016: 3). The well-proportioned coverage as well as cross-references within chapters make the articles written by different individuals an organic whole. Speculative research always coincides with current findings, outlooks, uncertainties and problems to be solved, such as the questions for definition writers at the conclusion of chapter 7, which are re-examinations of existing works as well as inspirations for future research.

Secondly, salutary lexicographic lessons are learned from historical and specialist lexicography. Many chapters are not purely historical accounts but history-inspired research, as lexicographical issues, although they are contemporary trends, also echo the past. The diachronic development of lexicography (general and specialised), especially the historical construction of the micro-structure of dictionaries sheds light on the present and future of lexicography. For instance, when the historical principle is at odds with the logical principle in sense ordering, Passow suggests "a conveniently ordered overview" (Durkin 2016: 168), a common practice afterwards, which was documented by Zgusta and with which Considine agrees. By the same token, some of the phenomena and issues in specialist lexicography are also universal to the general and learners' dictionaries, such as the gap between the original native use of a dead language and the indirect linguistic material consulted by lexicographers. This is the concern of any lexicographer, accordingly justifying the so-called marginal researches in specialist lexicography in the whole lexicography paradigm. Those inspiring lessons make the specialised dictionaries a boon for learners.

In addition, the volume presents a thorough methodology of lexicographical research. There is no unified writing template, as those chapters are written by different researchers. All articles, however, probe into the relevant academic study (in diachronic and/or synchronic sense) followed by the re-inspection and/or demarcation of the current issues and exhaustive study of the specific dictionary scenario or topics, particularly in the context of the electronic era. For instance, in chapter 18 the relation between orthography and graphematics is the prerequisite for research pertaining to German spelling dictionaries. The illustration combines direct quotations from existing dictionaries or corpora with references to works of predecessors or peers. With recommended readings, the points of view in the research papers seem more convincing and sophisticated.

Moreover, a chronology of major events of lexical dictionaries affixed to the content before the references pieces together the canonical lexicographic works in human history. From as early as about 3200 BC when the earliest Sumerian wordlist was written on clay tablets to 2009, the year when the Historical Thesaurus of the OED was published, the list provides a panorama of highlights in world lexicography, notably the dictionaries of English of all

sorts. It can serve both as an introduction to lexicography for beginners and amateurs as well as research clues for lexicographers and scholars.

In spite of the wide coverage and many intriguing questions authors raise, this book could have been even better from my point of view. On the one hand, it could have been more comprehensive and updated if it had incorporated more detailed microstructure discussions in Part IV and if the analysis of linguistic data had been more from the perspectives of usage-based theories such as cognitive linguistics. There are chapters and pieces of research advocating the significance of corpora and the Internet in the evidence and entry selection. The role of users played in decision-making about inclusion and exclusion as well as the format of dictionaries have been discussed in Part IV. However, also on the readers' expectation list are studies such as the semantics of concrete words and the multiword units, illustrative examples, collocations etc., and the exploration of microstructure and the user's survey from the perspective of cognitive linguistics, as proposed by Kövecses and Csábi in *Lexicography and Cognitive Linguistics*. On the other hand, many chapters represent an exhaustive study on topics related to lexicographic research and practice — so exhaustive that at some point the reader may lose track in the profound and sophisticated elaborations and the large number of references to various quotations and unparalleled compilation experiences or in the complex explanations of complicated linguistic phenomena. The reading collection is splendid, but is less easily digested. It would be more informative and useful if the language of some authors were less obscure and abstruse and more explicit and straightforward.

Interestingly, controversies arise when those sparking ideas are in disagreement for which there is no consensus today. Hanks claims the real definition in dictionaries is meaning potential, but being cautious of the distinction between norms and exploitation and the dependency of a given word's meaning on context in chapter 7, while Heuberger proposes componential analysis to "identify distinctive features" (Durkin 2016: 33) in chapter 3. Adamska-Salačiak, however, holds the view in chapter 9 that Šcerba's principles of providing L1 explanations of meaning in bilingual dictionaries fail in practical lexicography. This is true according to Hanks. However, the dominant translatory equivalence (or translational equivalence) in bilingual dictionaries is not always qualified for bearing the meaning potential; it reflects, instead, the most salient features of the L1 headword. To serve the end of comprehension and production for perplexed users when encountering a strange word or wrestling for an unknown expression in L2, the equivalence in the bilingual dictionary should manage to provide the meaning potential in L2 in a similar way to the one in monolingual learners' dictionaries.

