

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/10.5788/29-1-1527>

Lexikos 29

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/10.5788/29-1-1527>

Lexikos 29

Redakteurs

Editors

Danie Prinsloo

Dion Nkomo

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 29:2019

AFRILEX SERIES 29:2019



BURO VAN DIE WAT

STELLENBOSCH

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2019 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2019

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

ISBN 978-0-9946528-7-4
ISSN 2224-0039

Hierdie werk is gelisensieer ingevolge 'n Creative Commons License CC BY 4.0-lisensie.

Licensed under Creative Commons License CC BY 4.0.

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; SciELO SA; 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. 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)

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)
Prof. 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	ix
Foreword	x
<i>Danie Prinsloo and Dion Nkomo</i>	
'n Woord van AFRILEX	xi
A Few Words from AFRILEX	xii
<i>Herman L. Beyer</i>	
Redaksionele doelstellings	xiii
Editorial Objectives	xiv

Artikels / Articles

Die lexikografiese Behandling van Neologismen aus der Perspektive hispanophoner DaF-Lernender	1
<i>Vanessa González Ribao</i>	
Lexical Understanding of Native Bahasa Indonesia Speakers through Word Association to Improve Dictionary Definition	36
<i>Ami Pramesti Jewalani, Myrna Laksman-Huntley and Harwintha Yuhria Anjarningsih</i>	
Use of Hedges in Definitions: Out of Necessity or Theory-Driven?	55
<i>Mariusz Piotr Kamiński</i>	
The Lexicographic and Lexicological Aspects of a Web-Based Chrestomathy of Gothic and Anglo-Saxon Written Records	75
<i>Tinatín Margalitadze and George Meladze</i>	
Theoretical and Practical Reflections on Specialized Lexicography in African Languages	96
<i>Dion Nkomo</i>	

Die behandeling van die funksie dekodering in verskillende tipes woordeboeke <i>Anna Nel Otto en Jadé Blume</i>	125
Exploring the Properties of English Lexical Affixes by Exploiting the Resources of English General-Purpose Dictionaries <i>Torko Prčić</i>	151
New-line and Run-on Guiding Devices in Print Monolingual Dictionaries for Learners of English <i>Bartosz Ptasznik and Robert Lew</i>	180
African Language Dictionaries for Children — A Neglected Genre <i>Elsabé Taljard and Danie Prinsloo</i>	199
Connecting the Dots: Tradition and Disruption in Lexicography <i>Sven Tarp</i>	224
Lexicographical Contextualization and Personalization: A New Perspective <i>Sven Tarp and Rufus Gouwus</i>	250
Focus on the User: Front Matter in Slovenian Dictionaries <i>Marjeta Vrbinc and Alenka Vrbinc</i>	269

Projekte / Projects

Monitoring Academic Studies of Turkish Lexicography: A Bibliometric Study of 84 Years <i>Ferdi Bozkurt</i>	288
--	-----

Leksikonotas / Lexiconotes

La Création du Néologisme Français dans le Dictionnaire Général Bilingue Arabe–Français Moderne <i>Ibada Hilal</i>	316
--	-----

Reflections on the Making of the <i>Grand dictionnaire chinois–français contemporain</i> <i>Jianhua Huang and Hai Xu</i>	324
---	-----

Resensies / Reviews

Roderick McConchie and Jukka Tyrkkö (Editors). <i>Historical Dictionaries in Their Paratextual Context</i> <i>Miroslawa Podhajecka</i>	339
---	-----

Publikasieaankondigings / Publication Announcements

Voorskrifte aan Skrywers	349
Instructions to Authors	350

Voorwoord

Lexikos 29 is 'n gesamentlike poging van Danie Prinsloo en Dion Nkomo as mederedakteurs, doeltreffend bygestaan deur die Buro van die Woordeboek van die Afrikaanse Taal soos verteenwoordig deur Tanja Harteveld en Hermien van der Westhuizen. Ons het die tradisie van streng, dubbele anonieme keuring van artikels voortgesit en wil graag ons keurders hartlik bedank vir al die tyd wat hulle aan die keuringsproses gewy het en vir die toegewyde wyse waarop hulle die taak verrig het om te verseker dat *Lexikos* 'n joernaal van hoogstaande akademiese aard bly.

As egte internasionale joernaal, bevat hierdie jaar se uitgawe bydraes uit China, Denemarke, Duitsland, Georgië, Indonesië, Jordanië, Pole, Serwië, Slowenië en Turkye. Buiten artikels in Engels, is daar ook bydraes in Afrikaans, Duits en Frans. Met vier artikels vanuit Afrika in hierdie nommer, word die leksikografiese gesprek tussen Afrika en Europa voortgesit. Ons is dankbaar vir die voortgesette bydraes van die outeurs wat die sukses van *Lexikos* help verseker.

Ons stelsel van 'n span van roterende redakteurs sal in 2020 voortgesit word. Steve Ndinga-Koumba-Binza sal die redakteur in 2020 wees en Elsabé Taljard in 2021. Hulle sien uit daarna om bydraes vir die volgende uitgawes te ontvang.

Danie Prinsloo
Dion Nkomo
Redakteurs

Foreword

Lexikos 29 is a joint effort by Danie Prinsloo and Dion Nkomo as co-editors, strongly supported by the Bureau of the Woordeboek van die Afrikaanse Taal represented by Tanja Harteveld and Hermien van der Westhuizen. We continued the tradition of strict, double blind peer reviewing of articles and wish to sincerely thank all our reviewers for all the time they devoted to the review process and the sincere way that they put their minds to the task to ensure that *Lexikos* remains a journal of high academic quality.

Being a true international journal, this year's edition contains contributions from China, Denmark, Georgia, Germany, Indonesia, Jordan, Poland, Serbia, Slovenia and Turkey. In addition to articles in English, there are also contributions in Afrikaans, French and German. With four articles from Africa in this issue, the lexicographical dialogue between Africa and Europe is continued. We are grateful for the continued contributions by the authors which help ensure the success of *Lexikos*.

Our system of a team of rotating editors will be continued in 2020. Steve Ndinga-Koumba-Binza will be the editor for 2020 and Elsabé Taljard in 2021. They are looking forward to receiving contributions for the next issues.

Danie Prinsloo
Dion Nkomo
Editors

'n Woord van AFRILEX

Die redaksie van hierdie nommer van *Lexikos* is behartig deur een van die African Association for Lexicography (AFRILEX) se erelede, naamlik prof. Danie Prinsloo van die Universiteit van Pretoria. Hy is in sy taak bygestaan deur prof. Dion Nkomo van Rhodes Universiteit as mederedakteur (en toevallig ook AFRILEX se Sekretaris). My hartlike dank aan hierdie twee gewaardeerde kollegas vir hulle harde en dikwels ondankbare werk namens en vir die leksikografie en AFRILEX.

Die Buro van die WAT as uitgewer van *Lexikos* moet weer eens van harte bedank word vir sy voortgesette toewyding tot die metaleksikografiese diskoers. Me. Tanja Harteveld (resensieredakteur) en me. Hermien van der Westhuizen (tegniese ondersteuning) het soos gewoonlik die Buro se reputasie vir professionaliteit en kundigheid deeglik gestand gedoen.

Dit is my voorreg om namens die Raad en lede van AFRILEX die redaksionele span, die Buro van die WAT en bydraende outeurs van harte te bedank vir nommer 29 van *Lexikos*.

Herman L. Beyer
President: AFRILEX

A Few Words from AFRILEX

The editorship of this volume of *Lexikos* was managed by one of the honorary members of the African Association for Lexicography (AFRILEX), namely Prof. Danie Prinsloo from the University of Pretoria. He was assisted in his task by Prof. Dion Nkomo as associate editor (and co-incidentally also AFRILEX Secretary). My sincerest gratitude goes to these esteemed colleagues for their hard and often thankless work on behalf of and for lexicography and AFRILEX.

The Bureau of the WAT, as publisher of *Lexikos*, should be heartily thanked again for its unwavering dedication to the metalexigraphic discourse. As usual, Ms Tanja Harteveld (review editor) and Ms Hermien van der Westhuizen (technical support) thoroughly upheld the reputation of professionalism and expertise for which the Bureau is known.

It is my privilege to, on behalf of the Board and members of AFRILEX, sincerely thank the editorial team, the Bureau of the WAT and contributing authors for volume 29 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 uitgawe 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.

Die lexikografische Behandlung von Neologismen aus der Perspektive hispanophoner DaF-Lernender

Vanessa González Ribao, *Leibniz-Institut für Deutsche Sprache,
Mannheim, Deutschland (ribao@ids-mannheim.de)*

Zusammenfassung: Anhand von einigen medialen Kommunikationsverben wie *mailen* oder *twittern* wird das lexikografische Informationsangebot zu Neologismen auf seine Adäquatheit für die fremdsprachige Produktion untersucht. Die Untersuchung erfolgt aus der Perspektive eines spanischsprachigen DaF-Lernenden. Zur Analyse werden sowohl Neologismenwörterbücher und -datenbanken für das Deutsche als auch gängige, bilinguale Online-Wörterbücher für das Sprachenpaar Spanisch-Deutsch gezogen. Die Ergebnisse der lexikografischen Untersuchung werden exemplarisch mit korpusbasierten Daten aus einer Doktorarbeit verglichen. Die Befunde zeigen den Bedarf und die Notwendigkeit auf, die lexikografische Behandlung von (verbalen) Neologismen im spanisch-deutschen Kontext zu optimieren. Dabei soll — insbesondere — die fremdsprachige Textproduktion berücksichtigt werden.

Stichwörter: KOMMUNIKATIONSVERBEN, NEOLOGISMENLEXIKOGRAFIE, ONLINE-LEXIKOGRAFIE, BILINGUALE WÖRTERBÜCHER, SPANISCH-DEUTSCH, FREMDSPRACHIGE TEXTPRODUKTION, SPANISCHSPRACHIGE DAF-LERNENDER

Abstract: The Lexicographic Treatment of Neologisms from the Perspective of Hispanophone Learners of GFL. On the basis of some instrumental verbs of communication such as *mailen* (to mail) or *twittern* (to tweet) we examined the lexicological information offered on neologisms for its adequacy for text production in a foreign language. The study is carried out from the perspective of a Hispanophone learner of GFL. We analyzed neologism dictionaries and databases for German as well as bilingual online dictionaries for Spanish-German. We compared the results of the lexicographic investigation with corpus-based data from a doctoral thesis. The findings point out the need and necessity of improving the treatment of (verbal) neologisms in Spanish-German dictionaries. When doing so the text production in a foreign language should be particularly considered.

Keywords: VERBS OF COMMUNICATION, NEOLOGISM LEXICOGRAPHY, ONLINE LEXICOGRAPHY, BILINGUAL DICTIONARIES, SPANISH-GERMAN, TEXT PRODUCTION IN A FOREIGN LANGUAGE, HISPANOPHONE LEARNERS OF GFL

1. Einleitung

Die vorliegende lexikografische Untersuchung findet im Rahmen einer Dissertationsarbeit zum Thema Argumentstruktur (AS) und Bedeutung medialer Kommunikationsverben (MKV) des Deutschen und des Spanischen im Sprachvergleich statt (González Ribao 2019). Als MKV werden diejenigen Kommunikationsverben bezeichnet, die sich auf Sprachhandlungen beziehen, in denen die Kommunikation zwischen Sprecher und Hörer mithilfe eines technologischen Mediums (d.h. eines technischen Geräts bzw. über das Internet mithilfe einer Anwendungssoftware oder Webseite) erfolgt. Die in dieser Studie analysierten MKV werden in der Abbildung 1 aufgeführt. Ausgangspunkt der lexikografischen Untersuchung sind einschlägige lexikografische Studien für das betreffende Sprachenpaar, die gezeigt haben, dass v.a. bei Produktionssituationen mehr syntagmatische Daten als die in den herkömmlichen Wörterbüchern angeboten werden, notwendig sind (vgl. Meliss 2013, 2015a, 2015b, 2016). Da die Gruppe der MKV sowohl im Deutschen als auch im Spanischen zum größeren Teil aus Neologismen besteht (siehe Abb. 1), stehen diese neuen Verben im Mittelpunkt der Untersuchung. Das Ziel ist zu überprüfen, ob hispanophone DaF-lernende die notwendigen Informationen zu Neologismen für die Textproduktion im Deutschen finden. Daher stehen im Fokus der Untersuchung sowohl die deutschsprachige Neologismenlexikografie (vgl. Lemnitzer 2010) als auch die bilinguale Lernerlexikografie im spanisch-deutschen Kontext. Der Begriff „Lernerlexikografie“ ist relativ umstritten (vgl. Tarp 2008: 125-130, Engelberg und Lemnitzer 2009: 27-29). Ein Lernerwörterbuch wird hier als ein Wörterbuch definiert, das auf die Bedürfnisse eines Benutzers (vgl. Tarp 2013b: 463-465, Tarp 2008: 56-58) beim Spracherwerb ausgerichtet ist. Zur bilingualen Lernerlexikografie gehören demnach zweisprachige Wörterbücher, die für die Hilfestellung beim Erwerbsprozess einer Fremdsprache konzipiert sind (vgl. Tarp 2013a: 425-426). Als schriftliche und mündliche Produktion in der Fremdsprache zählen normalerweise folgende Situationen: Texterstellung, (Hin-)Übersetzung, Textbe- und überarbeitung sowie Textwiedergabe (vgl. Tarp 2008: 59-68 und 149-161, Engelberg und Lemnitzer 2009: 112-113 und 120-131, Wiegand et al. 2017: 341-342). Die Ergebnisse der vorliegenden Untersuchung zeigen die Notwendigkeit lexikografischer Ressourcen zu Neologismen für fremdsprachige Produktionszwecke aus der Perspektive spanischsprachiger DaF-Lernender auf.

DE	ES
<ul style="list-style-type: none">• telegrafieren• morsen• telefonieren• anrufen• faxen• simsen• funken• übertragen• ausstrahlen• senden• facebooken• twittern• posten• chatten• bloggen• mailen• skypen• whatsappen	<ul style="list-style-type: none">• telegrafiar• telefonar• faxear• radiar• transmitir• retransmitir• emitir• televisar• facebookear• tuitear• postear• chatear• bloguear• mailear• skypear• wasapear

Abb. 1: Die in der Doktorarbeit analysierten MKV des Deutschen und des Spanischen

Der Beitrag ist folgendermaßen aufgebaut: Zunächst wird die angewendete Methodologie kurz präsentiert (siehe Abschnitt 2). Eine Bestandsaufnahme und eine Analyse gängiger lexikografischer Ressourcen zu Neologismen im Deutschen werden in den Abschnitten 3 und 4 durchgeführt. Danach folgt eine Sichtung der Aufführung von Neologismen in einigen ausgewählten bilingualen Wörterbüchern für das betreffende Sprachenpaar Spanisch–Deutsch (siehe Abschnitt 5). Anschließend werden die Befunde der lexikografischen Recherche mit Korpusdaten aus der erwähnten Doktorarbeit exemplarisch verglichen. Dabei wird die Relevanz der Beschreibung syntagmatischer und argumentstruktureller Eigenschaften von Neologismen in dem lexikografischen Informationsangebot aufgezeigt (siehe Abschnitt 6). Zuletzt wird für die Wichtigkeit mehrsprachiger Neologismenlexikografie plädiert, die sich insbesondere auf die freie Produktion und Übersetzung ins Deutsche als Fremdsprache ausgerichtet (siehe Abschnitt 7).

2. Vorgehensweise und Untersuchungsparameter

Im Zentrum der Analyse steht die fremdsprachige Produktionssituation (siehe

Abschnitt 1). Diese Wörterbuchbenutzungssituation (vgl. Wiegand et al. 2010: 84) ist Ausgangspunkt und Bezugsrahmen der Untersuchung.

Für die Untersuchung einsprachiger Ressourcen wird angenommen, dass ein Sprecher des Spanischen als L1 ein neues Wort des Deutschen in einer konkreten kommunikativen Situation verwenden will. Der Benutzer¹ weiß — oder zumindest vermutet — dass dieses Wort existiert. Er benötigt aber weitere Informationen zu seinem Gebrauch. Deswegen sucht er Hilfe in einem einsprachigen Nachschlagewerk. Er will eventuell semantische Informationen zu Distributionsbeschränkungen bekommen. Vor allem erwartet er Informationsangaben zu dem syntagmatischen Kombinationspotenzial und Gebrauchskontext zu finden, die ihm bei der adäquaten Verwendung des Wortes in der Fremdsprache helfen. Nun ergibt sich die Frage, ob er die gewünschten Informationen in den vorliegenden einsprachigen Wörterbüchern findet.

Für die Untersuchung zweisprachiger Wörterbücher wird ebenfalls die bereits geschilderte Benutzungssituation in Betracht gezogen. Dazu wird eine weitere Benutzungssituation berücksichtigt, bei der angenommen wird, dass ein Sprecher des Spanischen als L1 das Konzept MEDIALE KOMMUNIKATION im Deutschen ausdrücken will. Ziel dieses Benutzers ist die freie Produktion in der Fremdsprache bzw. die umgekehrte Übersetzung. Die Suchstrategie, die der Benutzer normalerweise verwendet, beruht auf muttersprachlicher Introspektion: Er kennt die lexikalischen Einheiten, die in seiner Muttersprache das gegebene Konzept versprachlichen und fragt sich, wie bzw. ob man auch „so etwas“ auf Deutsch als L2 sagen kann. Dementsprechend sucht er Hilfe in einem bilingualen Wörterbuch. Er schlägt nach den verbalen Lexemen oder den Kernwörtern von komplexen Einheiten nach, die das Konzept in seiner Muttersprache lexikalisieren. Er will Entsprechungen suchen und finden und erwartet u.a. Informationsangaben zu dem syntagmatischen Kombinationspotenzial der Ziellemmata, die ihm dabei helfen, zunächst die passende Entsprechung auszuwählen und danach diese in der Fremdsprache adäquat zu verwenden. Nun ergibt sich hier ebenfalls die Frage, ob er die gewünschten Informationen findet.

Da die syntagmatischen Informationen in Wörterbüchern üblicherweise in Satzmuster und Beispielsätze einbezogen (bspw. *Jemand chattet mit Jemandem* oder *Peter chattet mit seiner Familie*) oder als morphosyntaktische Hinweise (Transitivität, Präpositionen, Kasusangaben, etc.) und Kontextangaben (zur Verdeutlichung des semantisch-syntaktischen Umfeldes) angeboten werden (vgl. Herbst 2009, Model 2010, Meliss 2015a, 2016, Fuentes Morán 2018), werden folgende Untersuchungsparameter in der Analyse besonders berücksichtigt: (i) Satzbauplan (SBP), (ii) Anzahl der Argumente und (iii) Valenz in syntaktischer sowie semantischer Hinsicht. Zudem wird das Interesse der Analyse zweisprachiger Wörterbücher auch auf (a) den Such-Auffindungsprozess und (b) die Disambiguierung der Ziellemmata gelegt.