As a whole, this book is a valuable addition to a set of introductory books on lexicography. The good organization of the selected topics into thematic sections and a logical sequence of chapters adds to the success of the presented subject matter. It functions as a mirror of contemporary lexicography, with

equilibrium in the layout and a structure of synchronic versus diachronic lexicography; the general concepts versus learners' and specialist lexicography, as well as special topics. The volume acts as "a guide to the most significant contours in the geography of the lexicographical world" (Durkin 2016: 1). Major issues confronting lexicography today are presented in an engaging and accessible way. It is interesting and inspiring for amateur readers and scholars of lexicography for its lucid but speculative diction and clear textual design.

References

- Fuertes-Olivera, Pedro A. and Sven Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: Walter de Gruyter.
- Van Sterkenburg, Piet (Ed.)**. 2003. *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins.
- Kövecses, Zoltán and Csábi, Szilvia.** 2014. Lexicography and Cognitive Linguistics. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics* 27(1): 118-139. Amsterdam/Philadelphia: John Benjamins.

Li Dehui
College of Foreign Languages and Cultures
Xiamen University
Xiamen
P.R. China
(Jessica8604@126.com)

Die Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. *Afrikaanse Woordelys en Spelreëls*. Elfde, verbeterde en omvattend herbewerkte uitgawe. 2017, xxxiii + 757 pp. ISBN: 978-1-86890-201-1 (Sagteband). Kaapstad: Pharos Woordeboeke. Prys: R450.

Rooier meestal mooier

Dat die samestelling van die AWS aan dieselfde prosesmatige uitdagings onderhewig sou kon wees as wat die samestellers van groot, grondliggende religieuse tekste in die gesig staar, is moontlik nie 'n vergesogte analogie nie — dit is weliswaar ook al monnikewerk genoem. Hier word immers in die eerste plek by wyse van sowel beskrywing as voorskrywing reëls saamgevat waarvolgens die taalgemeente, by wyse van spreke, hul talige handel en wandel moet inrig. Die boek kan buite konteks en sonder behoorlike bestudering maklik vir ongewenste politiekery aangewend word. Apokriewe inhoud (byvoorbeeld talige eendagsvlieë of ander idiosinkrasieë) moet uit die teks geweer word en as geheel moet die eindproduk samehangend, realisties, oortuigend en goed gefundeerd aandoen. Dit is 'n bron wat uit die aard van die saak moet poog om so na as moontlik daaraan te kom om alles vir almal te wees in die taal, terwyl hierdie bykans onmoontlike taak boonop moet geskied in 'n taalruimte wat toenemend omstrede is. Die Taalkommissie kry dit egter met die jongste uitgawe reg om vir 'n groter deel van die taalgemeenskap as tevore 'n normeringsbron daar te stel wat toeganklik, gebruikersvriendelik en steeds gesaghebbend is.

In hierdie resensie word kortliks na enkele hoofpunte en voorbeelde verwys ter illustrasie¹, maar dit is uiteraard onmoontlik om die teks werklik grondig te bespreek in hierdie beperkte bestek en hierdie verkenning moet daarom hoogstens as oriëntasie voor verdere eie verdieping gesien word. Die fokus val eweneens doelbewus nie primêr op die reëls nie, buiten waar volledige nuwe reëlhoofstukke bygevoeg is, en word hoofsaaklik vanuit die hoek van die woordelys as vertrekpunt benader. Laastens sal die insluiting van woorde uit die variëteite Kaaps en Oranjerivierafrikaans nie hier uitgesonder word nie, hoewel dit in betekenisvolle mate voorkom, aangesien dit reeds elders genoem is (vergelyk in hierdie verband byvoorbeeld Germishuys (2015)², Hendricks, aangehaal in Odendaal, 2017³, asook Luther, 2017⁴) en as sodanig meer omvattende aparte ondersoek regverdig.

Die 11de uitgawe van die AWS (hierna AWS11) is 'n lywige teks met 457 reëls teenoor die vorige uitgawe se 355. Dit is 742 bladsye in omvang teenoor die vorige uitgawe se 593. Die eerste indruk wat by die gebruiker geskep word, is van 'n groter boek⁵, maar nie 'n swaarder een nie. Opsigtelike inhoudelike toevoegings is die hoofstukke oor leestekens (13) en trappe van vergelyking (16) onder die reëls, die uiters nuttige en toeganklike "Gebruikersgids" wat 'n welkome uitbreiding is vergeleke die "Wenke vir gebruikers" van die vorige uitgawe, asook vier nuwe bylaes waarin omvattende bevestiging van Afrikaans as

internasionaal relevante taal en as wetenskapstaal neerslag vind (hier is onder meer landname, geldeenhede, chemiese elemente en geografiese eiename van die Ooste, alles gestandaardiseer vir die Afrikaanse gebruiker). Die boek het bruikbare kleurblokke en 'n toeganklike makrostruktuur. Die lettertipe en veral die ligblou blokke by reëls val veel minder strak op die oog, en as 'n naslaan-bron is dit 'n aangenamer boek as AWS10 om te gebruik, oftewel sêlfs aangenamer, aangesien AWS10 allerweë en met reg beskou is as die beste tot nog toe in hierdie verband.

Die terminologielys agterin AWS11 verdien voorts spesifieke vermelding. Hier is 26 nuwe definisies toegevoeg en hoewel die nota vooraan, soos in die verlede, dit stel dat hierdie definisies slegs betrekking het op die gebruik van die betrokke term in die AWS self, is dit nietemin vir die gewone gebruiker 'n nuttige algemene hulpbron en vir studente bykans onontbeerlik. Nuwe insluitings wat veral vir gebruikers nuttig behoort te wees, sluit in "affek", "fleksie"/ "infleksie", etlike terme wat met dialoog te doen het, die veel voorkomende nuwe nosie "meerledige eienaam" of "multiwoordeienaam", "parentese" en "selfnoemfunksie", asook verwysings na tipes sinne en tipes voegwoorde. Daar is egter ook insluitings waarvan die definisies moontlik heel voor die hand liggend of selfs minder relevant is, byvoorbeeld "getal", "nommer" en "syfer" onderskeidelik, of die onderskeid tussen 'n "lys" en 'n "reeks". Dit is wel jammer dat daar in die huidige tydsgees, iets wat volgens die teks deur elke teenswoordige uitgawe van die AWS daarin weerspieël word, by die definisie van "Standaardafrikaans" in die terminologielys 'n uitspraak gemaak word oor gebruiksdomeine vir "die gesproke vorm" daarvan, terwyl dit onder "Werkswyse" tereg gestel word dat die AWS "slegs die ortografiese konvensies (spelling en skryfwyse)" bereël. Die eerste sin onder die definisie van "Standaardafrikaans" sou, vir die doeleindes van hierdie terminologielys, voldoende wees. Dit is egter verblydend dat soveel klem by meer as een punt in die teks daarop gelê word dat Standaardafrikaans die dinamiese en lewende produk van uitgebreide en ingewikkelde prosesse is wat deurentyd voortgaan en waarvoor alle variëteite van die taal belangrik is.

Die nuwe bylaes is veral vir taalpraktisyns nuttig en ruim potensieel 'n klomp voormalige onsekerheid, onduidelikheid en gebrek aan eenvormigheid uit die weg. Die nuwe hoofstuk oor die trappe van vergelyking is eweneens 'n welkome formalisering van 'n taalverskynsel wat tot op hede nêrens so samehangend in een band saamgevat is nie. Die gebruik van die woord "meestal" in sekere reëls, sonder dat daar voorbeelde gebied word van gevalle waar die betrokke reël dan nou nié geld nie, is egter in hierdie hoofstuk effe steurend.

Die hoofstuk oor leestekens is ook 'n nuttige nuwigheid, veral die reëls oor hoof- of kleinletters na die dubbelpunt, die reëls oor konomiegebruik, asook dié vir die hantering van uitgeligte en aangehaalde inligting. Vir studentegebruikers en redigeerders van akademiese tekste sal die afdeling oor die gespesialiseerde gebruik van leestekens ook besonder bruikbaar wees. Oor reël 13.84 het hierdie leser egter 'n vraag. Onder taalwerkers is daar lank reeds onsekerheid

oor die gebruik van 'n opeenvolgende uitroepteken en vraagteken aan die einde van een sin en dit is verblydend dat dit nou bereël is. Volgens die nuwe reël kan die uitroepteken direk ná die vraagteken gebruik word om affek in 'n vraagsin aan te dui. Gedagtig daaraan dat die vraagteken volgens reël 13.63 gebruik word om 'n vraag *af te sluit* (my kursivering), moet daar egter gevra word of reël 13.84 dit nie eerder moes stel dat die uitroepteken eerste kom en die vraagteken laaste wanneer die twee tesame gebruik word nie (d.w.s. "!" eerder as "?!").