3. Bestandsaufnahme lexikografischer Ressourcen zu Neologismen im Deutschen

Unter den lexikografischen Ressourcen zu Neologismen muss man zwischen einsprachigen und zweisprachigen Angeboten unterscheiden.

Das einsprachige Angebot an lexikografischen Ressourcen zu Neologismen für das Deutsche basiert grundsätzlich auf Korpusarbeit (vgl. Engelberg und Lemnitzer 2009: 238-243) und ist namentlich in universitären und sprachwissenschaftlichen Forschungseinrichtungen zu finden. Das bevorzugte Medium ist das Internet. Dort zeigen Neologismendatenbanken eine stärkere Präsenz gegenüber (herkömmlichen) Online-Wörterbüchern. Diese Beobachtung schließt sich an die aktuelle Diskussion zu der verschwommenen Grenze zwischen Datenbanken und Wörterbüchern in der Online-Lexikografie an (vgl. Lew 2011: 246-248, Samaniego Fernández und Pérez Cabello de Alba 2011: 306-309). Die hier untersuchte Auswahl berücksichtigt diejenigen Online-Ressourcen, die bei der Abfrage in Google nach „Neologismenwörterbücher online“ auf der ersten Google-Seite erscheinen, sowie diejenigen Neologismenwörterbücher in Print-Format, die nicht älter als ca. 10 Jahre sind und für gewöhnlich in deutschen Universitätsbibliotheken zur Verfügung stehen². Die untersuchten Ressourcen sind das *Neologismenwörterbuch* des IDS-Mannheim sowie seine zweibändige Printversion *Neuer Wortschatz. Neologismen im Deutschen 2001–2010* (vgl. Steffens und Al-Waldi 2015), die Neologismensammlung *Die Wortwarte* von Lemnitzer aus der Berlin-Brandenburgischen Akademie der Wissenschaften und das *Deutsche Neologismenwörterbuch* von Quasthoff (2007). Sie werden im Teilabschnitt 4 genauer vorgestellt und untersucht.

Bilinguale Neologismenwörterbücher für das Deutsche werden in der Regel im Rahmen eines umfangreichen lexikografischen Forschungsprojekts erstellt. Das IDS-Mannheim hat — zum Beispiel — die Materialien aus dem Projekt *Neuer Wortschatz* für das Sprachenpaar Deutsch–Russisch erweitert. Das *Deutsch–russische Neologismenwörterbuch* von Steffens und Nikitina (2016) — das auch online über das Plattform OWID^{plus} konsultierbar ist, richtet sich primär an russischsprachige DaF-Lernende, die eine Bedeutungserklärung suchen und ihren Wortschatz erweitern möchten. Weitere zweisprachige Neologismenwörterbücher im deutschen Kontext sind das *Deutsch–chinesisches Neologismenwörterbuch* von Dou (2004) und das *Polnisch–deutsches Wörterbuch der Neologismen* von Worbs, Markowski und Meger (2007). Das erste ist für die Rezeption (vgl. Engelberg und Lemnitzer 2009: 118-123) und den Wortschatzaufbau im Deutschen von chinesischsprachigen DaF-Lernenden konzipiert, während die Zielgruppe des zweiten Muttersprachler des Deutschen sind, die Polnisch lernen bzw. studieren. Konkret für das betreffende Sprachenpaar Spanisch–Deutsch liegt das *Compendio temático de neologismos Alemán — Español: deutsche Neubildungen* von Gierden Vega et al. (2010) vor. Es handelt sich um ein Kompendium neuer Wörter des Deutschen, deren Bedeutungserklärung auf Spanisch geschrieben ist. Es werden ggf. die spanischen Äquivalente sowie

Gebrauchsbeispiele angeboten. Die stark ausgeprägte didaktische Funktion des Werks erkennt man an der thematischen Anordnung des Wortschatzes sowie an dem ergänzenden theoretischen Teil zur Wortbildung im Deutschen.

Sämtliche, hier dargestellten, bilingualen Wörterbücher — mit Ausnahme des *Polnisch-deutschen Wörterbuchs der Neologismen* — zielen auf die Erweiterung sozio-kultureller und historisch-politischer Kenntnisse der DaF-Lernenden durch den Wortschatzaufbau ab und haben somit nur diese spezifische Adressatengruppe im Visier. Daraus folgt, dass alle gerade erwähnten Wörterbücher primär auf folgende lexikografische Funktionen (vgl. Tarp 2013b: 465) ausgerichtet sind: kommunikative Funktionen bezüglich fremdsprachiger Rezeption und vor allem kognitive Funktionen (vgl. Tarp 2008: 44-54 bezüglich kommunikativer und kognitiver Situationen und Tarp 2008: 81-88 zu den entsprechenden Funktionen). Das heißt, dass sie dafür konzipiert sind, die Bedürfnisse eines Fremdsprache-Lernenden beim Textverstehen (vgl. Tarp 2008: 147-149) und bei Erwerb bzw. Erweiterung seiner Kenntnisse in einem bestimmten thematischen Bereich (vgl. Tarp 2008: 163-166) zu befriedigen. Deshalb fallen diese Wörterbücher aus der vorliegenden Untersuchung heraus³.

4. Informationsangebot in den aktuellen Neologismenwörterbüchern und Datenbanken für das Deutsche

In Anbetracht der im Abschnitt 2 dargestellten Benutzersituation und der angeführten Parameter sind folgende einsprachige lexikografische Ressourcen zu Neologismen im Deutschen untersucht worden. Für jede Ressource folgen eine Beschreibung und ein Kommentar zu ihrer Adäquatheit für die betreffende Benutzersituation.

4.1 Das *Neologismenwörterbuch*

Das *Neologismenwörterbuch*⁴ ist Bestandteil des wissenschaftlichen lexikografischen Portals OWID des IDS-Mannheim (vgl. Müller-Spitzer: 2010). Es erfasst neue Wörter, die seit den 90er Jahren in das Deutsche eingegangen sind. Als empirische Grundlage dient das virtuelle Projektkorpus „neo“, das vorwiegend Presstexte aus DeReKo enthält, sowie eine selbst erstellte Wortkartei exzerpiertener Print-, Hör- und Internetbelege. Dies wird durch wissenschaftliche Sekundärliteratur angereichert. Die Stichwortauswahl wird nach dreierlei Selektionskriterien getroffen. Die Wort-Kandidaten müssen (i) für den erfassten Zeitraum neu sein, (ii) der Allgemeinsprache angehören und (iii) dem deutschen Sprachgebrauch entsprechen.

Das Wörterbuchnetz OWID (vgl. Engelberg und Müller-Spitzer 2013: 1030-1033) bietet neben der Suchfeld-Option eine vielfältige Auswahl an Suchfunktionen an. Die Suche kann thematisch nach Fach- bzw. Sachgebieten oder chronologisch nach Dekaden eingeschränkt werden. Ebenfalls kann man durch

die verschiedenen alphabetischen Register Zugang zu dem gesuchten Wort bekommen. Folgende Listen werden angeboten: Stichwort, Phraseologismen, Strichlemmata und verdeckte neue Wörter (sie sind keine Stichwörter. Sie sind in den Wortartikeln enthalten). Zudem besteht die Möglichkeit, eine erweiterte Suche zu tätigen. Hierbei kann die Suche durch folgende Filter verfeinert werden: Neologismtyp, Erscheinungsjahrzehnte, Wortart und Wortartmerkmale, Wortbildung und Wortbildungsproduktivität. Die Suche nach bestimmten Teilen bzw. Komponenten des Wortes ist gleichfalls möglich mithilfe der Kommandos „beginnt mit“, „enthält“ und „endet auf“⁵.

In der aktuellen online Fassung des *Neologismenwörterbuchs* (Stand: März 2018) sind alle in der oben erwähnten Doktorarbeit untersuchten, neuen MKV des Deutschen inventarisiert (siehe Abb. 1). Am Beispiel des Verbs *simsen* wird die Struktur eines Artikels beschrieben. Der Artikel besteht aus vier verschiedenen Informationsblöcken: (a) allgemeine Informationen (siehe Abb. 2), (b) Bedeutung und Verwendung (siehe Abb. 3), (c) Grammatik (siehe Abb. 4) und (d) weitere Informationen (siehe Abb. 5). Diese modulare Darstellung der lexikografischen Daten knüpft an den in dem lexikografischen Institut der Universität Aarhus geprägten Begriff *monofunctional dictionary* an (vgl. Bergenholtz 2012: 260). Sie vertreten die Auffassung, dass ein gutes Wörterbuch eine einfache Auffindung und einen schnellen Zugriff auf die gesuchten Daten ermöglichen soll. Dies geschieht, wenn die Daten auf die entsprechenden Benutzerbedürfnisse in einer bestimmten Situation ausgerichtet sind und folglich eine Datenüberflutung vermieden wird (vgl. Bergenholtz 2011). Die im Rahmen der *Functional Theory* gezeigte lexikografische Praxis tendiert zur Erstellung von mehreren verschiedenen funktionsgezielten Wörterbüchern auf der Grundlage einer gemeinsamen Datenbank (vgl. Bergenholtz 2012). Beispiele davon sind die *Danish Music Dictionaries* (vgl. Bergenholtz und Bergenholtz 2011). In derselben Linie entwickelt sich auch die lexikografische Lernplattform *Base Lexicale du Français* (vgl. Verlinde 2011). Poly- oder multifunktionale Wörterbücher müssen dann so transparent und flexibel bei der Darstellung ihrer Daten sein, dass man je nach Situation das passende monofunktionelle Wörterbuch daraus machen kann bzw. die benötigten Daten auswählen kann (vgl. Tarp 2009: 48-61). Bei elektronischen Wörterbüchern heißt das, dass der Benutzer vordefinieren kann, welche lexikografische Daten er (nicht) erhalten soll (vgl. Begriff „Benutzerschnittstelle“ in Wiegand et al. 2010: 679), so wie es bei dem *Neologismenwörterbuch* der Fall ist. Ferner ist das gesamte Portal [OWID] hinsichtlich seiner Benutzerfreundlichkeit (vgl. Wiegand et al. 2010: 677) und Effizienz (~ *Usability*) positiv bewertet (vgl. Heid 2011). Im Folgenden wird jeder Block detailliert dargestellt.

(a) Allgemeine Informationen: Schwerpunkt dieses Blocks sind die Informationen zur Entstehung, Frequenz und Morphologie des Wortes, wie in Abbildung 2 zu sehen ist. Unter dem Stichwort findet man Angaben zu der Lesart und zu der Dekade, in der das Wort aufgekommen ist, sowie ein

Häufigkeitsdiagramm, das die zeitliche Verteilung der Gebrauchshäufigkeiten zu *simsen* ab 1990 darstellt. Danach folgen Informationen zu dem Neologismen-
typ, der Schreibung samt vorhandenen alternativen Schreibweisen und der
Wortbildung.

The screenshot shows the dictionary entry for 'simsen' in the 'Neologismenwörterbuch'. At the top, there is a blue square icon and the word 'simsen' in bold. Below it, the definition 'Lesart: 'schreiben und senden'' is given. To the right, there is a link for 'Benutzerhinweise'. Underneath, a small bar chart shows the frequency of the word over time. The entry is categorized as 'Neologismus der 90er Jahre'. Below the chart, there are sections for 'Neologismen-typ' (Neulexem), 'Formvariante' (smsen, SMSen), 'Schreibung' (Worttrennung: sim|sen), and 'Wortbildung' (Wortbildungsart/-typ: Ableitung (Konversion), Basis: SMS (Nomen)). There are also callout boxes providing pronunciation information: 'Ausspracheangabe zu den Formvarianten: [ɛs|ɛm'ɛsɛn]' and 'unter Einschub eines eufonischen - i'.

Abb. 2: Bildschirmausdruck des Eintrags zu *simsen* im *Neologismenwörterbuch* [22.02.18]

(b) Bedeutung und Verwendung: Dieser Block enthält eine kurze Definition, eine Sammlung typischer Verwendungen und eine Liste sinnverwandter Ausdrücke (siehe Abb. 3). Besonders innovativ ist das Angebot der als „typischer Verwendungen“ etikettierten syntagmatischen Verbindungen, die eine hohe Frequenz nachweisen, ohne notwendigerweise idiomatisch zu sein. Sie veranschaulichen den typischen Gebrauch eines Wortes im syntagmatischen Zusammenhang und liefern vielfältige Informationen zu den lexikogrammatistischen Eigenschaften des Wortes. Beispielsweise lässt sich aus den typischen Verwendungen von *simsen* erschließen, dass es oft durch Adverbien modifiziert wird (siehe *ununterbrochen simsens*, *leidenschaftlich gern simsens*), dass es häufig mit einem satzförmigen Direktobjekt auftritt (siehe *simsens*, *wann die Versammlung stattfindet*) und dass es üblicherweise in einem informellen bzw. familiären

Kontext vorkommt (siehe *dem Freund die Neuigkeiten simsen, mit der Freundin simsen*).

Aufkommen: seit 2000 in Gebrauch

The screenshot shows a dictionary entry for the verb 'simsen'. At the top, there are three tabs: 'Bedeutung und Verwendung' (selected), 'Grammatik', and 'Weitere Informationen'. The main content is organized into several sections:

- Bedeutungsangabe:** A blue vertical bar is followed by the text 'eine SMS mit Hilfe eines PCs oder eines Handys schreiben und senden'.
- Belegblock:** Labeled 'anzeigen »'.
- Typische Verwendungen:** Lists several phrases: 'dem Freund die Neuigkeiten simsen', 'mit der Freundin simsen', 'simsen, wann die Versammlung stattfindet', 'ununterbrochen simsen', and 'leidenschaftlich gern simsen'.
- Sinnverwandte Ausdrücke:** Divided into two sub-sections:
 - Kohyponym(e):** A list of related terms: chatten, emailen, facebooken, mailen, twittern, whatsappen, zwitschern.
 - Hyperonym(e):** A list of superordinate terms: texten.

Abb. 3: Bildschirmausdruck des Eintrags zu *simsen* im *Neologismenwörterbuch* [22.02.18]

(c) **Grammatik:** Dieser Block bietet Informationen zu Wortart, Konjugation, Syntax und Wortbildungsproduktivität an (siehe Abb. 4). Die angebotene Beschreibung der syntaktischen Umgebung durch Phrasenmusterangaben informiert den Benutzer über die Eigenschaften obligatorischer und fakultativer Argumente des gesuchten Wortes, hier *simsen*. Sie zeigt — in abstrakter Form — das Valenzpotential des Verbs *simsen*, das aus den konkreten Realisierungen der Angabe „typische Verwendungen“ nur indirekt zu entnehmen ist (siehe Abb. 3). Beide komplementären Informationsangaben sind unerlässlich für die erfolgreiche Textproduktion in der Fremdsprache.

Aufkommen: seit 2000 in Gebrauch

Bedeutung und Verwendung **Grammatik** Weitere Informationen

Grammatische Angaben

Wortart: **Verb** (schwach)

Konjugation

Präteritum:	simste
Partizip Perfekt:	gesimst
Perfektbildung:	mit haben
Passiv:	bildbar

meist im substantivierten Infinitiv belegt;
finite Verbformen erst nach 2000

Syntaktische Umgebung: jemand simst (jemandem) (etwas)
jemand simst (mit jemandem)
jemand simst, dass ...
jemand simst, ob ...
jemand simst, wer/wie/was/... ...

Wortbildungsproduktivität

Präverbfügung: *ansimsen, zurücksimsen*

Präverbfügungen erst nach 2000

Abb. 4: Bildschirmausdruck des Eintrags zu *simsen* im *Neologismenwörterbuch* [22.02.18]

(d) Weitere Informationen: In diesem Block findet man — normalerweise — weitere Angaben zu den paradigmatischen Informationen des Stichwortes und zu seiner Präsenz in anderen Wörterbüchern (siehe Abb. 5). Diese Informationen sind ebenfalls von großer Relevanz für die Textproduktion. Sie machen den Benutzer auf die Existenz vielfältiger Ausdrucksmöglichkeiten für dasselbe Konzept, das das konsultierte Stichwort lexikalisiert, aufmerksam. Die Darstellung des Stichwortes im Zusammenhang mit bedeutungsverwandten und teiläquivalenten Ausdrücken hilft dem Benutzer sowohl beim Wortschatzaufbau (kognitive Situation) als auch bei der stilistischen Verfeinerung seines Textes in Situationen der Textproduktion.


Aufkommen: seit 2000 in Gebrauch

Bedeutung und Verwendung Grammatik Weitere Informationen

Sprachreflexives
Alternativ zu dem jüngeren Verb *simsen* werden solche typischen Verwendungen mit *SMS* (in der Lesart 'Kurzmitteilung') gebraucht wie: *eine SMS schicken, eine SMS senden, eine SMS schreiben*. Sie sind in den IDS-Textkorpora (derzeit noch) häufiger belegt als *simsen*.

simsen gehörte 2001 zu den von der Gesellschaft für deutsche Sprache in Wiesbaden benannten "Wörtern des Jahres".

Vgl. im Portal OWID im Wörterbuch "Kommunikationsverben" den Wortartikel



Für Tätigkeiten, die sich auf Kommunikation im Internet beziehen, sind seit den 90er Jahren folgende Bezeichnungen aufgekommen und im Neologismenwörterbuch Stichwort geworden: *anmailen, bloggen, chatten, emailen, entfolgen, entfreunden, facebooken, gruscheln, leaken, liken, mailen, posten, retweeten, simsens, skypeen, texten, twittern, voipen, whatsappen, wischen, youtuben, zwitschern*.

Vorkommen in Wörterbüchern

- Duden - Fremdwörterbuch (2001)
- Duden - Wörterbuch der New Economy (2001)

Abb. 5: Bildschirmausdruck des Eintrags zu *simsen* im *Neologismenwörterbuch* [22.02.18]

Zusammenfassend kann man festhalten, dass das *Neologismenwörterbuch* ein lexikografisches Werkzeug ist, das sich sowohl für Muttersprachler als auch für DaF- und DaZ-Lernende zur Rezeption und Produktion deutscher Texte sowie zur Kenntniserweiterung in bestimmten Bereichen eignet, da der Benutzer je nach Bedürfnis und Situation die benötigten Daten selektieren kann.