By die afdeling oor Omgangsafrikaans is nie wesentlike veranderinge aan die riglyne aangebring nie, buiten dat daar in terme van bewoording voorsiening gemaak is daarvoor dat Omgangsafrikaanse woorde nie net aan Engels nie, maar ook aan ander tale, ontleen kan word. Die leser kry egter (steeds) voorbeelde slegs uit Engels. AWS10 se "element van humor, spot, kleinering of platheid" in verwysing na verafrikaansings soos "baisiekel" word in die nuutste AWS 'n "geurtjie van veral humor, spot, geringskatting of platheid". Regdeur die teks kom soortgelyke idiomatiese belyning of vernuwing voor, soos geïllustreer met die sinvolle woordkeuse "geringskatting" in plaas van "kleinering", hoewel daar ook soms vreemde en oënskynlik idiosinkratiese woordkeuses in dieselfde kader voorkom, soos "geurtjie" in plaas van die meer neutrale en formele "element". 'n Sinvolle toekomstige insluiting in hierdie afdeling sou die spelling van Afrikaanse vloek- of taboeoorde, of minstens riglyne daarvoor, wees. Hoewel die gebruiker begrip het vir oorwegings van fatsoenlikheid en sensitiwiteit, en met dank kennis neem van die reëls wat wys hoe om byvoorbeeld die asterisk te gebruik om bepaalde letters in hierdie soort woorde te versteek, is die realiteit dat taalpraktisyns gereeld op verskillende forums probleme met die spelling en skryfwyse van sekere algemene vloek- en taboeoorde vermeld. Hierdie woorde kom byvoorbeeld wel in gepubliseerde literêre werk voor, asook in ander gesaghebbende taalhulpbronne, en probleme met hul spelling of skryfwyse val direk binne die Taalkommissie se mandaat.

In die woordelys self sluit AWS11 nagenoeg 2000 nuwe lemmas in. Weens lengtebeperkings word hier slegs op 'n seleksie uit hierdie nuwe insluitings gewys ter illustrasie van bepaalde neigings. Die seleksie kom ook slegs uit die lukraak verkose letters "A" en "W", ter wille van bondigheid, hoewel verdere voorbeelde in bykans elke letter se geval voorkom.

Die insluiting van 'n woord soos "Afrikaanssintaksis" skep die indruk dat een van die potensiële teikengebruikers vir die AWS die redelik gevorderde taalstudent is. By die woord "agroëkonomies" is 'n addisionele reëlverwysing toegevoeg wat verwys na die gebruik van die deelteken om uitspraakverwarring te voorkom; soortgelyk word by "weimaraner" nou nie net kruisverwys na die reël oor vreemde woorde nie, maar ook na reël 9.7 (d) oor hoof- of kleinlettergebruik by die name van o.a. diererasse. Hierdie uitbreiding van reëlkruisverwysings by woorde in die woordelys is 'n welkome toevoeging, soos ook die aanvanklike toevoeging van hierdie verwysings in die vorige uitgawe. Die insluiting van lemmas soos "Amsterdam-Johannesburg/Johannesburg-Amster-

dam-vlug *of* Amsterdam-Johannesburg/Johannesburg-Amsterdamvlug" en "wit leuen *of* witleuen *of* wit leuentjie *of* witleuentjie" is waarskynlik nie soseer voorbeelde van insluiting ter wille van erkenning nie, maar eerder van komplekse woorde waaruit die gebruiker dan ook ander soortgelyke komplekse se spelling kan aflei. Eersgenoemde geval verwys ook terug na die nuwe hoofstuk oor leestekens en is dus waarskynlik om daardie rede ook in die woordelys opgeneem. Laastens is die skrapping van die onsensitiewe "witmens" en "witman" met behoud van slegs "wit mens" in die woordelys besonder verblydend; so ook die gepaardgaande kruisverwysing na die reël wat dit stel dat die kleuradjektiewe los geskryf word. (Dieselfde patroon geld in die gevalle "swartmens"/"swartman" vs. "swart mens" en "bruinmens"/"bruinman" vs. "bruin mens".) Hieragter sit meer as bloot 'n taalkundige gegewe — hier stel die Taalkommissie dit ook baie duidelik dat velkleur in Afrikaans nie meer 'n tipe mens onderskei nie, maar bloot op een van die betrokke mens se arbitrêre attribute dui. Dit is 'n lofwaardige standpunt om in te neem, soos ook met die insluiting en erkenning van die woord "selfdood".