4.2 Die Wortwarte

Das Projekt *Die Wortwarte* sammelt seit 2000 neue Wörter, die in den Onlineausgaben einiger Zeitungen sowie in den Onlinemagazinen *Perlentaucher* und

heisse.de erschienen sind. Der Auswahlprozess erfolgt in zwei Etappen: Erst werden die Neuwort-Kandidaten mithilfe der in den Projekten des Programmbereichs Korpuslinguistik des IDS-Mannheim angebotenen Werkzeugen maschinell selektiert. Als Referenz für die Vorauswahl dienen die Wortlisten aus DeReKo (DeReWo). Danach sortiert Lemnitzer von Hand alle Ausdrücke, die keine neugebildeten Wörter sind, aus und von den letzteren wählt er nur diejenigen aus, die man nicht in die Kategorie der Okkasionalismen einschließen kann (vgl. Einleitung und Hintergrund in www.wortwarte.de). Die Wortsuche kann entweder alphabetisch, thematisch oder chronologisch eingegrenzt werden. Über ein Suchfeld, in das man das gesuchte Wort eingeben kann, verfügt die Webseite nicht. Folgende Verben aus der Gruppe der neuen MKV sind in der Neologismensammlung enthalten (Stand: März 2018): *bloggen*, *mailen*, *simsen* und *twittern*. Bei der Suche nach den neuen MKV wurde das alphabetische Kriterium eingesetzt. Im Folgenden wird das Informationsangebot der *Wortwarte* anhand des Eintrags zum Verb *mailen* exemplarisch beschrieben (siehe Abb. 6).

mailen, Verb

[Informationstechnologie](#); [bei Google](#) [bei Wikipedia](#) zuletzt 160473mal gesehen (am 19.9.2005)

26 Oktober 2000 17 : 02 Hallo Gerald Meier, da es uns auch wirklich wichtig ist, dass Du uns wieder lieb hast, rödeln unsere Techniker seit ca. einer Woche an Deinem Problem. Wenn Du näheres über den genauen Stand der Dinge wissen willst, dann [maile](#) unseren Webmaster an : technik@meinberlin.de ! Ich drücke die Daumen ! Tschüß - verregnete Grüße Sabine vom meinberlin-Team Homepage : meinberlin.de 5 gerald meier - berlin Donnerstag 26 Oktober 2000 10 : 17 wann kümmert sich endlich jemand um den Zugang zum webkalender ! ?
<http://www.meinberlin.de/tmh/gaestebuch/index.html>

Abb. 6: Eintrag zu *mailen* in der *Wortwarte* [22.02.2018]

Wie in Abbildung 6 zu beobachten ist, besteht ein Eintrag in der *Wortwarte* aus folgenden Teilen:

- Stichwort („mailen“)
- Angabe zur Wortart („Verb“). Bei Substantiven sind an dieser Stelle Informationen zu Genus sowie zu Genitiv- und Pluralbildung zu finden.
- Sachgebiet-Angabe (Informationstechnologie)
- Angabe zu Frequenzanzahl mit Datum („bei Google bei Wikipedia zuletzt 160473mal gesehen (am 19.09.2005)“)
- Belegbeispiel mit Quelle zur Illustration des Wortes im Gebrauchskontext.

Die *Wortwarte* leistet eine empirisch fundierte Arbeit zur Sprachdokumentation, die ermöglicht, die Entwicklung des gegenwärtigen Wortschatzes des Deutschen zu beobachten. Die Neologismensammlung ist jedoch für Textproduktionszwecke wenig nützlich, da u.a. Angaben zu semantischer und syntagmatischer Kombinatorik fehlen. Außerdem kann die angebotene quantitative Information nur bedingt als Referenz genommen werden, da sie auf Online-Quellen beruht und schnell veraltet. Sie hat eher einen impressionistischen Charakter.

4.3 Das Deutsche Neologismenwörterbuch

Das *Deutsche Neologismenwörterbuch* von Quasthoff (2007) steht im Zusammenhang mit dem Projekt *Deutscher Wortschatz* der Abteilung für Automatische Sprachverarbeitung am Institut für Informatik der Universität Leipzig. Es erfasst ein Inventar von 2284 Stichwörtern, die sich durch ihre zunehmende Häufigkeitsschwelle und Mindestanzahl in den Korpora des Projekts im Zeitraum von 2000 bis Ende 2006 im Vergleich zu den Jahren 1995-1999 auszeichnen. Das umfangreiche Korpus besteht aus Zeitungstexten sowie Online-Ausgaben anderer Medien (vgl. Quasthoff 2007: 9). Das konsequent angewendete Häufigkeitskriterium führt zu einer Auffassung des Begriffs *Neologismus* im weitesten Sinne. Dementsprechend beinhaltet das Wörterbuch sowohl neu erschienene Wörter als auch etablierte Wörter, die vor Kurzem — und oft aus extralinguistischen Gründen — in der Alltagssprache auffällig frequent vorkommen (vgl. Quasthoff 2007: 5). Bezüglich der neuen MKV konnte kein Verb der Gruppe in dem Wörterbuch gefunden werden. De facto sind lediglich folgende drei Verben inventarisiert: *abschaufeln*, *durchregieren* und *nachsteuern*. Dennoch ist — im Zusammenhang mit den neuen MKV — das deverbale Substantiv *Bloggen* (neben den Substantiven *Blog*, *Blogger* und *Blogosphäre*) im Wörterbuch registriert. Der Eintrag zu *Bloggen* wird hier benutzt, um die Eintragsstruktur exemplarisch zu beschreiben (siehe Abb. 7).

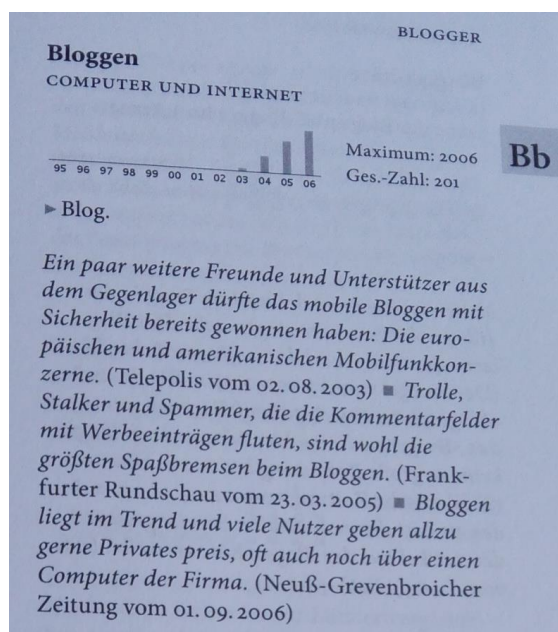


Abb. 7: Eintrag zu *Bloggen* in dem *Deutschen Neologismenwörterbuch* (Quasthoff 2007: 109)

Der in Abbildung 7 illustrierte Eintrag zu *Bloggen* ist ein Verweiseintrag. Unter dem Stichwort als Überschrift findet man Information zum Sachgebiet (Angabe „Computer und Internet“), ein Häufigkeitsdiagramm und drei Belegbeispiele, die den Gebrauch des Wortes in den Medien veranschaulichen. Das Häufigkeitsdiagramm zeigt die Häufigkeit von *Bloggen* für jedes Jahr zwischen 1995 und 2006 im Vergleich. Überdies wird das Jahr der größten Häufigkeit (Angabe „Maximum: 2006“) zusammen mit der Häufigkeit des Wortes im Gesamtzeitraum (Angabe „Ges.-Zahl: 201“) hervorgehoben. Für weitere Informationen wird auf den Eintrag zu *Blog* verwiesen, in dem ein Erläuterungstext zu den Ursachen des auffälligen Wortgebrauchs zu finden ist. Haupteinträge können zudem Informationen zu alternativen Schreibweisen sowie eine kurze Definition enthalten. Dennoch müssen nicht alle Angaben immer vorhanden sein (vgl. Quasthoff 2007: 7).

Das *Deutsche Neologismenwörterbuch* von Quasthoff (2007) dient zum Erwerb und zur Erweiterung des enzyklopädischen Wissens und leistet eine sogfältige, lexikografische Dokumentationsarbeit, indem es den Zeitgeist der späten 90er und früheren Nullerjahre durch die Entwicklung des Wortschatzes widerspiegelt. Als Hilfsmittel für die Textproduktion im Deutschen — zumal als Fremdsprache — wurde das Werk jedoch nicht konzipiert.

5. Neologismen in bilingualen Wörterbüchern für das Sprachenpaar Spanisch–Deutsch

Das Angebot an bilingualen allgemeinsprachlichen Wörterbüchern für das Sprachenpaar Spanisch–Deutsch ist relativ umfangreich (vgl. Domínguez Vázquez 2013, Meliss 2016). Die vorliegende Untersuchung beschränkt sich auf das online Angebot, da verschiedene lexikografische Studien und Umfragen zur Wörterbuchnutzung zeigen, dass sich das Internet als Medium in den letzten Jahren durchgesetzt hat (vgl. Domínguez Vázquez et al. 2013: 145-146, Meliss 2015b: 405-406, Fernández Méndez et al. 2016: 80 sowie Scheller-Boltz und Weinberger 2017: 10). Im Einzelnen werden folgende im Internet angebotene Wörterbücher der Analyse unterzogen: Die beiden Online-Wörterbücher Spanisch ↔ Deutsch aus den lexikografischen Portalen von PONS und LEO und die Online-Version des Printwörterbuches Langenscheidt Spanisch. Die Auswahl lässt sich folgendermaßen begründen. Sowohl die oben zitierten Benutzerumfragen im Spanisch–Deutschen Kontext als auch andere lexikografische Studien europäischer Sprachen im Vergleich bestätigen, dass die Benutzer — bevorzugt und am Häufigsten — die Online-Wörterbücher von PONS und LEO zu Rate ziehen (vgl. Meliss 2015b: 412-413, Fernández Méndez et al. 2016: 80 und Scheller-Boltz und Weinberger 2017: 11). Ergänzend soll die Analyse der Online-Version von Langenscheidt dazu dienen, Wörterbücher mit unterschiedlicher Nutzerbeteiligung (vgl. Wiegand et al. 2010: 90-91, Abel und Meyer 2016) miteinander zu vergleichen. Konkret wird das Informationsangebot der kollaborativen Wörterbücher (vgl. Abel und Meyer 2016: 253-262)

PONS und LEO mit dem von dem rein institutionellen Wörterbuch Langenscheidt, in denen der Benutzer lediglich indirekt beteiligt wird (vgl. Abel und Meyer 2016: 263-272), kontrastiert (vgl. Fuertes-Olivera 2009: 103-107, Lew 2011: 230-231).

Die Untersuchung erfolgt aus der Perspektive eines Benutzers, dessen Zweck die Textproduktion in der Fremdsprache ist. Daher steht das syntagmatische Informationsangebot im Mittelpunkt des Untersuchungsinteresses, wie im Abschnitt 2 erklärt wurde. Die Analyse soll anhand der neuen MKV des Deutschen erfolgen. Sämtliche Verben sind jedoch nicht in allen drei untersuchten Wörterbüchern zu finden. Ein Überblick über die MKV, die in den genannten Wörterbüchern inventarisiert sind, liefert Abbildung 8.

Ausrichtung ES→DE			Ausrichtung DE→ES				
Wörterbücher			Stichwort		Wörterbücher		
de.pons .com	leo.org	Langen scheidt -Online	ES	DE	de.pons .com	leo.org	Langen scheidt -Online
∅ _B	∅	∅	<i>bloguear</i>	<i>bloggen</i>	∅	∅	∅
∅	∅	∅	<i>chatear</i>	<i>chatten</i>	∅	∅	∅
∅	∅	∅	<i>postear</i>	<i>posten</i>	∅ _B	∅	∅
∅ _B	∅	∅	<i>tuítear</i>	<i>twittern</i>	∅ _B	∅	∅
∅	∅	∅	<i>facebookear</i>	<i>facebooken</i>	∅	∅	∅
∅!	∅!	∅!	<i>wasapear</i>	<i>whatsappen</i>	∅	∅	∅
∅	∅	∅	<i>skypear</i>	<i>skypen</i>	∅! _B	∅!	∅
∅	∅	∅	<i>mailear</i>	<i>mailen</i>	∅! _B	∅!	∅!
				<i>simsen</i>	∅!	∅!	∅!

∅ = Kein Eintrag, ∅ = beide Ausrichtungen, ∅! = nur eine Ausrichtung, _B = Benutzereintrag

Abb. 8: Überblick über die MKV, die in PONS, LEO und Langenscheidt-Online inventarisiert sind⁶

Die im Rechteck eingeschlossenen MKV scheinen sich gegenüber den restlichen MKV im Wortschatz etabliert zu haben, da sie in den drei Wörterbüchern aufgeführt werden (siehe Symbol „∅“ in Abb. 8). Einige Einträge verdankt man jedoch dem kollaborativen Benutzer und nicht der Wörterbuchredaktion, wie der tiefgestellte Buchstabe B in Abbildung 8 zeigt⁷. Das Verb *simsen* gehört auch zu den konventionalisierten Neologismen. Es unterscheidet sich von den anderen Verben darin, dass sich sein potenzielles Äquivalent *esemesear* in dem spanischen Wortschatz nicht festgesetzt hat. Aus diesem Grund bleibt es aus der vorliegenden Studie ausgeschlossen, da es in keinem Wörterbuch des Spanischen zu finden ist. Die Verben *skypen* und *mailen* sind ebenfalls in dem deutschen Wortschatz integriert, denn sie besitzen einen Eintrag im Wörterbuch. Das spanische Verb *mailear* hingegen ist in keinem der untersuchten Wörterbücher registriert, während das Verb *skypear* nur im Wörterbuch Langenscheidt-Online zu finden ist. Das umgekehrte Szenario stellt sich bei

dem Sprachenpaar *whatsappen-whastappear* bzw. *wasapear* ein. Die spanischen Benutzer haben die an das orthografische Sprachsystem des Spanischen angepasste Form *wasapear* anerkannt und sie in das Wörterbuch eingetragen. Im Gegensatz dazu haben die deutsche Entsprechung zu *whatsappen* und die Verben *skypen* und *skypear* noch keinen Platz in den betreffenden Wörterbüchern gefunden. Diejenigen MKV, die nur für eine Sprache, d.h. entweder für das Spanische oder das Deutsche, abrufbar sind, tragen die Kennzeichnung „!“ in Abbildung 8.

Die Befunde zu den obigen Einträgen werden im Teilabschnitt 5.1 ausführlich präsentiert. Im Teilabschnitt 5.2 werden die entsprechenden Schlussfolgerungen gezogen.

5.1 Informationsangebot zu Neologismen in bilingualen Wörterbüchern

In Anbetracht der im Abschnitt 2 dargestellten Benutzersituation ist das Informationsangebot zu Neologismen in den drei erwähnten zweisprachigen Wörterbüchern für das Sprachenpaar Spanisch–Deutsch untersucht worden. Anhand ausgewählter Wörterbuchartikel wird die Nützlichkeit dieses Angebotes bei (a) der Produktion in der Fremdsprache, (b) dem Such-Auffindungsprozess und (c) der Disambiguierung der Ziellemmata gezeigt.

5.1.1 Neologismen in PONS-Online-Portal

Das über das PONS-Online-Portal verfügbare bilinguale Wörterbuch Spanisch ↔ Deutsch ist — im Prinzip — auf die Produktions- und Rezeptionsbedürfnisse beider Benutzergruppen ausgerichtet. Es hat zwar das Printwörterbuch als Ausgangspunkt, ist aber als Online-Wörterbuch neu konzipiert worden (vgl. Wiegand et al. 2010: 78-82, Engelberg und Storrer 2016: 34-35). Es wird grundsätzlich redaktionell betreut, bietet jedoch dem Benutzer die Möglichkeit, sich direkt und indirekt zu beteiligen (vgl. Abel und Meyer 2016: 252-253). Insofern kann der Benutzer neue Einträge erstellen und über die gegebene bzw. fehlende Informationen im Forum diskutieren. Direkten Zugang zu anderen lexikografischen Lern- und Forschungsressourcen bekommt der Benutzer ebenfalls. Dies garantiert aber nicht, dass das gesuchte Wort auch in den anderen Ressourcen zu finden ist. De facto muss der Benutzer immer eine neue Konsultation durchführen, da die Verlinkung nicht direkt zum Stichwort sondern auf die Startseite des ausgewählten Nachschlagewerks führt. In der von Engelberg und Müller-Spitzer (2013: 1029-1033) vorgeschlagenen Portaltypologie befindet sich das PONS-Portal an der Schnittstelle zwischen Wörterbuchsuchmaschinen und –netzen, da es eine Vereinheitlichung des Layouts und der Zugriffstruktur aber eine sehr hohe Eigenständigkeit der beinhalteten Wörterbücher aufweist (vgl. Engelberg und Storrer 2016: 54).

Im Folgenden werden die Einträge zu *posten/postear* (siehe Abb. 9), zu *mailen* (siehe Abb. 10), zu *twittern/tuitear* bzw. *twittear* (siehe Abb. 11) und zu *wasapear* [whatsappen] (siehe Abb. 12) aus dem PONS-Portal in Verbindung mit der Frage nach der Nützlichkeit des Informationsangebots zu den Schwerpunkten (a)–(c) exemplarisch analysiert.

Der Eintrag zu *postear* und der zu *posten* haben verschiedene Autorschaft. Der erste ist von der PONS-Redaktion verfasst worden, während der letzte von einem Benutzer erstellt worden ist, wie in Abbildung 9 zu sehen ist. Dies lässt sich ebenfalls an dem unterschiedlichen Informationsangebot beider Artikel bemerken. Der Benutzer findet nur Hinweise zu der semantischen und syntaktischen Valenz des Verbpaars *posten/postear*, wenn er von dem Spanischen ausgeht (Suche nach *postear* = „postear algo“ → „etw posten“).



Abb. 9: Bearbeitete Bildschirmausdrucke der Einträge zu *postear* und *posten* in Pons-Online [06.03.18]

Für Rezeptionszwecke hingegen ist die Konsultation in beiden Ausrichtungen erfolgreich, da sowohl der ELE- als auch der DaF-Lernende den Ausdruck in seiner Muttersprache findet, der ihm hilft, die Bedeutung in der Fremdsprache zu deuten. Der Eintrag zu *posten* bietet zusätzlich eine ausführliche Bedeu-

tungserklärung an. Bezüglich der Ausrichtung DE→ES lässt sich das Stichwort *posten*, das mehrdeutig ist, anhand von kategoriellen Informationen (bspw. „VERB“ vs. „SUBST m“) und Angaben zum Sachgebiet bzw. Kontextgebrauch („einkaufen“, „Botengänge machen“ oder „INET“) gut disambiguieren.