Etlike begrippe wat verwys na nuwe ontwikkelinge in veral die wetenskap word ook ingesluit, byvoorbeeld die eienaamvorm "Antroposeen, -e" plus afgeleide adjektief "Antroposeens, -e". Sover dit nuwe lemmas uit gebruiksvariëteite betref, is die woord "antie" byvoorbeeld ingesluit (terloops, dit geld ook "ant *of* tant" plus die aanwysing dat dié vorme voor eiename gebruik word). Volgens die beginsel dat Standaardafrikaans uit alle variëteite gevoed word en sekere woorde ter wille van erkenning insluit, is dit 'n sinvolle keuse, hoewel daar bevraagteken kan word of dit werklik 'n voorbeeld van formele, geskrewe Standaardafrikaans is of sal word.

Buiten die toevoeging van lemmas as sodanig, word daar in sommige gevalle ook betekenisleiding toegevoeg waar dit voorheen nie gestaan het nie, of minder uitgebreid was. Vergelyk byvoorbeeld die lemma "angora" in AWS11:

angora, -s *of* angorabok, -ke 1.11, 9.7 (d), 20.5
angora (*soort dier*), -s 1.11, 9.7 (d)
angora (*soort tekstielstof*) 1.11
angoraooi, -e 9.7 (d), 12.1 (a)

Daarteenoor staan in die vorige uitgawe slegs die volgende:

angora, -s 1.11
angorabok, -ke 18.5
angoraooi, -e

In heelparty gevalle is die toegevoegde leiding wel nuttig, maar daar is ook gevalle soos die bostaande waar die leser wonder of dit werklik nodig is om soveel addisionele betekenisleiding te verskaf waar daar eintlik geen (nuwe) probleem met *spelling of skryfwyse* voorkom nie en waar die betrokke voorbeelde in elk geval almal tot dieselfde woordsoort behoort en dus logieserwys

aan dieselfde spelreël onderhewig is. Bosman et al. (2017⁶) het oor AWS10 daarop gewys dat die aanbod van mikrostrukturele inligting by lemmas miskien met meer sorg bedink kan word en dit blyk ook die geval te wees by sekere elemente in AWS11. Die gebruiker wonder in hierdie verband ook oor lemmas soos "wit suiker" en "witsuiker" wat skynbaar as wisselvorme opgeneem is, sonder betekenisleiding of kruisverwysing na 'n reël, en oor iets soos "wel" wat net so uit AWS10 oorgebring is tesame met die etiket (*bw.*) — lewer hierdie spesifieke vorm werklik betekenisvolle spelling- of skryfprobleme op?

Lemmas wat nuutskeppings, leenwoorde en/of verwysings na tegnologiese ontwikkelinge omvat, verdien aparte vermelding. In hierdie verband is daar heelparty uiters sinvolle insluitings waarvan gebruikers met dank kennis behoort te neem. Voorbeelde sluit in die nou reeds gebruiklike "hommeltuig", die werkwoord- en naamwoordvorme van "whatsapp" (maar nie ander moontlikhede soos "instagram" of "snapchat" nie, hoewel hul spelling dienooreenkomsdig afgelei sou kon word), "selfie", "Wi-Fi/"wi-fi", "meem" en "venue". Aan die ander kant is daar ongelukkig ook die ortografies skrikwekkende "googleloer", die nou helaas verewigde vertaalfout "gladdejantjie" (vir "smoothie", wat gelukkig self ook ingesluit is) en die semanties uiters twyfelagtige dog kennelik veel voorkomende "potgooi" (hier wonder die gebruiker telkens of dit vir die Taalkommissie moontlik sou wees om die taalgemeenskap teen sigself se soms absurde maaksels te beskerm, hoewel daar seker min is wat gedoen kan word as gebruikers sulke onsinnighede algemeen genoeg gebruik).