Ein anderes Szenario stellt der Artikel zu *mailen* dar. Wie in Abbildung 10 zu sehen ist, findet der Benutzer zwei verschiedene Einträge zu *mailen*, die sich in der Information zu der AS voneinander unterscheiden. Beide führen zu sehr ähnlichen aber jedoch nicht identischen Entsprechungsvorschlägen im Spanischen. Wenn ein DaF-Lernender das Verb *mailen* in diesem Wörterbuch nachschlägt, um Informationen zu der adäquaten Verwendung dieses Verbes zu erhalten, muss er entscheiden können, welche von den angebotenen Varianten (d.h. die Dativ- oder die Präpositionalvariante von *mailen*) sich am besten für seinen Zweck eignet. Da keine Gebrauchsbeispiele angeboten werden, kann ihm diese Entscheidung schwer fallen. Wenn er sich für die Präpositionalvariante entscheiden sollte, muss er erst die Angaben „mailen (an) (akk)“ richtig interpretieren können. Er muss schon vorher wissen, dass sich die Angabe „(akk)“ auf die Kasusrektion der in der Angabe „(an)“ erwähnten Präposition bezieht. Es muss sich dann um einen fortgeschrittenen DaF-Lernenden oder einen kundigen bzw. sehr erfahrenen Benutzer (vgl. Wiegand 1998: 505-508) handeln. Wenn aber diese Kenntnisse fehlen, muss der Benutzer nicht zwangsläufig zu dem richtigen Schluss kommen, dass das Verb *mailen* mit einem Präpositionalobjekt auftritt, dessen Präposition (an) Akkusativ regiert“. Er kann gleichfalls auf die falsche Interpretation kommen, dass das Verb *mailen* ein Akkusativobjekt und ein Präpositionalobjekt fordert, bei dem der passende Kasus zu der Präposition *an* unbekannt ist“. Über die semantische Valenz von *mailen* oder die mögliche lexikalische Besetzung der Argumente erhält der Benutzer an dieser Stelle ebenfalls keine Auskunft. Die syntagmatischen Informationen sind unvollständig. Das ist ein Beispiel von der Problematik hinsichtlich der Kodifizierung von Valenzangaben in Wörterbüchern, die schon Herbst (2009: 287-289) in Bezug zu Lernerwörterbüchern für das Englische angesprochen hat.

mailen <small>VERB</small>		Benutzereintrag	
◀ jdm mailen	enviar un e-mail (a alguien)	◀	+
◀ mailen (an) (akk)	enviar un sms, mail... (a alguien)	◀	+

Abb. 10: Eintrag zu *mailen* in Pons-Online [06.03.18]

Darüber hinaus kann der Benutzer nur Zugriff zu dem Verb *mailen* erhalten, wenn er schon vorher das Verb kennt und gezielt nach ihm sucht. Ein spanischsprachiger Benutzer, der sowohl die Verben *enviar* [senden], *mandar* [schicken] und *escribir* [schreiben] als auch die nominalen Komponente der Konstruktio-

nen *correo* [Post] und *e-mail* [E-Mail] nachschlägt, wird nicht unmittelbar zu *mailen* geführt. Erst nach weiteren Konsultationen und über das deutsche Substantiv *Mail* gelangt er an den Benutzereintrag zu *mailen*.

Die in Abbildung 11 abgebildeten Einträge zu *tuitear* und *twittern* bestätigen ebenfalls, dass die Äquivalenzrelationen bzw. das Äquivalenzangebot in PONS-Online unsystematisch sein kann. Beispielsweise kommt der Benutzer von *twittern* im Deutschen sowohl zu *tuitear* als auch zu *twittear* im Spanischen aber nicht umgekehrt. Die Schreibvariante *twittear* entspricht keinem spanischen Eintrag oder Stichwort im Wörterbuch. Die Suche nach *tuitear* führt zum Verb *twittern* ohne weitere morphosyntaktische Informationen (z.B. kein Hinweis auf Transitivität). Dieses unsystematische und unvollständige Informationsangebot hat möglicherweise seine Ursache in der Autorschaft der Beiträge, da sie von Benutzern und nicht vom Redaktionsteam erfasst worden sind. Daraus schließt sich, dass PONS als ein kollaborativ-institutionelles Wörterbuch verstanden werden soll, bei dem die Benutzerbeiträge — *a priori* — redaktionell unbearbeitet bleiben (vgl. Lew 2011: 237, Abel und Meyer 2016: 260-261).



Abb. 11: Bearbeitete Bildschirmausdrucke der Einträge zu *tuitear* und *twittern* in Pons-Online [06.03.18]

Ähnliche Kritikpunkte kann man bezüglich des Eintrags zu *wasapear* [whatsapp] vorbringen, obwohl es sich **nicht** um einen Benutzereintrag handelt (siehe Abb. 12). Dem Benutzer wird eine Entsprechung zu dem nachgeschlagenen spanischen Stichwort *wasapear* geliefert. Das syntagmatische Informationsangebot — ohne ergänzende Gebrauchsbeispiele — erweist sich als unzureichend für die Produktionszwecke des DaF-Lernenden. Beispielsweise wird dem DaF-Lernenden die Information zu dem Kasus des Reflexivpronomens *sich* in dem Ausdruck *sich über WhatsApp schreiben* nicht geliefert. Die Angabe ist folglich für einen Benutzer mit geringem Sprachniveau mangelhaft.



Abb. 12: Eintrag zu *wasapear* in PONS-Online [06.03.18]

5.1.2 Neologismen in Leo-Portal

Die Leo GmbH bietet vielfältige zweisprachige Wörterbücher in Verbindung mit dem Deutschen sowie einen Trainer- und Forumbereich für die jeweiligen Sprachen. Die lexikografische Ressourcen des LEO-Portals sind seit ihrem Ursprung als reines Onlineangebot mit kooperativem Charakter konzipiert (vgl. Engelberg und Storrer 2016: 39 und Abel und Meyer 2016: 283). Die Benutzereinträge werden von Sprachexperten geprüft. In diesem Sinne weisen sich die LEO-Wörterbücher als semi-kollaborativ aus (vgl. Abel und Meyer 2016: 261-263). Dazu gehört auch das bilinguale Wörterbuch Spanisch \leftrightarrow Deutsch. Zusätzlich stellt LEO externe Verlinkungen zu anderen Wörterbüchern und didaktischen Sprachressourcen zur Verfügung, über die der Benutzer weitere Informationen zu dem gesuchten Ausdruck erhalten kann. Allerdings steht diese Verlinkung nicht für alle Stichwörter bereit, da es kein systematisches, redaktionell ausgearbeitetes Angebot ist, sondern automatisch generiert wird.

Im Folgenden werden die Einträge zu *twittern/tuitear* (siehe Abb. 13), zu *posten/postear* (siehe Abb. 14 und Abb. 15), zu *mailen* (vgl. Abb. 16) und zu *wasapear* [whatsappen] (siehe Abb. 17) aus LEO illustriert und die Nützlichkeit des entsprechenden Informationsangebots in Verbindung mit den Schwerpunkten (a)–(c) kommentiert.

Mit Unabhängigkeit von der Ausgangssprache wird der Benutzer in der Regel zu einem einzigen zweisprachigen Artikel geführt. Dies könnte damit zusammenhängen, dass LEO die typischen Eigenschaften eines virtuellen Wörterbuchportals aufweist (vgl. Engelberg und Storrer 2016: 53). Sein Informationsangebot ist so vereinheitlicht, dass die Eigenständigkeit der integrierten Wörterbücher sehr schlecht zu erkennen ist. Beispielsweise kommen sowohl der Benutzer, der nach *tuitear*, als auch der Benutzer, der nach *twittern* nachgeschlagen hat, zu demselben Doppelseintrag zu *tuitear/twitter*, der in Abbildung 13 gezeigt wird. Das Wörterbuch LEO zeichnet sich u.a. dadurch aus, dass es dem Benutzer ausführliche morphosyntaktische Informationen anbietet. Zum Beispiel kann der Benutzer in dem Artikel zu *tuitear/twittern* den Info-Button anklicken und sich die ganze Konjugation des jeweiligen Verbs tabellarisch zeigen lassen, wie in Abbildung 13 illustriert wird. Diese Daten sind für die Textproduktion zwar hilfreich aber nicht ausreichend. Konkrete Angaben zur Valenz oder Argumentanzahl von *twittern* findet der Benutzer nämlich nicht. Der Hinweis „– in sozialen Netzwerken“ ist konfus. Diese

Präpositionalphrase kann als typische Verwendung bzw. häufiger Kookkurrenzpartner des Verbs (*in sozialen Netzwerken twittern*) interpretiert werden. Es kann sich aber auch um eine Kontext- bzw. Sachgebiet-Angabe handeln. Weitere Gebrauchsbeispiele erhält man für dieses Verb nicht.

Verben
 tuitear [COMP] - en redes sociales | twittern | twitterte, getwittert | - in sozialen Netzwerken

Weitere Aktionen:
 Neue Diskussion starten | Gespeicherte Vokabeln sortieren | Suchhistorie

Verb tuitear [INFORM.] - en redes sociales
 twittern | twitterte, getwittert | - in sozialen Netzwerken

Verbtabelle für: tuitear		Verbtabelle für: twittern	
Hilfsverb/Verbo auxiliar: haber		Hilfsverb: haben	
Modi Modo indicativo :: Modo subjuntivo :: Modo imperativo :: Formas no personales		Modi Indikativ :: Konjunktiv :: Imperativ :: Unpersönliche Zeiten	
Modo indicativo	Modo subjuntivo	Indikativ	Perfekt
Presente	Pretérito perfecto compuesto	Präsens	Perfekt
(yo) tuiteo	(yo) he tuiteado	ich twittere	ich habe getwittert
(tú) tuiteas	(tú) has tuiteado	du twitterst	du hast getwittert
(él/ella/usted) tuitea	(él/ella/usted) ha tuiteado	er/sie/es twittert	er/sie/es hat getwittert
(nosotros) tuiteamos	(nosotros) hemos tuiteado	wir twittern	wir haben getwittert
(vosotros) tuiteáis	(vosotros) habéis tuiteado	ihr twittert	ihr habt getwittert
(ellos/ellas/ustedes) tuitean	(ellos/ellas/ustedes) han tuiteado	sie twittern	sie haben getwittert
Pretérito imperfecto	Pretérito pluscuamperfecto	Präteritum	Plusquamperfekt
(yo) tuiteaba	(yo) había tuiteado	ich twitterte	ich hatte getwittert
(tú) tuiteabas	(tú) habías tuiteado	du twittertest	du hattest getwittert
(él/ella/usted) tuiteaba	(él/ella/usted) había tuiteado	er/sie/es twitterte	er/sie/es hatte getwittert
(nosotros) tuiteábamos	(nosotros) habíamos tuiteado	wir twitterten	wir hatten getwittert
(vosotros) tuiteabais	(vosotros) habíais tuiteado	ihr twittertet	ihr hattet getwittert
(ellos/ellas/ustedes) tuiteaban	(ellos/ellas/ustedes) habían tuiteado	sie twitterten	sie hatten getwittert
Pretérito perfecto simple (definido)	Pretérito anterior	Futur I	Futur II
(yo) tuiteé	(yo) hubé tuiteado	ich werde twittern	ich werde getwittert haben
(tú) tuiteaste	(tú) hubiste tuiteado	du wirst twittern	du wirst getwittert haben
(él/ella/usted) tuiteó	(él/ella/usted) hubo tuiteado	er/sie/es wird twittern	er/sie/es wird getwittert haben
(nosotros) tuiteamos	(nosotros) hubimos tuiteado	wir werden twittern	wir werden getwittert haben
(vosotros) tuiteasteis	(vosotros) hubisteis tuiteado	ihr werdet twittern	ihr werdet getwittert haben
(ellos/ellas/ustedes) tuitearon	(ellos/ellas/ustedes) hubieron tuiteado	sie werden twittern	sie werden getwittert haben

Abb. 13: Bearbeitete Bildschirmausdrucke der Einträge zu *twittern* und *tuitear* in leo.org [06.03.18]

Eine Ausnahme bilden die Einträge für das Verbpaar *posten/postear*, da die Suche nach dem jeweiligen Verb je nach Ausgangssprache zu unterschiedlichen Einträgen leitet, wie die Abbildungen 14 und 15 veranschaulichen. Der Grund dafür ist, dass sich das Stichwort *posten* im Deutschen auf zwei homophone Wörter beziehen kann: das Substantiv *der Posten* und das Verb *posten*, wie Abbildung 15 bezeugt. Die Suche nach *postear* hingegen führt direkt zu dem spanischen Eintrag „postear algo“ mit seinem deutschen Pendant „etw. ^{Akk} posten“ (siehe Abb. 14). Da findet der Benutzer semantische und syntaktische Valenzangaben, die er für die adäquate Verwendung des Verbs und seiner Argumente im Deutschen benötigt. Dazu steht eine kurze Bedeutungserklärung „- etwas in einem Weblog schreiben“ zur Verfügung, die gleichzeitig als Kontexthinweis dient. Weitere morphosyntaktische Informationen kann der Benutzer über den Info-Button erhalten.

Zu diesen Informationen gelangt auch der Benutzer, der „posten“ nachgeschlagen hat. Er bekommt sie allerdings auf eine andere Art und Weise dargestellt. Der Artikel zum deutschen Stichwort „posten“ teilt das Informationsangebot in zwei Blöcke, wie man in Abbildung 15 sehen kann. Der erste trägt die Überschrift „Substantive“. Der zweite lautet „Verben“. Logischer-

weise hätte der Benutzer denken können, dass er, wenn er einmal den „Substantive“-Block gemieden und sich direkt auf den „Verben“-Block gerichtet hat, dort die Informationen erhalten würde, die ausschließlich das Verb *posten* betreffen. Das ist jedoch nicht der Fall. Der Benutzer muss erst diejenigen verbalen Ausdrücke und Kollokationen aussortieren, die mit dem Verb *posten* nicht in Verbindung stehen. Dann bleiben drei passende Einträge übrig. Diese sind in Abbildung 15 eingerahmt worden. Die drei Einträge auf der deutschen Seite beinhalten fast identische Angaben. Sie korrelieren allerdings mit drei verschiedenen Einträgen auf der spanischen Seite. Der erste Entsprechungsvorschlag zu „*etw.^{Akk} posten*“ ist „*postear algo*“ [etwas posten], der zweite „*publicar un post*“ [einen Post veröffentlichen] und der dritte „*publicar algo*“ [etwas veröffentlichen]. Die drei Einträge werden durch weitere Angaben kontextuell eingegrenzt (z.B. „*in sozialen Netzwerken*“). Unbekannt bleibt noch der Grund, warum das vielfältige Angebot an Ausdrucksmöglichkeiten nur für das Spanische und nicht auch für das Deutsche vorhanden ist.



Abb. 14: Eintrag zu *postear* in leo.org [06.03.18]



Abb. 15: Bearbeiteter Ausschnitt aus dem Eintrag zu *posten* in leo.org [06.03.18]

Die Abbildungen 16 und 17 dienen zur Illustration der Einträge zu dem deutschen Verb *mailen* und dem spanischen Substantiv *e-mail* einerseits, und zu dem spanischen Verb *wasapear* und dem Substantiv *WhatsApp* andererseits. Sie bestätigen, dass die Auffindung komplexer Einheiten in LEO normalerweise erfolgreich ist und dass man auch über komplexe Einheiten Zugang zu einfachen Einheiten erhält. Beispielsweise führt die Suche nach dem spanischen Substantiv *e-mail* sowohl zu den komplexen Einheiten *escribir un e-mail a alguien* [jemandem eine E-Mail schreiben] und *enviar algo a alguien por correo electrónico* [jemandem etwas per elektronischer Post senden] im Spanischen und *jemandem etwas per E-Mail senden* im Deutschen als auch zum Verb *mailen* (siehe Abb. 16). Genauso gelangt man über das deutsche Stichwort *WhatsApp* zu dem deutschen Ausdruck *per WhatsApp schreiben* bzw. *jemandem (Nachrichten) über WhatsApp schreiben* und zu dem spanischen Ausdruck *mandar wasaps a alguien* [jemandem WhatsApp-Nachrichten schicken] aber auch zum Verb *wasapear* (siehe Abb. 17).

Suche nach *mailen* DE → ES

Verben

escribir un e-mail a alguien | jmdm. mailen | mailte, gemailt | [ugs.]

Weitere Aktionen

Neue Diskussion starten | Gespeicherte Vokabeln sortieren | Suchhistorie

Suche nach *e-mail* ES → DE

Substantive

el emilio [COMP] [ugs.] | die E-Mail Pl.: die E-Mails englisch

correo electrónico [COMP] | die (auch: das) E-Mail Pl.: die E-Mails englisch

el e-mail auch: E-mail, email [COMP] | die E-Mail Pl.: die E-Mails englisch

correo electrónico [COMP] | das E-Mail Pl.: die E-Mails englisch besonders (Österr.; Schweiz)

el e-mail auch: E-mail, email [COMP] | das E-Mail Pl.: die E-Mails englisch (Südt.; Österr.; Schweiz)

la cuenta [COMP] | der (auch: das) Account Pl.: die Accounts - E-Mail etc.

Verben

enviar algo a alguien por correo electrónico | (jmdm.) etw.^{Akk} per E-Mail senden

escribir un e-mail a alguien | jmdm. mailen | mailte, gemailt | [ugs.]

Verb **escribir un e-mail a alguien** | jmdm. mailen | mailte, gemailt | [ugs.]

Substantiv **el emilio [INFORM.] [col.]** | die E-Mail Pl.: die E-Mails englisch

Abb. 16: Bearbeitete Bildschirmausdrucke der Einträge zu *mailen* und *e-mail* in leo.org [12.03.18]



Abb. 17: Bearbeitete Bildschirmausdrucke der Einträge zu *wasapear* und *WhatsApp* in leo.org [12.03.18]

5.1.3 Neologismen in Langenscheidt-Online

Das im Internet angebotene zweisprachige Wörterbuch Spanisch ↔ Deutsch von Langenscheidt fußt auf seinem gedruckten Pendant und soll eher als eine digitalisierte Version des Printwerks verstanden werden, die in dem neuen Portal des Langenscheidt Verlages integriert ist. Das Wörterbuchportal funktioniert wie eine Suchmaschine. Trotzdem zeigt es hinsichtlich Layout Ähnlichkeiten zu Wörterbuchnetzen auf (vgl. Engelberg und Storrer 2016: 52-53). Das Wörterbuch befindet sich noch in dem Medienwandel-Prozess (vgl. Wiegand et al. 2010: 88-89), da es das Potenzial des digitalen Formates nicht optimal ausgeschöpft hat (vgl. Debus-Gregor und Heid 2013: 1009-1011, Engelberg und Storrer 2016: 34-35). Beispielsweise verfügt es weder über externe Verlinkungen zu anderen lexikografischen Ressourcen noch über interne Verlinkungen zu ergänzenden Materialien oder zu Diskussionsforen. Der Benutzer hat dennoch die Möglichkeit, durch ein Feedback-Formular seine Meinung zu äußern.