Spelling in die woordelys word ook in etlike gevalle verander om met veranderinge aan reëls te strook. Hiervan is die geval "anglis" uit AWS10, nou "Anglis" in AWS11, 'n voorbeeld wat saamhang met die redelik ingrypende en meestal sinvolle herbereëling van eiename om voorsiening te maak vir die hantering van onder meer die sogenaamde "multiwoordeienaam". Reël 9.19 in AWS10 het dit gestel dat hoof- of kleinletters gebruik kan word by byvoeglike naamwoorde en selfstandige naamwoorde wat uit eiename afgelei word, terwyl daar in AWS11 by die reëls oor eiename en afleidings 'n uitsondering ingeskryf word vir afleidings uit "multiwoordeienaam, saamgestelde eiename wat met 'n koppelteken vas geskryf word, geografiese eiename en taalbenamings". Die nuwe reël 9.19 stel dit dan vervolgens dat die hoofletter(s) van die basis in sodanige gevalle behoue bly, terwyl die nuwe reël 9.18 die res van die ou reël 9.19 inkorporeer. As gevolg van, of minstens tesame met, hierdie taalkundig verklaarbare en verstaanbare herbereëling, word die ou reël 9.20 ("Afleidings op **-s**, **-isme**, **-istiek** en **-isties** wat uit eiename gevorm is, kan met 'n hoofletter of 'n kleinletter begin.") geskrap. In die plek daarvan staan egter die volgende opmerking wat in 'n mate die netjieser aanpak kelder: "'n Gebruik het in (veral) die taalkunde ontstaan om sommige afleidings van taalbenamings met kleinletters te skryf; dit moenie as foutief beskou word nie, byvoorbeeld **anglisme** (*naas Anglisme*)", ensovoorts. Daar sou gevra kon word waarom daar nie beter daarvoor voorsiening gemaak is in 'n reël wat in elk geval wysiging ondergaan het as dit in die taalkunde gebruiklik is nie? Dit sou moontlik meer wenslik wees as die huidige teenstrydigheid wat daaruit voortspruit dat iets

soos "Suid-Afrikanisme" volgens die reël met 'n hoofletter gespel moet word, terwyl byvoorbeeld "neerlandisme"/"nederlandisme" (naas "Neerlandisme"/"Nederlandisme") volgens die betrokke opmerking ook nie as verkeerd beskou moet word nie. Die hele gedeelte oor taalbenaminge en hul afleidings moet dalk in die volgende uitgawe grondig herbekyk word. Nietemin is die insluiting van byvoorbeeld die religiebenaming "Wicca" en die afgeleide naamwoorde "Wiccan"/"wiccan", met kruisverwysing na dieselfde 9.18, weer 'n sinvolle voorbeeld van sowel erkenning (van die betrokke groepering), paradigmatische volledigheid en die moontlikheid van afleidings uit ander eiename op 'n soortgelyke patroon. Melding moet ook gemaak word van die insluiting van heelparty woorde uit die Islam-leefwêreld, byvoorbeeld in die vorm van "sjaria of sjari'a", en ook uit die leefwêreld van verskeie ander geloofs- en/of kultuurgroepe regdeur AWS11.

Die bostaande voorbeelde en oorsig dien slegs as inleiding tot wat AWS11 aan die Afrikaanse taalgemeenskap bied. Dit kan vir alle taalgebruikers, maar veral vir diegene wat gretig is om (dit wat hulle dink) Standaardafrikaans (is) te verguis, lonend wees om hierdie uitsonderlike taalhulpbron self in die hand te neem, dit grondig te bestudeer en werklik bewus te raak van wat dit is en hoe 'n formidabele stuk werk ten grondslag daarvan lê — dit is die nuutste vergestaltung van 'n moderne, omvattende, dinamiese, gestandaardiseerde taalvariëteit wat as hulpbron aan alle lede van die taalgemeenskap beskikbaar is.