Im Folgenden werden die Einträge zu *twittern/tuitear* (siehe Abb. 18), zu *posten/postear* (siehe Abb. 19), zum deutschen *mailen* und zum spanischen *wasapear* [whatsappen] (siehe Abb. 20) aus Langenscheidt in Verbindung mit der Frage nach der Nützlichkeit des Informationsangebots zu den Schwerpunkten (a)–(c) kommentiert.

Abbildung 18 präsentiert die Einträge zu *tuitear/twittern* aus Langenscheidt-Online im spanisch–deutschen Vergleich. Auf den ersten Blick wird schon auffällig, dass der deutsche Eintrag zu *twittern* ausführlicher als der spanische Eintrag zu *tuitear* ist. Er enthält über die Entsprechungsvorschläge in der Zielsprache (vgl. „usar Twitter“ und „tuitear“) und die Angaben zu Aussprache und Wortart des Stichwortes (vgl. „[ˈtvit̪ərm] v/i“) hinaus zusätzliche

Gebrauchsbeispiele zum Stichwort aber auch zu den vorgeschlagenen Äquivalenten in der Zielsprache, eine kurze Bedeutungserklärung des Stichwortes (vgl. „≈ den Internetdienst Twitter nutzen“) sowie Markierungen zu Sprachregister der vorgeschlagenen Äquivalenten in der Zielsprache (vgl. „UMG“) und Sachgebiet des Stichwortes (vgl. „INTERNET“). Das heißt, dass der Benutzer, der zunächst nach dem spanischen Verb *tuitear* gesucht hat, noch eine weitere Konsultation tätigen muss, um an diese Daten zu kommen. Auffällig wird auch auf den zweiten Blick, dass die Angaben zu Transitivität bzw. Intransitivität im Widerspruch zu den syntagmatischen Informationen stehen, die aus den Beispielen hergeleitet werden können. Laut der Angabe zur Wortart, die neben dem Stichwort „twittern“ steht, ist das Verb intransitiv (vgl. „v/i“). Im unteren Beispiel wird das Verb aber transitiv verwendet (vgl. „er hat das Ergebnis des Spiels getwittert“).



Abb. 18: Bearbeitete Bildschirmausdrucke der Einträge zu *tuitear* und *twittern* in Langenscheidt-Online [22.02.18]

Auf Abbildung 19, die die Einträge zu *posten/postear* illustriert, lässt sich ebenfalls feststellen, dass der Zugriff zu den Gebrauchsbeispielen ausschließlich über das Deutsche als Ausgangssprache erfolgt. Der spanische Eintrag liefert zwar Informationen zu Transitivität (vgl. „v/t“) und zu der semantischen bzw. lexikalischen Füllung des Direktobjekts (vgl. „Kommentar“), aber zu den anschaulichen Gebrauchsbeispielen für beide Sprachen gelangt man erst, wenn man auf das deutsche Übersetzungsäquivalent klickt und zum deutschen Eintrag weitergeleitet wird.



Abb. 19: Bearbeitete Bildschirmausdrucke der Einträge zu *posten* und *postear* in Langenscheidt-Online [22.02.18]

Das letzte Beispiel liefert Abbildung 20, die die Einträge zum spanischen *wasapear* [whatsappen] und deutschen *mailen* im Vergleich präsentiert. Im Eintrag zu *mailen* gelangt man nach den Überschrift-Informationen direkt zum Beispielbereich ohne weitere Aufschlüsselung der Entsprechungsvorschläge. Den Beispielen kann der Benutzer relevante Informationen zu Semantik, Syntax und teilweise Gebrauchskontext des Verbs entnehmen. Dies bestätigt die Beobachtung (siehe oben), dass illustrative Beispiele nur in deutschen Einträgen angeboten werden. Der spanische Eintrag zu *wasapear* [whatsappen] seinerseits richtet sich eher auf Rezeptionszwecke, da die angebotenen deutschen Entsprechungen bzw. Übersetzungsäquivalente zu *wasapear* den Status als feste Verbindungen oder typische Verwendungen nicht besitzen. Sie scheinen nicht konventionalisiert zu sein. Sie können vielmehr als Bedeutungserklärung angesehen werden. Diese Hypothese wird durch weitere Konsultationen bekräftigt. Die betreffenden Ausdrücke „WhatsApp nutzen“ und „WhatsApp-Nachrichten verschicken“ können weder über das Stichwort „WhatsApp“, „nutzen“ oder „verschicken“ im Wörterbuch aufgerufen werden. Im Gegensatz dazu gelangt man immer zu den spanischen Kollokationen *mandar/enviar algo a alguien por correo electrónico* [jemandem etwas per elektronischer Post schicken/senden] und *mandar/enviar un mail/e-mail a alguien* [jemandem eine Mail/E-Mail schicken/senden], die als Entsprechungen zu *mailen* gelten. Sie sind sowohl unter ihren Nominalkomponenten „correo“, „mail“ und „e-mail“ als auch unter ihren Verbalbasen „mandar“ und „enviar“ im Wörterbuch abrufbar.



Abb. 20: Bearbeitete Bildschirmausdrucke der Einträge zu *wasapear* und *mailen* aus Langenscheidt-Online im Vergleich [14.03.18]

5.2 Schlussfolgerung zu Informationsangebot in bilingualen Online-Wörterbüchern

Anhand der neuen MKV des Deutschen und Spanischen ist die Aufführung von Neologismen in drei im Internet angebotenen zweisprachigen Wörterbüchern für das Sprachenpaar Spanisch–Deutsch mit Fokus auf der Produktion in L2 untersucht worden. Das lexikografische Angebot ist auf Adäquatheit und Nützlichkeit bei (a) der Produktion in der Fremdsprache, (b) dem Such-Auffindungsprozess (insbesondere von komplexen Einheiten) und (c) der ausgangs- und zielsprachigen Disambiguierung überprüft worden. Nach einer kritischen Beurteilung der Ergebnisse dieser Untersuchung aus kontrastiver Perspektive lassen sich folgende Schlüsse ziehen:

(a) Das Angebot an morphosyntaktischen und lexikosemantischen Daten erweist sich für die richtige bzw. kontextadäquate Anwendung der Ziellemmata in der Fremdsprache in einigen Fällen als unzureichend. Der Benutzer ist oft gezwungen, weitere Konsultationen zu tätigen, um auf die gesuchten Informationen zu gelangen. Ferner ist dieses Informationsangebot manchmal bei den Benutzereinträgen in PONS-Online unvollständig und in Leo.org unsystematisch. In Langenscheidt-Online sind die spanischen und deutschen Einträge mit unterschiedlicher Sorgfältigkeit bzw. Ausführlichkeit ausgearbeitet. Allen drei Wörterbüchern mangelt es an anschaulichen Beispielen, obwohl bestätigt wurde, dass Benutzer einen schnellen Zugriff zu grammatischen Informationen finden, wenn diese in Beispielen dargestellt werden (vgl. Bogaards und Van der Kloot 2001: 117-118). Die angebotenen Beispiele scheinen Kompetenz- und keine Korpus- bzw. Originalbeispiele zu sein (vgl. Engelberg und Lemnitzer 2009: 236-238, Prinsloo 2013). Ein Vorteil von Kompetenzbeispielen im Vergleich zu Original- bzw. nicht modifizierten Korpusbeispielen ist, dass sie kürzer sind und einfacher zu verstehen sind. Auf der anderen Seite fehlen denen oft kontextuelle Informationen und sie werden als künstlich empfunden (vgl. Prinsloo 2013: 512-513).

(b) Die Auffindung komplexer Einheiten hat sich meistens als erfolgreich bestätigt. Obwohl diese oft erst nach mehreren Konsultationen erfolgte und der gesuchte Ausdruck meistens im Artikeltext verborgen ist. Die Ausnahme bilden die Benutzereinträge in PONS-Online, da die Korrelation dieser Beiträge miteinander und mit anderen Redaktionseinträgen fehlerhaft ist. Die Suche nach komplexen Einheiten hat in den drei Wörterbüchern jedoch nicht immer zu einfachen Lexemen in der Zielsprache (DaF) geführt.

(c) Die angebotenen Daten erlauben — in der Regel — die Ausgangslemmata semantisch adäquat zu disambiguieren. Die syntagmatische und pragmatische Disambiguierung sowie die korrekte Auswahl der Ziellemmata können hingegen in einigen Fällen schwer fallen. Dies ist besonders der Fall in Leo.org, da

unterschiedliche Lesarten und sogar Homonyme nicht ausreichend gut differenziert sind und gemeinsam in einem einzigen Artikel dargestellt werden.

Zusammenfassend bieten die untersuchten zweisprachigen Wörterbücher zwar morphosyntaktische Daten und Gebrauchsbeispiele, aber dieses Angebot ist für die adäquate Produktion in der Fremdsprache nicht immer behilflich; insbesondere im Falle von Neologismen, da in diesen Einträgen meistens die Bedeutungserläuterung im Fokus steht. Darüber hinaus sind keine relevanten Unterschiede in dem Sprachangebot der drei Wörterbücher festgestellt worden. Bezüglich der Benutzerbeteiligung scheint das semi-kollaborative Wörterbuch von LEO einen höheren Grad an Stabilität und Zuverlässigkeit als das kollaborativ-institutionelle Wörterbuch von PONS aufzuweisen, da bei dem letzteren den Eindruck erweckt wird, dass die Benutzereinträge nicht auf Qualität geprüft sind. Diese werden mit dem Etikett „Benutzereintrag“ versehen, das sie von den anderen Einträgen unterscheidet. Die Benutzereinträge in PONS sind in der Makro- und teilweise in der Mikrostruktur des Wörterbuches integriert, da sie dasselbe Layout wie Redaktionseinträge aufweisen und ebenfalls aufrufbar sind. Sie sind aber in der Mediostruktur nicht integriert, da sie nicht miteinander und mit den anderen Einträgen vernetzt sind. Diese Integration erfolgt erst, wenn sie von der Redaktion überprüft und übernommen werden. Dann sind sie von den Redaktionseinträgen nicht mehr zu unterscheiden.

6. Exemplarischer Vergleich der lexikografischen Daten mit korpusbasierten Daten

Auf Grundlage einer Korpusuntersuchung medialer Kommunikationsverben des Deutschen und des Spanischen, die im Rahmen des im Abschnitt 1 erwähnten Dissertationsprojekts durchgeführt worden ist, lässt sich aus empirischen Korpusdaten feststellen, dass gerade Informationen zu dem Gebrauchs-kontext, dem lexikosyntaktischen Kombinationspotenzial und der Argumentstruktur unter anderen syntagmatischen Informationen besonders wichtig sowohl für die Übersetzung als auch für die freie Textproduktion in der Fremdsprache sind, da die festgestellten Divergenzen zwischen den Verben beider Sprachen genau darin liegen. Pro Verb wurden durchschnittlich 150 Belege aus den entsprechenden *Corpora from the Web* (COW) für das Deutsche und das Spanische (*realise* 2012 und 2014) der Freien Universität Berlin (vgl. Schäfer 2015). Zwei ausgewählte Beispiele anhand der Verben *posten* und *twittern* sollen zur Illustration dienen.

(i) Das Lokativ-Argument bei *posten*:

In der betreffenden Korpusuntersuchung ist festgestellt worden, dass das deutsche Verb *posten* in fast der Hälfte seiner Realisierungen mit einem Lokativ-Argument (LOC) vorkommt (siehe Abb. 21)⁸. Dieses Argument kann als Adverbial- oder Präpositionalphrase (AP oder PP) realisiert werden. Bezüg-

lich der präpositionalen Realisierung des Lokativ-Arguments weist das deutsche Verb eine breitere Vielfalt an Präpositionen auf, obwohl eine Dominanz der Präposition *in* zu erkennen ist.

Lokativ-Argument im		47%		
Zusammenhang mit <i>posten</i>				
K_{NP} V M_{NP} LOC_{PP} bei		1,3%	transitiv	aktiv
K_{NP} V M_{NP} LOC_{PP} unter		1%		
K_{NP} V M_{NP} LOC_{PP} an		1%		
K_{NP} V M_{NP} LOC_{PP} in		4,5%		
K_{NP} V M_{NP} LOC_{PP} in + Akk		2%		
K_{NP} V M_{NP} LOC_{PP} auf		2,2%		
K_{NP} V M_{NP} LOC_{AP} hier/da		10%		
K_{NP} V P_{NS} LOC_{PP} in		9%		
K_{NP} V T_{PP} über LOC_{PP} in		0,5%	intransitiv	
K_{NP} V T_{PP} zu LOC_{PP} in		0,5%		
K_{NP} V LOC_{PP} in		5%		
K_{NP} V LOC_{AP} hier		4%		
K_{NP} V LOC_{AP} dort		0,5%		
K_{NP} V LOC_{AP} hier FIN_{umzu}		0,5%		
M_{NP} V LOC_{PP} in		2%		
M_{NP} V LOC_{PP} auf		1%		
M_{NP} V LOC_{AP} dort		0,5%		
M_{NP} V LOC_{AP} hier		1,3%		
M_{NP} V LOC_{AP} hier K_{PP} von		0,5%		
T_{PP} über V LOC_{PP} in		0,5%		
V LOC_{PP} in		0,5%		

Abb. 21: Realisierungen von *posten*, in denen ein Lokativ-Argument vorkommt und ihre Häufigkeit

Diese Informationen stellen die Relevanz des Lokativ-Arguments bei dem Verben *posten* heraus. Sie sind jedoch in keinem der konsultierten Wörterbücher (siehe Abschnitt 4. und 5) aufgeführt worden. Die entsprechenden Einträge zu *posten* enthalten keine Angaben zum Lokativ-Argument. Lediglich aus den „typischen Verwendungen“ im *Neologismenwörterbuch* des IDS kann man schließen, dass *posten* ein Lokativ-Argument als Präpositionalphrase realisieren kann (vgl. Eintrag zu *posten* im *Neologismenwörterbuch*).

(ii) Realisierung der MESSAGE bei *twittern*:

Abbildung 22 liefert korpusbasierte Daten zu den Realisierungen von *twittern*. Das Verb kann sowohl transitiv als auch intransitiv verwendet werden. Das MESSAGE-Argument (M) wird meistens durch eine Nominalphrase (NP) realisiert. Es kann aber auch als Nebensatz (P_{NS}) oder als Hauptsatz bzw. Zitat (DR_{HS}) kodiert werden. Zudem können MESSAGE- und TOPIK-Argument (T) alternieren. Das Lokativ-Argument zeigt ebenfalls bei den dokumentierten Realisierungen von *twittern* eine verhältnismäßig starke Präsenz.

In den konsultierten Wörterbüchern sind die Informationen zu Realisierungsmöglichkeiten des MESSAGE-Arguments und zu den anderen möglichen Argumenten von *twittern* — wenn vorhanden — in den Beispielen verborgen. Lediglich in dem *Neologismenwörterbuch* des IDS sind etwas expliziter Angaben dazu unter den „typischen Verwendungen“ und bei der Beschreibung der syntaktischen Umgebung des Verbs *twittern* zu finden (vgl. Eintrag zu *twittern* im *Neologismenwörterbuch*). De facto etikettieren es fast alle konsultierten Wörterbücher, die das deutsche Verb *twittern* inventarisieren, als intransitives Verb.

Häufige Realisierungen von <i>twittern</i>	91%		
K_{NP} V M_{NP}	20%	transitiv	aktiv
K _{NP} V DR _{HS}	7 %		
K _{NP} V P _{NS}	4%		
K_{NP} V	22%	intransitiv	
K _{NP} V T _{PP über/von/zu}	9%		
M_{NP} V	22%		
M _{NP} V LOC _{AP/PP}	4%		
V	3,5%		

Abb. 22: Realisierungen von *twittern*, die häufiger als 3% sind

Die Befunde bestätigen die Notwendigkeit einer lexikografischen Beschreibung von Neologismen, in der neben ihrer Bedeutung auch ihre syntagmatischen Eigenschaften und ihre kontextuelle Lage u.a. miteinbezogen sind. Dies gewährt bereits das *Neologismenwörterbuch* des IDS für das Deutsche. In der sprachvergleichenden (Lerner-)Lexikografie hingegen ist momentan ein

derartiges Werk nicht vorhanden. Die Lernende einer Fremdsprache verstehen aber Neologismen als wichtigen Bestandteil des gängigen Alltagswortschatzes und müssen sie in der Praxis korrekt einsetzen. Die am häufigsten verwendeten zweisprachigen Online-Wörterbücher im spanisch-deutschen Kontext decken diesen Bedarf nur teilweise, da sie in der Regel Benutzerszenarios für die fremdsprachige Rezeption und die Rückübersetzung anvisieren (vgl. Tarp 2008: 147-149 und 161-163, Wiegand et al. 2010: 23-24 und 83-84).

7. Plädoyer für ein zweisprachiges Neologismenwörterbuch Spanisch ↔ Deutsch

Die bereits präsentierten Untersuchungsergebnisse zeigen, dass die Erstellung eines zwei- bzw. mehrsprachigen Neologismenwörterbuchs sinnvoll und in dem konkreten Fall für das Sprachenpaar Spanisch ↔ Deutsch auch notwendig ist. Sie stimmen gleichzeitig mit den Ergebnissen anderer aktueller lexikografischer Studien überein, die auch eine ungenaue Berücksichtigung der möglichen Wörterbuchbenutzungssituationen sowie ein teilweise unvollständiges Informationsangebot bestätigen (vgl. Meliss 2013, 2015a, 2015b, 2016). Davon zeugen ebenfalls vielfältige einschlägigen Benutzerumfragen (vgl. Domínguez Vázquez et al. 2013, Meliss 2015b, Fernández Méndez et al. 2016), in denen die Benutzer selbst behaupten, einige Informationen vermisst zu haben, die vor allem das syntagmatische Umfeld und die Kombinatorik des gesuchten Wortes betreffen (vgl. Meliss 2015b: 420-421).