Eindnotas

1. Sekere gegewens en getalle word met dank aangehaal uit dr. Frikkie Lombard se aanbieding *Maak kennis met AWS*¹¹. Besikbaar by: <http://www.litnet.co.za/die-aws-100-jaar-middaguur-gesprek-video/>.
2. Germishuys, G. 2015. 'Skanghagha' nou ook in AWS. *Volksblad*, 20 April.
3. Odendaal, G. 2017. LitNet Akademies-resensie-essay: *Kaaps in fokus*. Besikbaar by: <http://www.litnet.co.za/litnet-akademies-resensie-essay-kaaps-fokus/>.
4. Luther, J. 2017. *Nee, my ou suikerpot, die lewe is nie net skanghagha nie*. Besikbaar by: <https://viva-afrikaans.org/lees-luister/blog/item/290-sabela>.
5. Vir die doeleindes van hierdie resensie word slegs van die gedrukte weergawe gebruik gemaak. Die bron is ook elektronies verkrygbaar, en die woordelys is ook apart elektronies beskikbaar via die Virtuele Instituut vir Afrikaans se databasis.
6. Bosman, N., E. Taljard en D. Prinsloo. 2017. Honderd jaar *Afrikaanse Woordelys en Spelreëls* — 'n Oorsig en waardering. Deel 1: Die woordelys. *Tydskrif vir Geesteswetenskappe* 57(2-1): 285-301.

Marius Swart
Departement Afrikaans en Nederlands
Universiteit Stellenbosch
Stellenbosch
(mariusswart@sun.ac.za)

Publikasieaankondigings / Publication Announcements

Monika Bielińska und/and Stefan J. Schierholz (Herausgeber/Editors). *Wörterbuchkritik — Dictionary Criticism*. 2017, vi + 409 pp. ISBN 978-3-11-054473-2, ISSN 0175-9264. Lexicographica. Series Maior 152. Berlin: Walter de Gruyter. <https://www.degruyter.com/view/product/488665>. Preis/Price: €99,95.

Philip Durkin. *The Oxford Handbook of Lexicography*. 2016, xxiii + 698 pp. ISBN: 978-0-19-969163-0. Oxford: Oxford University Press. Price £95.00. (Review in this issue.)

Ekaterina V. Rakhilina, Jean-Marie Merle et Irina Kor Chahine. *Verba sonandi. Représentation linguistique des cris d'animaux*. 2017, 342 pp. ISBN 979-10-320-0102-8, ISSN 1158-629X. Langues et Langage 28. Aix-en-Provence: Presses Universitaires de Provence. <https://presses-universitaires.univ-amu.fr/verba-sonandi-1>. Prix: €29,00.

Stefan J. Schierholz, Rufus Hjalmar Gouws, Zita Hollós und Werner Wolski (Herausgeber). *Wörterbuchforschung und Lexikographie*. 2016, xii + 331 pp. ISBN 978-3-11-047219-6, ISSN 0175-9264. Lexicographica. Series Maior 151. Berlin: Walter de Gruyter. <https://www.degruyter.com/view/product/469617>. Price: €119,95.

Die Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. *Afrikaanse Woordelys en Spelreëls*. Elfde, verbeterde en omvattend herbewerkte uitgawe. 2017, xxxiii + 757 pp. ISBN: 978-1-86890-201-1 (Sagteband). Kaapstad: Pharos Woordeboeke. Prys: R450,00. (Resensie in hierdie nommer.)

Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer und/and Werner Wolski (Herausgeber und Bearbeiter/Editors and Compilers) unter Mitarbeit von/with the Collaboration of Ekaterina Butina-Koller (Russisch/Russian), Rute Costa (Portugiesisch/Portuguese), Milka Enčeva (Bulgarisch/Bulgarian), Ma Theresa Fuentes Morán (Spanisch/Spanish), Laura Giacomini (Italienisch/Italian), Rufus H. Gouws (Afrikaans; Englisch/English), Franz Josef Hausmann und/and Maria Hegner (Französisch/French), Regina Hessky und/and Zita Hollós (Ungarisch/Hungarian), Pavel Petkov (Bulgarisch/Bulgarian), Giovanni Rovere (Italienisch/Italian), Stefan J. Schierholz (Portugiesisch/Portuguese) und/and Pavlina Zlateva (Bulgarisch/Bulgarian). *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary*

Research. Mit englischen Übersetzungen der Umtexthe und Definitionen sowie Äquivalenten in neun Sprachen / With English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages. 2. Bd.: / Vol. 2: 2017, 942 pp. ISBN 978-3-11-034094-5. Berlin/New York: Walter de Gruyter. Preis/Price €399,00.

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://lexikos.journals.ac.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 geskifte 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 rekenaarskryf 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 enigen 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 geïntermeer 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://lexikos.journals.ac.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**.