Danksagung

Ich möchte mich bei Danie Prinsloo für Ihre unendliche Geduld und die Unterstützung bei der Überarbeitung dieses Artikels bedanken. Ebenfalls bedanke ich mich bei den beiden anonymen Gutachtern für Ihre hilfreiche Kommentare.

Endnoten

1. In der vorliegenden Studie wird der allgemeine Begriff „Benutzer,“ (vgl. Wiegand et al. 2010: 675) sowohl zur Bezeichnung des potenziellen Benutzers (Wiegand 1998: 504, Tarp 2008: 54-55) als auch des Benutzers-in-actu (Wiegand et al. 2010: 678) in den konkreten Beispielen verwendet. Außerdem haltet es sich immer um einen fremdsprachigen Benutzer, da zumindest ein Teil der Wörterbuchbenutzung in einer Fremdsprache erfolgt (vgl. Wiegand et al. 2017: 341). Ferner soll nochmal betont werden, dass sich die hier ausgewählte generische Form des Maskulinums sowohl auf männliche Benutzer und Benutzerinnen als auch auf Wörterbuchbenutzer bezieht, die sich mit keinem beider sozialen Geschlechter identifizieren. Die Variante „die Wörterbuch benutzende Person“ bleibt ausgeschlossen, da sie sich lediglich auf den Benutzer-in-actu beschränkt (vgl. Wiegand 1998: 500-501).

2. Insbesondere hinsichtlich des Print-Formats sollte man die bei den Neologismenwörterbüchern problematische Aktualitätsfrage berücksichtigen (vgl. Engelberg und Lemnitzer 2009: 58-60).
3. Da die betreffenden Wörterbücher in Bezug auf die Bedürfnisse ihres potenziellen Benutzers (vgl. Tarp 2008: 54-58), d.h. in Bezug auf den Benutzungsanlass und -grund (Wiegand et al. 2010: 681 und 685) beschrieben worden sind, eignet sich die angeführte Terminologie der an dem lexikografischen Institut der Universität Aarhus entwickelten *Function Theory* (vgl. Tarp 2008) insbesondere, denn in diesem theoretischen Rahmen der potenzielle Benutzer eine zentrale Rolle spielt. Zur Diskussion über die relevantesten Unterschiede zwischen der *Function Theory* und der weitverbreiteten lexikografischen Theorie von Wiegand (1998) siehe Tarp 2008: 39-43 und 80-97.
4. Die Printversion des Wörterbuches, die die Neologismen der Nuller- und Zehnerjahre bis 2010 erfasst (vgl. Steffens und Al-Waldi 2015) wurde für die Analyse nicht berücksichtigt, weil die Online-Version inhaltlich darüber hinaus geht und dazu die neuen Wörter der letzten Jahre aufführt.
5. Siehe Engelberg und Lemnitzer (2009: 99-112) für allgemeine Informationen zum Thema Suche in Online-Wörterbüchern.
6. Letzte Konsultation wurde am 06.03.2018 getätigt.
7. Ein Vorteil kollaborativer Benutzerbeteiligung ist gerade die Schließung inhaltlicher Lücken; insbesondere bezüglich Neologismen, Fachsprache und dialektaler Varietäten (vgl. Abel und Meyer 2016: 278-280). Dennoch sind alle in den kollaborativen Wörterbüchern inventarisierten MKV auch in dem institutionellen Wörterbuch Langenscheidt abrufbar (siehe Abb. 8).
8. Erklärung zu Abkürzungen für die Argumentenbeschreibung: K = Kommunikator, M = Message, DR = direkte Rede, P = Proposition, T = Topik, LOC = interner Lokativ, FIN = Finalität, V = Verb, NP = Nominalphrase, AP = Adverbialphrase, PP = Präpositionalphrase, HS = Hauptsatz, NS= Nebensatz.

Primäre und sekundäre Quellen

- Abel, A. und C.M. Meyer. 2016. Nutzerbeteiligung. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016: 249-290.
- Bergenholtz, H. 2011. Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries. Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.). 2011: 30-53.
- Bergenholtz, H. 2012. Concepts for Monofunctional Accounting Dictionaries. *Terminology* 18(2): 243-263.
- Bergenholtz, H. und I. Bergenholtz. 2011. A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool. Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.). 2011: 187-207.
- Bogaards, P. und W.A. van der Kloot. 2001. The Use of Grammatical Information in Learners' Dictionaries. *International Journal of Lexicography* 14(2): 97-121.
- COW: *Corpora from the Web*. <http://corporafromtheweb.org/>.
- Debus-Gregor, E. und U. Heid. 2013. Design Criteria and 'Added Value' of Electronic Dictionaries for Human Users. Gouws, R.H. et al. (Hrsg.). 2013: 1001-1013.
- DeReKo: *Das Deutsche Referenzkorpus*. <http://www1.ids-mannheim.de/kl/projekte/korpora/>.
- Die Wortwarte. <http://www.wortwarte.de/>.

- Domínguez Vázquez, M.J. (Hrsg.).** 2013. *Trends in der deutsch-spanischen Lexikographie*. Frankfurt a.M.: Peter Lang.
- Domínguez Vázquez, M.J., M. Mirazo Balsa und V. Vidal Pérez.** 2013. Wörterbuchbenutzung: Erwartungen und Bedürfnisse. Ergebnisse einer Umfrage bei Deutsch lernenden Hispanophonen. Domínguez Vázquez, M.J. (Hrsg.). 2013: 135-172.
- Dou, X.** 2004. *Deutsch-chinesisches Neologismenwörterbuch. De-Han xinci cidian*. Beijing: The Commercial Press.
- Engelberg, S. und L. Lemnitzer.** 2009. *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.
- Engelberg, S. und C. Müller-Spitzer.** 2013. Dictionary Portals. Gouws, R.H. et al. (Hrsg.). 2013: 1023-1035.
- Engelberg, S. und A. Storrer.** 2016. Typologie von Internetwörterbüchern und -portalen. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016: 31-63.
- Fernández Méndez, M., M. Franco Barros und I. Sueiro Orallo.** 2016. El uso de los diccionarios de/ para ELE y DaF: análisis de una encuesta desde una perspectiva contrastiva. Robles i Sabater, F., D. Reimann und R. Sánchez Prieto (Hrsg.). 2016. *Sprachdidaktik Spanisch-Deutsch: Forschungen an der Schnittstelle von Linguistik und Fremdsprachendidaktik*: 73-92. Romanistische Fremdsprachenforschung und Unterrichtsentwicklung 6. Tübingen: Narr Francke Attempto.
- Fuertes-Olivera, P.A.** 2009. The Function Theory of Lexicography and Electronic Dictionaries: WIKTIONARY as a Prototype of Collective Free Multiple-Language Internet Dictionary. Bergenholtz, H., S. Nielsen und S. Tarp (Hrsg.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 99-134. Linguistic Insights: Studies in Language and Communication 90. Bern: Peter Lang.
- Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.).** 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.
- Gierden Vega, C., B. Eggelte, B. Heinsch und D. Hofmann.** 2010. *Compendio temático de neologismos Alemán — Español: deutsche Neubildungen*. Stuttgart: Ibidem.
- González Ribao, V.** 2019. *Das Zusammenspiel von Verb- und Musterbedeutung der medialen Kommunikationsverben im Sprachvergleich Deutsch-Spanisch-Englisch. Eine konzeptuell orientierte lexikologische Studie*. Unv. Diss. Santiago de Compostela: Universität Santiago de Compostela, Spanien. Online: <http://hdl.handle.net/10347/18728>.
- Gouws, R.H., U. Heid, W. Schweickard und H.E. Wiegand (Hrsg.).** 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Handbücher zur Sprach- und Kommunikationswissenschaft 5.4. Berlin/Boston: De Gruyter Mouton.
- Heid, U.** 2011. Electronic Dictionaries as Tools: Towards an Assessment of Usability. Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.). 2011: 287-304.
- Herbst, T.** 2009. Item-specific Syntagmatic Relations in Dictionaries. Nielsen, S. und S. Tarp (Hrsg.). 2009: 281-308.
- Klosa, A. und C. Müller-Spitzer (Hrsg.).** 2016. *Internetlexikographie. Ein Kompendium*. Berlin/Boston: De Gruyter.
- Langenscheidt: Online-Angebot Spanisch ↔ Deutsch.* <https://de.langenscheidt.com/deutsch-spanisch/>.
- Lemnitzer, L.** 2010. Neologismenlexikographie und das Internet. *Lexicographica* 26: 65-78.
- LEO: Online-Wörterbuch Spanisch ↔ Deutsch.* <https://www.leo.org/alem%C3%A1n-esp%C3%B1ol>.

- Lew, R.** 2011. Online Dictionaries of English. Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.). 2011: 230-250.
- Meliss, M.** 2013. Das zweisprachige Wörterbuch im bilateralen deutsch-spanischen Kontext. Alte und neue Wege. Domínguez Vázquez, M.J. (Hrsg.). 2013: 61-87.
- Meliss, M.** 2015a. Zum Kombinationspotenzial von Verben in einsprachigen DaF-Lernerwörterbüchern — kritische Bestandsaufnahme und neue Anforderungen. *Zeitschrift für Deutsch als Fremdsprache* 1(2015): 14-27.
- Meliss, M.** 2015b. Was suchen und finden Lerner des Deutschen als Fremdsprache in aktuellen Wörterbüchern? Auswertung einer Umfrage und Anforderungen an eine aktuelle DaF-Lernerlexikographie. Roelcke, T. (Hrsg.). 2015. *Wörterbücher für Deutsch als Fremdsprache — Probleme und Perspektiven*: 401-432. Info Daf-Themenheft 42(4).
- Meliss, M.** 2016. Wie viele und welche bilingualen Online-Wörterbücher brauchen wir für den DaF-Bereich? Erstellung eines Kriterienrasters und erste Bestandsaufnahme aus der Sicht eines hispanophonen Lernerkontextes. Cuartero Otal, J., J.P. Larreta Zulategui und C. Ehlers (Hrsg.). 2016. *Querschnitt durch die deutsche Sprache aus spanischer Sicht. Perspektiven der kontrastiven Linguistik*: 187-210. Sprachwissenschaft 30. Berlin: Frank & Timme.
- Müller-Spitzer, C.** 2010. OWID — A Dictionary Net for Corpus-Based Lexicography of Contemporary German. Dykstra, A. und T. Schoonheim (Hrsg.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010*: 445-452. Leeuwarden: Fryske Akademy. *Neologismenwörterbuch*. <http://www.owid.de/wb/neo/start.html>.
- Nielsen, S. und S. Tarp (Hrsg.)**. 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*. Terminology and Lexicography Research and Practice 12. Amsterdam/Philadelphia: John Benjamins.
- PONS: Online-Wörterbuch Spanisch ↔ Deutsch*. <https://de.pons.com/>.
- Prinsloo, D.J.** 2013. New Developments in the Selection of Examples. Gouws et al. (Hrsg.). 2013: 509-516.
- Quasthoff, U. (Hrsg.)**. 2007. *Deutsches Neologismenwörterbuch*. Berlin/New York: De Gruyter.
- Samaniego Fernández, E. und B. Pérez Cabello de Alba**. 2011. Conclusions: Ten Key Issues in e-Lexicography for the Future. Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.). 2011: 305-311.
- Schäfer, R.** 2015. Processing and Querying Large Web Corpora with the COW14 Architecture. Bański, P., H. Biber, E. Breiteneder, M. Kupietz, H. Lungen und A. Witt (Hrsg.). 2015. *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)*: 28-34. Mannheim: Institut für Deutsche Sprache.
- Scheller-Boltz, D. und H. Weinberger**. 2017. Lexikografie und Innovation. Zur Vielfältigkeit in der gegenwärtigen Lexikografie und warum wir Lexikografie heute anders verstehen müssen. Scheller-Boltz, D. und H. Weinberger (Hrsg.). 2017. *Lexikografische Innovation. Innovative Lexikografie. Bi- und multilinguale Wörterbücher in Gegenwart und Zukunft: Projekte, Konzepte, Visionen*: 7-18. Hildesheim: Olms.
- Steffens, D. und D. Al-Waldi**. ³2015. *Neuer Wortschatz: Neologismen im Deutschen 2001–2010*. Mannheim: Institut für Deutsche Sprache.
- Steffens, D. und O. Nikitina**. ²2016. *Deutsch-russisches Neologismenwörterbuch. Neuer Wortschatz im Deutschen 1991–2010. Немецко-русский словарь неологизмов. Новая лексика в немецком языке 1991–2010*. Mannheim: Institut für Deutsche Sprache. Online: <http://www.owid.de/drnwb/> [08.2018]

- 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: Max Niemeyer.
- Tarp, S.** 2009. Reflections on Data Access in Lexicographic Works. Nielsen, S. und S. Tarp (Hrsg.). 2009: 43-62.
- Tarp, S.** 2013a. New Developments in Learners' Dictionaries III: Bilingual Learners' Dictionaries. Gouws, R.H. et al. (Hrsg.). 2013: 425-431.
- Tarp, S.** 2013b. Lexicographic Functions. Gouws, R.H. et al. (Hrsg.). 2013: 460-468.
- Verlinde, S.** 2011. Modelling Interactive Reading, Translation and Writing Assistants. Fuertes-Olivera, P.A. und H. Bergenholtz (Hrsg.). 2011: 275-286.
- Wiegand, H.E.** 1998. *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Berlin/New York: De Gruyter.
- Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer und W. Wolski (Hrsg.).** 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Band/Volume 1: Systematische Einführung/Systematic Introduction A-C.* Berlin/New York: De Gruyter.
- Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer und W. Wolski (Hrsg.).** 2017. *Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Band/Volume 2: D-H.* Berlin/New York: De Gruyter.
- Worbs, E., A. Markowski und A. Meger.** 2007. *Polnisch-deutsches Wörterbuch der Neologismen. Neuer polnischer Wortschatz nach 1989.* Wiesbaden: Harrassowitz.

Lexical Understanding of Native Bahasa Indonesia Speakers through Word Association to Improve Dictionary Definitions*

Ami P. Jewalani, *Department of Linguistics, Faculty of Humanities,
Universitas Indonesia, West Java, Depok, Indonesia (ami.pramesti@ui.ac.id)*,

Myrna Laksman-Huntley, *Department of Linguistics, Faculty of
Humanities, Universitas Indonesia, West Java, Depok, Indonesia
(laksman@ui.ac.id)*

and

Harwintha Y. Anjarningsih, *Department of Linguistics, Faculty of
Humanities, Universitas Indonesia, West Java, Depok, Indonesia
(harwintha@ui.ac.id)*

Abstract: This study examined lexical cognitive word associations of Bahasa Indonesia native speakers. A word association task using 30 cue nouns was conducted with 45 educated adult native speakers of Bahasa Indonesia aged 20–29, after which the generated data was classified based on the extensive semantic taxonomy. It was found that most responses related to the cue words were associated with lexical features, followed by entity features, situation features, taxonomic category, and introspective features, all of which suggested that this group of Bahasa Indonesia speakers related words to other words that shared similar lexical features, and especially with words that usually come after the target words. It was also found that the participants rarely associated feelings to the cue words as there were very few introspective feature associations. While this was a limited study focused on a specific population and only used nouns, the results could be of assistance in developing dictionaries and thesauri, or could be used as preliminary data to build databases, such as WordNet in Bahasa Indonesia. As there have been few studies focused on Bahasa Indonesia word associations, this study could also be used for future comparative word association studies.

Keywords: WORD ASSOCIATION, SEMANTIC PROPERTIES, CONCEPTUAL PROCESSING, SEMANTIC TAXONOMIC CODING, BAHASA INDONESIA, COGNITION, DICTIONARY DEFINITIONS, WORD RETRIEVAL, LEXICAL ACCESS

* A shortened version of this article was presented as a paper at the Third Asia-Pacific Research in Social Sciences and Humanities Conference 2018 (APRISH 2018), which was hosted by the Faculty of Social and Political Sciences, Universitas Indonesia, in Jakarta, Indonesia, 13–15 August 2018.

Résumé: Compréhension lexicale des locuteurs natifs de Bahasa Indonesia à travers d'une association de mots pour améliorer les définitions de dictionnaires.

Cette étude a examiné les associations de mots lexicaux cognitifs en Bahasa Indonesia de ses locuteurs natifs. Une tâche d'association de mots utilisant 30 noms de mots-clés a été réalisée en impliquant 45 locuteurs indonésiens bien instruits âgés de 20 à 29 ans à la suite de quoi les données générées ont été classées selon la vaste taxonomie sémantique. Cette étude a révélé que la plupart des réponses liées aux mots-clés étaient associées à des caractéristiques lexicales, suivies des entités, des caractéristiques de situation, de la catégorie taxonomique et des caractéristiques introspectives, ce qui suggère que ce groupe de locuteurs indonésiens associe des mots avec d'autres mots partageant des caractéristiques lexicales similaires, en particulier avec les mots qui viennent habituellement après les mots cibles. Par ailleurs, il est à constater que les participants associaient rarement les sentiments aux mots-clés car il y avait très peu d'associations présentant des caractéristiques introspectives. Bien qu'il s'agisse d'une étude limitée centrée sur une population spécifique et utilisant uniquement des noms, les résultats pourraient être indispensables pour le développement de dictionnaires et de thésaurus, ou pourraient s'utiliser comme données préliminaires pour créer des bases de données, telles que Wordnet en Bahasa Indonesia. Puisque peu d'études ont été consacrées aux associations de mots en Bahasa Indonesia, cette étude pourrait également servir à de futures études comparatives sur les associations de mots.

Mots-clés: ASSOCIATION DE MOTS, PROPRIÉTÉS SÉMANTIQUES, TRAITEMENT CONCEPTUEL, CODAGE TAXONOMIQUE SÉMANTIQUE, BAHASA INDONESIA, COGNITION, DÉFINITIONS DE DICTIONNAIRE, RECHERCHE DE MOTS, ACCÈS LEXICAL

1. Word relations for concept understanding

In all languages, words are used to label concepts that are understood by others who speak that language. Generally, concepts are mentally connected to other concepts, such as the concept of food to the concept of eating, or food to hunger, or food to dining room. Language and situated simulation (LASS) theory states that the "conceptual system is organized around situated action" (Barsalou 2003: 522); in other words, accessing one concept brings to mind other associated concepts and accessing a word automatically brings to mind other associated words.

Language and cognition studies have found that understanding these concepts and word associations can give some indication as to how these are stored in the memory, and sociolinguistic studies have examined these word and concept associations to understand how lexical items are understood in different cultures and languages, including Bahasa Indonesia. Therefore, this study examines how Indonesians understand their language through a word association task, which "is the simplest possible form of the linguistic processing that occurs during conceptual processing" (Barsalou, Santos, Simmons and Wilson 2008: 249).

A word association test was used by Bahar, Johnstone and Sutcliffe (1999) to investigate student cognitive structures in a genetics context, for which Eng-

lish genetics cue words, such as gene, chromosome, and mutation were used to generate other cue words to understand how the students linked the words to others. De Deyne and Storms (2008) also used a Dutch word association test to explore semantic networks and their properties, from which they found that the associates were invariably nouns regardless of the word category cues, and most of the elicited words were related to a certain situation or in taxonomic ways to the cue words. The second responses also revealed that entity features became more available indicating that while the first words were taxonomically related, the following associations were conceptual (De Deyne and Storms 2008: 223).

Pranoto and Afrilita (2018) explained the manner in which the network of mental lexicon models in someone's cognition structures can be seen through a lexical relation in psycholinguistics. Therefore, word association tasks can reveal how certain concepts are stored in a certain language community as they require participants to produce words they feel are closely connected to a particular word. Word associations are subject to the culture of the speaker of a certain language since, as Pranoto and Afrilita (2018: 30) stated, "the mental lexicon develops according to the events experienced by the language learner." For example, the word *candle* for Bahasa Indonesia would not be connected to *church* as most Indonesians are Moslem, so this word would be more likely associated with *power failure* — a situation when Indonesians use candles to light the room.

Geeraerts (1987) classified dictionaries into three types based on the theories used in compiling them and their functional intentions. One of them is the large-scale dictionary that is designed to scientifically describe linguistic variations and spread knowledge of the words which uses prototypical concepts as the definitions. The *Kamus Besar Bahasa Indonesia* (KBBI) is one of the examples. In the same way, the associated words from native Bahasa Indonesia speakers could be seen as prototype definitions and could be used by lexicographers to assess a word's meaning in the speakers' minds based on prototype theory. Van Sterkenburg (2003) claimed that while the definitions in dictionaries are in both whole sentences and phrases, the sentence format was more natural and spontaneous in defining words and concepts. There are several headwords with short definitions in the KBBI. For example, the lemma *jawab* 'answer', in the KBBI V offline, is defined as *sahut, balas* 'response, reply'; therefore, more information is needed for this to be fully understandable. Adding information about the context in which *jawab* is defined as *balas* would assist learners to understand better when and how to use the word. Therefore, it is intended that the current study could assist lexicographers develop better definitions by revealing the associated concepts in the minds of Bahasa Indonesia speakers.

This study is also expected to assist in developing a thesaurus of Bahasa Indonesia. In one Bahasa Indonesia thesaurus, Endarmoko (2007), the headword synonyms are written in numerical order. For example, the synonyms for the lemma *bulat* 'round' are written as follows:

bulat *a* 1 bundar; 2 a cukup, genap integral, jangkap (*Jw*), kafi (*Ar*), komplet, lengkap menyeluruh, padu, penuh, sempurna, solid, tamam (*kl*), utuh; tunggal

However, as the difference between 1 and 2 is not obvious, using examples in sentences or from the domain could be added. Therefore, the associated words and conceptual relationships elicited in this study could be a reference for determining an example or the domain in which the words are used.

As words can be related to other words in many ways, to understand the relationships between the cue and the participant associated words, the semantic taxonomic coding adapted by De Deyne and Storms (2008) was used.

2. Semantic taxonomic coding

Depending on culture, language, background and experience, people connect words in many ways. *Sky*, for example, can be related to *cloud*, *stars*, and *moon* or *blue* or even to the word *romantic*. De Deyne and Storms (2008) developed a semantic taxonomy that had a wide range of categories that can be used to determine how one word is linked to another word by meaning and external factors, such as the location where the word is usually found. There were five main categories in De Deyne and Storms' semantic taxonomy: (1) entity features, (2) situation features, (3) taxonomic categories, (4) lexical features, and (5) introspective features as well as several subtypes, all of which are helpful in understanding the cognitive processes associated with the way words are used. Below is the elaboration of De Deyne and Storms' semantic taxonomy categories and their subclasses that is exemplified using Bahasa Indonesia.

2.1 Entity features

De Deyne and Storms (2008) explained that a word association is put into the entity features category if a word is linked to its associate based on their specific concrete features regardless of the situation, for example, *kapur* 'chalk' and *putih* 'white'. They further classified this category into ten subclasses that explain the specific features shared with the intended word: (1) external component, (2) external surface feature, (3) internal component, (4) internal surface feature, (5) behavior, (6) material, (7) quantity, (8) associated abstract entity, (9) systemic feature, and (10) larger whole.

The external feature is related to the outer part of the concept labeled by the main word while the internal feature is related to the inner part (De Deyne and Storms 2008). In addition, the surface feature is regarding the entity that exists in the concept of the word itself or is not its component. For example, with regard to the cue word *cabai* 'chili', *merah* 'red' is categorized as an external surface feature whereas *pedas* 'spicy' is classified as an internal surface feature.

Additionally, the component is associated with the words denoting composition. For instance, *roda* 'tire' is the external component of *mobil* 'car' whereas *bensin* 'fuel' is the internal component.

Moreover, according to De Deyne and Storms (2008), the behavior feature is related to words denoting natural characteristics, such as *bumi* 'earth' that is associated with *berputar* 'rotate' while the material feature is used to classify the relation of two words when one of the words is the material of the object denoted by the other word such as *karung* 'sack' is made of *plastik* 'plastic'. Moreover, the quantity feature designates the amount or number of a particular concept that usually exists, such as *tangan* 'hands' are *dua* 'two'. Words can also have abstract associations with other words, such as *rumah* 'house' and *keluarga* 'family'; systemic characteristics association, such as *asap* 'smoke' and *uap* 'steam', or larger whole type associations, such as a *rumah* 'house' being the larger part of a *lantai* 'floor'.

2.2 Situation features

This type of class, which is divided into thirteen subtypes, defines words of which the properties are linked by situations (De Deyne and Storms 2008). (1) The function subtypes indicate words that are related by a function, for example *berjalan* 'to walk' is a function of the word *kaki* 'leg'. Words can also be linked by (2) action properties, such as a *bola* 'ball' is *ditendang* 'kicked' or (3) object properties, such as *kaki* 'feet' and *sepatu* 'shoes'. (4) The person property can also be associated with a word in a particular situation, such as *anak-anak* 'children' that is associated with the word *gambar* 'picture'. (5) Living thing properties link living things to other words, such as *kucing* 'cat' and *pasir* 'sand', (6) social organization properties relate words, such as *murid* 'students' to *sekolah* 'school', and (7) social artifact properties link words, for instance *gambar* 'picture' with the artifact *lukisan* 'painting'.

In addition to the previous features, De Deyne and Storms (2008) stated that (8) location features, (9) time features, and (10) events features connect words through the situational properties; *cabai* 'chili' that can be found in the *pasar* 'market', a *karung* 'sack' that is used to celebrate *Hari Kemerdekaan* 'Independence Day', and a *stadion* 'stadium' connected to *piala dunia* 'world cup', are all examples of the location, time, and event subtypes, respectively. Words are also connected by (11) manner properties — how an event or action is performed or to what an entity is transformed, (12) physical states of a situation, and (13) quantities; for example, the association between *jawab* 'answer' and *jelas* 'clear' (manner), *lilin* 'candle' and *panas* 'hot' (physical state), and *balap* 'race' and *banyak mobil* 'many cars' (quantity).

2.3 Taxonomic categories

Another category used to display the manner in which words are connected to

each other, as defined by De Deyne and Storms (2008), is taxonomic categories in which words are connected taxonomically in more general terms (superordinate), more specific terms (subordinate), or have a shared superordinate meaning (coordinate). Words are also linked as a synonym or antonym of the target word, or as a specific example of it. *Mobil* 'car', for instance, has *sedan* 'sedan' as a subordinate and *sepeda motor* 'motorcycle' as a coordinate as these are linked through the superordinate *kendaraan* 'vehicle'. An example of words related to the antonym and synonym, respectively, is *tanya* 'ask' and *balas* 'response' for the word *jawab* 'answer', and the word *Toyota* is an individual example of the word *mobil* 'car'.

2.4 Introspective features

The feelings of speakers toward certain words labeling a concept can be a reason for an association. People's positive or negative evaluations of a word and their emotions regarding a concept are two of the several types of introspective features in De Deyne and Storms' (2008) classification. *Nyaman* 'comfortable' and *damai* 'peaceful' are emotions speakers may feel when thinking about the word *house*. Meanwhile, *ramah* 'friendly' could be an evaluation of the *sopir* 'driver' concept. Representation, one of the introspective feature subtypes, refers to the representational state in the mind of a situational participant, such as their beliefs, goals, desires, ideas, and perceptions (De Deyne and Storms 2008). The word *patah* 'broken' associated with the word *sayap* 'wings' is an example of representation. The possibilities of how a concept denoted by a word can have happened (cause and effect) can be grouped as contingency, such as the *asap* 'smoke' that is caused by an *api* 'fire'. Words can relate to other words through comparisons that require cognitive operation, such as defining a *kuda* 'horse' that is similar to a *keledai* 'donkey'. Introspective features also include word associations that negate a property, such as a *burung unta* 'ostrich' that *tidak bisa terbang* 'cannot fly'. Finally, another subcategory of introspective features is quantity, in terms of numerosity, frequency, intensity, or typicality of an introspection or one of its properties (De Deyne and Storms 2008), such as the *matahari* 'sun' that elicits the word *sangat panas* 'very hot'.

2.5 Lexical features

De Deyne and Storms (2008) posited another feature of how words are associated with one another: lexical feature. It explains that words are connected to each other by their lexical features, such as the words that are located after or before (forward and backward completion) the word, its fragment, or orthographic similarity; for example, *boneka* 'doll' is a forward completion of the word *rumah* 'house', *susun* 'stack' is a backward completion of the word *rumah*

'house', *mata* 'eye' is a fragment of the word *kacamata* 'glasses', and *baku* 'basic' and *bahu* 'shoulder' are related as they share similar orthographic forms.

Furthermore, De Deyne and Storms (2008) added that words can also relate to one another through mediation or because they are an expression used in language; for example, *kapur* 'chalk' and *semen* 'cement' are mediated by the concept *pembangunan* 'construction', and *bunga* 'flower' and *desa* 'village' are related through an expression in Bahasa Indonesia (i.e., to denote the most beautiful lady in the region — *bunga desa*). Comments about the characteristics of words, called metacomments, are also considered word associations; for example, the word *duplikasi* 'duplication' comes to mind for the word *kupu-kupu* 'butterfly'.

3. Method

3.1 Participants

Forty-five adult native speakers of Bahasa Indonesia (20–29 years old) participated in the study selected by a purposive sampling technique. As all were Universitas Indonesia students, it was assumed they had a wide Bahasa Indonesia vocabulary. Participants in this study joined the experiment on their own accord after viewing the advertisement broadcast through WhatsApp. Nevertheless, before the experiment began, they were additionally asked if they were willing to be involved in the experiment. No participants were language impaired or hearing impaired and all had normal or corrected-to-normal vision. Moreover, since this study was conducted in Universitas Indonesia, the participants' ethnicity varied. Most participants were Javanese, some others were Sundanese, Betawi, and Minang. This research was approved by DRPM (Directorate of Research and Community Service) of Universitas Indonesia.

3.2 Materials

This study used 30 nouns as cue words taken from previous research done by Sari (2019). The words were ascertained as nouns based on the dictionary *Kamus Besar Bahasa Indonesia V* (KBBI V). Brysbaert, Wijnendaele and De Deyne (2000) noted that the earlier individuals acquire words, the easier they are to access from the lexicon; therefore, with this in mind, this study selected words that are acquired earlier in life. The age of acquisition of the cue words is five years, or in other words, they have already been acquired by five-year-old children. It was confirmed by five kindergarten teachers who teach 34 children aged five years, that those cue words have been acquired by their students. The cue words were selected based on nouns that exist in close proximity to everyday life. Table 1 displays all the cue words used in the study.

Cue words		
<i>bola</i> (ball)	<i>lantai</i> (floor)	<i>lilin</i> (candle)
<i>pasir</i> (sand)	<i>asap</i> (smoke)	<i>paku</i> (nail)
<i>tempat</i> (place)	<i>besi</i> (iron)	<i>sopir</i> (driver)
<i>rumah</i> (house)	<i>roda</i> (tire)	<i>balap</i> (race)
<i>jawab</i> (answer)	<i>nilai</i> (score)	<i>karpas</i> (carpet)
<i>kaki</i> (foot)	<i>kapur</i> (chalk)	<i>kaus</i> (t-shirt)
<i>bunga</i> (flower)	<i>susun</i> (stack)	<i>sayap</i> (wing)
<i>mobil</i> (car)	<i>karung</i> (sack)	<i>setir</i> (steering wheel)
<i>gambar</i> (picture)	<i>cabai</i> (chili)	<i>badut</i> (clown)
<i>tangan</i> (hand)	<i>senam</i> (gymnastics)	<i>bantal</i> (pillow)

Table 1: Nouns that are in close proximity to everyday life as cues

3.3 Procedure

The research took place in a classroom. The participants were shown the cue word on a screen for fifteen seconds and asked to write down as many words or concepts as they could think of after reading the cue. However, for this study, only the first responses were used to ensure the analyzed associated words were not influenced by words other than the cues or other words accessed after the cues. The cue words were presented randomly.

3.4 Data processing and analysis

The collected data — the first words written by participants after viewing the cue words — were classified by the first rater (i.e., the first author) based on semantic taxonomic coding adapted by De Deyne and Storms (2008). For example, it was decided that the word *cepat* 'fast' was connected to the cue word *asap* 'smoke' through its manner properties; in other words, as *smoke* moves *fast* in certain situations, the word *cepat* 'fast' was classified as situation-manner. Once the category for each response was determined, the classification was subsequently assessed by four different assessors to ensure the validity of interpretation. All assessors were lecturers in the language and linguistics field whose

ages ranged between 30 and 58 years. Each of the four assessors was given a questionnaire in the form of tables containing responses of participants that had already been categorized. They were required to assess whether the categorization in the tables was correct or not. If the latter is applied, then the assessors were asked to classify the response based on De Deyne and Storms' (2008) classification. The classification that has been completed was determined as valid if three out of four assessors confirmed them as correct.

4. Word association categories of responses

The response received from participants was around 1,350 responses. However, two responses were excluded as one participant did not give any response to one word and the other response was not related to the cue word at all — the word *dari* 'from' for cue word *senam* 'gymnastics'. Another 33 remaining categorized responses could not be counted as valid as the number of raters agreeing to a category was fewer than three people.

Based on the categorization adapted by De Deyne and Storms (2008), most of the 1,315 responses that were confirmed as valid by the assessors were related to the cue word by their entity properties (25.9%). For example, *tangan* 'hand' and *jempol* 'thumb' were followed by responses related to their lexical features (23.7%), such as responses *tanggal lahir* 'date of birth' for cue word *tempat* 'place'. From Figure 1, it can additionally be noted that situation feature was the category to which a large number of responses were related (21.3%) after lexical features. The word *cincin* 'ring' was an example of a response that was related by situation feature to the cue word *tangan* 'hand'. As much as 18% of the responses were categorized to be related to the cue words according to their taxonomical categories, such as *permadani* 'rug' for the cue word *karpet* 'carpet'. The remaining responses (10.8%) were associated with the cue words by the mental state attached to the cue words, named by De Deyne and Storms (2008) as introspective features — *nilai* 'score' and *jelek* 'bad'.

As the majority of cue words were associated with entity and lexical features, the results indicated that the participants tended to first associate the cue words, nouns in this case, with the information related to the concept characteristics as an entity than with their application in a phrase, which suggested that younger Indonesians may store words close to words that share the same properties or entities connected to the cues or their collocations. The relatively few words associated with introspective properties might have been an indication that these words were not associated with any particular mental state in their lexical experiences.

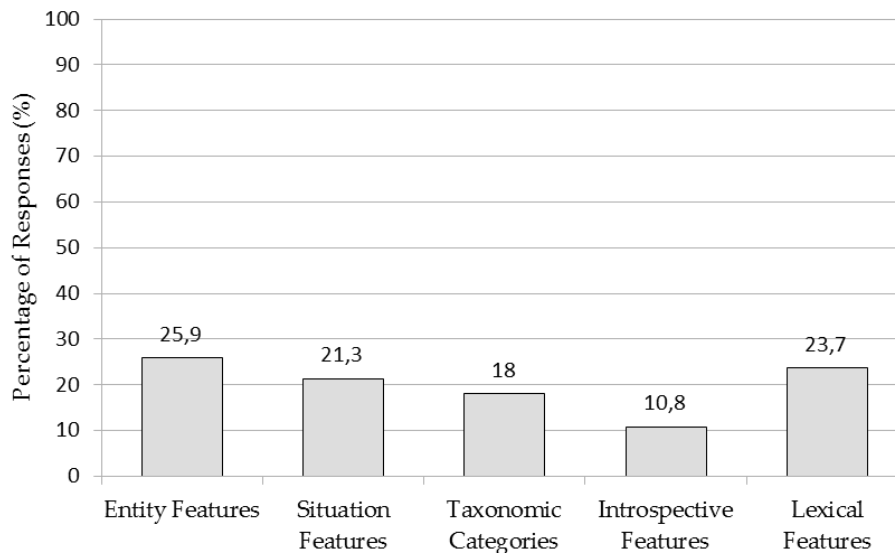


Figure 1: Elicited words sorted into main classes

4.1 Entity features

As mentioned before, most words elicited from the participants were associated with the cue words by their entity features. Around 41.3% of the responses to the cues that were related to entity was classified as being linked by the external surface features of the cue words; for example, the words *karpét* 'carpet' elicited the words *berbulu* 'furry', *kasar* 'rugged', *keras* 'hard', *merah* 'red', and *tebal* 'thick', all of which were related to the surface characteristics — the texture, color, and the size — of the carpet. The second largest entity feature responses (11.7%) were associated with parts of the cue words; for example, the word *kendaraan* 'vehicle' and *mobil* 'car' were evoked by *roda* 'tire'. A further 11.1% of responses were associated with materials; for instance, the word *lantai* 'floor' evoked *marmèr* 'marble' and *ubin* 'tile'.

Two subtypes in this category had similar proportions, namely systemic features and internal surface features, at 8.5 % and 8.2%, respectively. A systemic feature example was the cue word *meleleh* 'melt' being associated with *lilin* 'candle', and an example of an internal surface feature was the word *pedas* 'spicy' being associated with *cabai* 'chili'. The external component was the feature by which 7.6% responses are related to the cue words. For instance, the word *kelopak* 'petal' is the external component of *bunga* 'flower'.

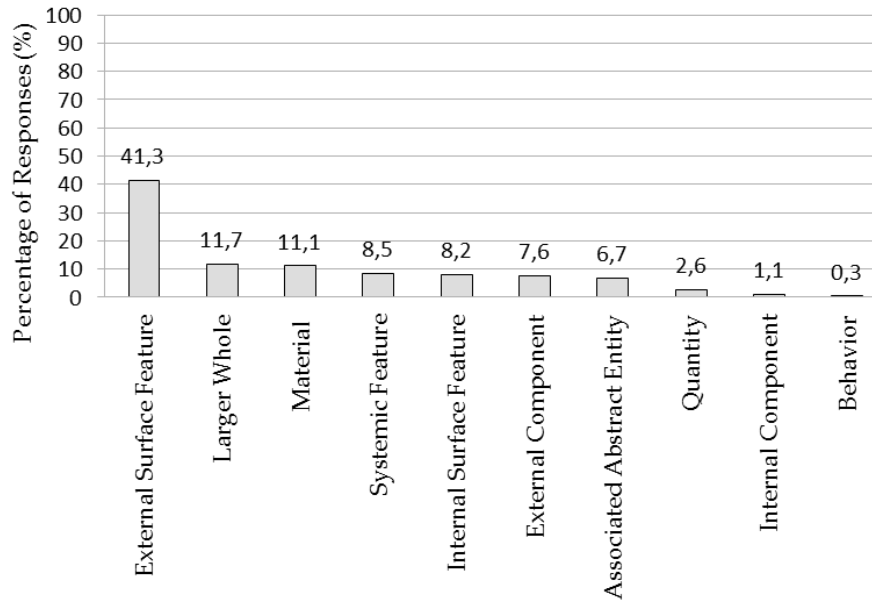


Figure 2: Percentage of responses classified in the entity features category

Entity features, which are abstract association properties, such as *kuntilanak* (a ghost that comes from an Indonesian myth who is deterred by sticking a nail on its head) for the cue word *paku* 'nail' and *karung* 'sack' with *lomba* 'contest' (sacks are often used in contests in Indonesia, especially while celebrating Independence Day), were found in 6.7% of responses. Words denoting quantity were found in 2.6% of cue word responses, such as *dua* 'two' and *sepasang* 'pairs' from the cue word *kaki* 'foot'. Internal component connections were only found in 28 out of 341 responses (9.8%), such as *putik* 'pistil' from the cue word *bunga* 'flower'. There were very few behavioral entity responses (0.3%); for example, the term *dapat berputar* 'able to spin' was given in response as a natural characteristic of *roda* 'tire'.

These results gave some indications regarding the thinking of young Indonesians on the identification of entity features. As there was a higher proportion of external concept responses and less focus on the internal aspects of the entity, it could be discerned that these educated young Indonesians may have stored/defined words related to an entity from the outside to the inside and then to the natural behavior. However, more detailed research would be necessary to confirm this supposition because the very little published research in Bahasa Indonesia discusses word storage from different perspectives not directly comparable to the one employed here (as in Pranoto and Afrilita 2018).

4.2 Lexical features

As can be seen in Figure 3, the majority of the words associated with cue words were forward completion (72.1%), followed by backward completion (16.7%). Almost all the cue words that evoked forward or backward completion suggested that these young educated Indonesian participants were able to easily access the associated collocates. Furthermore, as most responses were associated with collocates located to the right of the cues, this might have been because of the Bahasa Indonesia phrase pattern *Diterangkan Menerangkan* or Head + Modifier (Suharianto in Dewi 2013), which locates the head to the left of the modifier. Nouns in Bahasa Indonesia can be used as both a head and a modifier in a phrase. Nevertheless, the result showing that most of the participants think of words that modify the cue words indicates that most participants consider the cue words as the headword of a phrase. When the cue word *rumah* 'house' was given, for example, participants were more likely to mention other words whose role is as the modifier of the cue words, such as *makan* 'eating' – restaurant; *sakit* 'sick' – hospital; or *duka* 'sad' – funeral home, rather than convert the cue words to the modifier by mentioning *atap* 'top'.

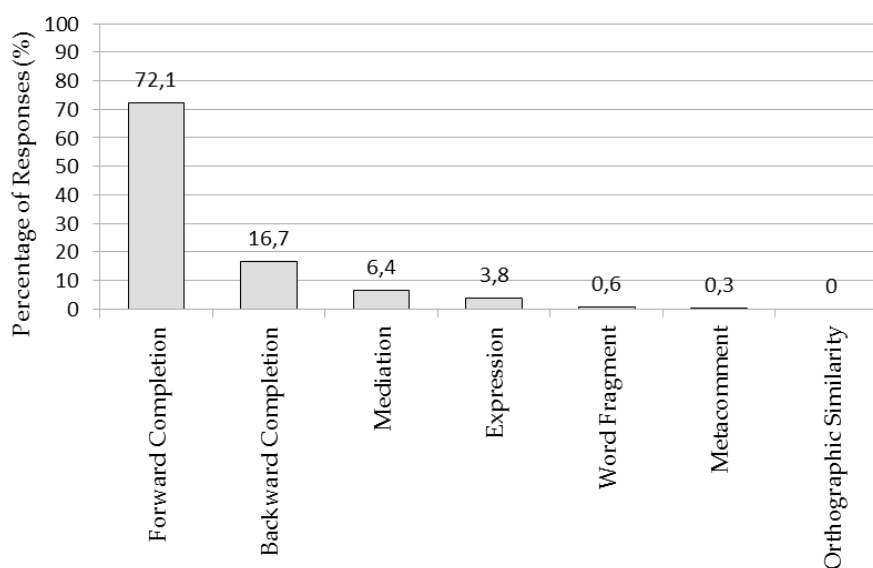


Figure 3: Percentage of first word association connected to the cues by lexical features

Around 6.4% of the responses in the lexical features categories were connected to the cues through situation or entity mediation; for example, the word *asap* 'smoke' induced the word *awan* 'cloud' as both words were mediated through

their shapes — the shape of smoke is similar to clouds. A few words (3.8%) stimulated by the cues were categorized as expressions since the cues were used in idiomatic phrases to express concepts apart from the literal meaning; for example, the cue *tangan* 'hand' evoked the word *panjang* 'long', which in combination is an expression in Bahasa Indonesia for a thief, thereby indicating that these expressions are closely stored. However, there were fewer words that were arising from the cue words associated with fragments (0.6%) and metacomments (0.3%), while there were no responses that were related due to orthographic similarity to the cue words, although the stimuli were presented in their written forms.

4.3 Situation features

As mentioned, 20.51% of the words related to the cue words had situation properties. As Figure 4 shows, most of these responses (30.4%) were related to another object in a situation; for example, the word *bola* 'ball' evoked the word *gawang* 'goal post'; or location (60/283 responses or 21.2%); for example, *pantai* 'beach' and *laut* 'sea' were associated with the cue word *pasir* 'sand'. There were 15.5% of words denoting function from the cue words, such as *menulis* 'to write' from the cue word *tangan* 'hand'.

Around 6.4% of responses identified people closely related to the cue words, such as *sopir* 'driver' from the cue word *setir* 'steering wheel' — named as a person. As with the person category, 18 responses were related to the cue words as they were the physical state of the cue words. Meanwhile, 6% of responses denoted actions, such as *mekar* 'bloom' from the cue word *bunga* 'flower.' There were few manner-related responses, namely *berurutan* 'in sequence' from the cue word *susun* 'stack' and time *mati lampu* 'power failure' from the cue word *lilin* 'candle' words, at only 4.2% and 3.5%.

The social artifacts and event classes appeared in even fewer situation feature responses (2.1% and 2.8%); for example, the word *karya* 'creation' was related to the cue word *gambar* 'picture' as a social artifact, and 17 *Agustus* (Indonesian Independence Day) was associated with the cue word *karung* 'sack'.

The classes with the fewest situation feature responses were living things, and buildings. The word *kucing* 'cat' was linked to the word *pasir* 'sand' as cats use sand as their toilet, and only one cue word elicited a building in their relation; *rumah* 'house' elicited by *paku* 'nail'.

The data analysis revealed that participants most often associated the words based on the situation with other objects in a certain situation, the location, and the function, which suggested that other features were not closely attached to the word in the mental lexicon.

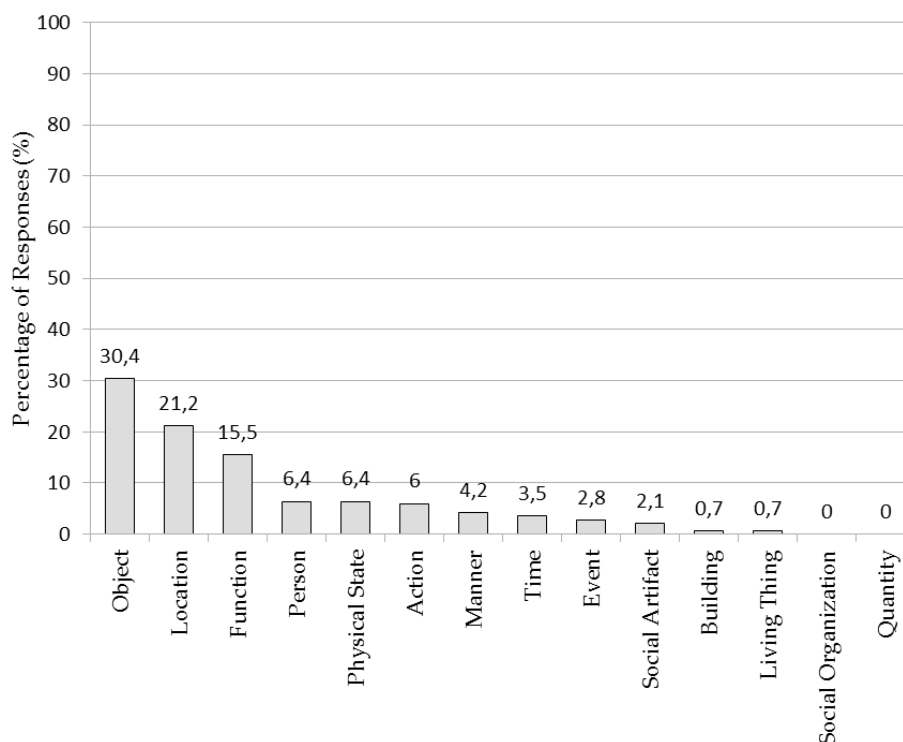


Figure 4: Responses related to the cue words by situation

4.4 Taxonomic categories

Besides associating words with their lexical features, entity, and situation properties, people often recalled other words that had taxonomic connections. The majority of responses (27.4%) that connected the cue words by the taxonomic category were subordinate to the cue words (Figure 5); for example, *anggrek* 'orchid', *cemara* 'pine', *matahari* 'sun', *mawar* 'rose' and *melati* 'jasmine' for the cue word *bunga* 'flower'. Superordinates, such as *logam* 'metal' from cue word *besi* 'iron', were elicited by 23.2% of responses in this category, and coordinates (e.g. *debu* 'dust' and *tanah* 'soil' from the cue word *pasir* 'sand') were elicited in 16.9% of the responses.

Cue word antonyms emerged in the taxonomic category responses more often than coordinates — 43 times out of 237, for example, *acak* 'random' was given as the opposite of the cue word *susun* 'stack'. Synonyms and specific cue examples were given in 7.09% of responses in this category, such as *kemudi* that had the same meaning as *setir* 'steering wheel' and *pengemudi* 'driver' that had a similar meaning to *sopir* 'driver'. Examples of words given to designate a spe-

cific example of the cue words were Dufan, McD, and Joker for the cue word *badut* 'clown'.

The result shows that it could be possible that the participant group stored words in more taxonomic ways than in meaningful relationships with other words, except for antonyms, since superordinates, subordinates, and coordinates of the cue words emerged far more often than their synonym and the specific examples. Moreover, the small number of specific examples suggested that these may be stored at quite a distance from the intended word and not common to anyone since, as mentioned earlier, mental lexicons are dependent on what has been experienced by a person.

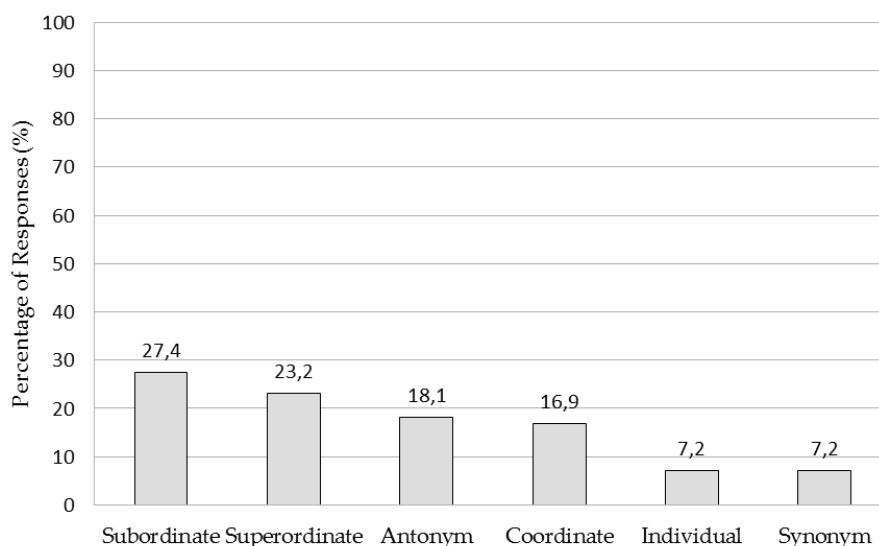


Figure 5: Proportion of the responses linked by the taxonomic categories

4.5 Introspective features

Introspective word association represents the mental state of the participants toward the cue words (De Deyne and Storms 2008). As Figure 6 shows, the majority (64.8%) of the responses in this category that had introspective features were related to the participants' positive or negative evaluations of the cues; for example, the word *mobil* 'car' elicited *keren* 'cool', *hedon* 'hedonic', *mahal* 'expensive', and *mewah* 'luxurious'. Cause and effect features — labeled as contingency features — were found in 21.1% of responses; for example *bugar* 'fit', *capek* 'exhausted', *keringat* 'sweat', and *sehat* 'healthy' from the cue word *senam* 'gymnastics'. Emotional features were found in 12.5% of responses; for example, *takut* 'afraid', *seram* 'scary', and *lucu* 'funny' were responses to the cue

word *badut* 'clown'. The representation feature was found in only a small number of responses (4/144 or 2.2%); for example, the cue word *sayap* 'wing' elicited *patah* 'broken' and *enak* 'tasty'.

These results indicated that the introspective responses were mainly in the form of participants' judgement, positive or negative, and that words related to the cue words for causal effect were stored closer than those related to emotion. The absence of responses that are related to the cue words by cognitive operation, negation or quantity may suggest that when thinking of nouns, Bahasa Indonesia speakers rarely associate them with something that does not exist within them or state their comparison to other nouns.

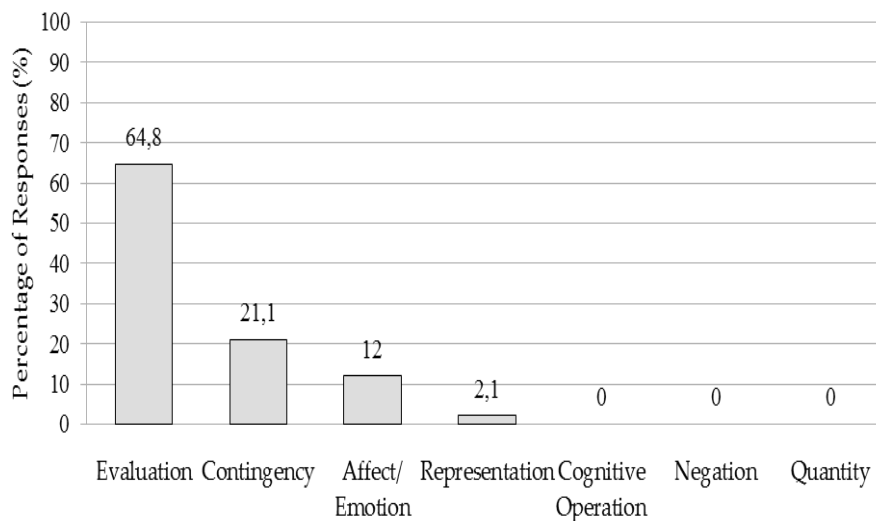


Figure 6: Introspective feature responses

5. Conclusion

Observing the way speakers of a certain language think about words and concepts can reveal the ways that these concepts are stored in the mental lexicon. This research has revealed information about how young native speakers of Bahasa Indonesia process words in their lexicon by connecting them with words that share lexical features or share one or more entity feature aspects. Generally, it was found that the Bahasa Indonesia participants, with reference to the entity features, perceive nouns based on their outside characteristics first. There was also a tendency for them to access nouns according to how they were used in speech, such as in phrases.

Therefore, the physical properties, situation, taxonomic categories, and feelings influenced how the words were processed in the mental lexicon. The appearance of these features in the retrieved words indicates the contexts attached to the words as conceptualized by native speakers of Bahasa Indonesia.

In conclusion, knowing how the native speakers of a certain language understand a word in their language can assist in formulating dictionary definitions so that language learners and native speakers can understand the best possible way that new words can be learned and stored. Lexicographers could use this type of research information to include more outside characteristics for the words listed in the dictionary, such as defining the meaning of a noun to ease the understanding of those particular words. Further, the words associated with the lemma in the dictionary entry could be used as the best example of how the lemma is used in a sentence and as a prototype of the language users' cognition.

Knowing how native speakers of Bahasa Indonesia understand a word in their language and its characteristics could also assist in developing a Bahasa Indonesia thesaurus to assist writers or learners understand how words in Bahasa Indonesia are used. The words generated in this study could also be used as examples of the domains in which the synonyms belong. Knowing how words are related to another in the minds of certain cultures could be a base to create a tool that relates words, such as WordNet, which has not yet been developed for Bahasa Indonesia.

However, as the cue words used in this study were only nouns, different results were found from the extensive study done by De Deyne and Storms (2008). Further, although the cue words were all nouns, it is undeniable that there might be a possibility that the participant consider some of those as verbs since the classification of word categories according to the speakers of Bahasa Indonesia is vague. This is corroborated in Jewalani's study (2019), which found that many of the participants — Bahasa Indonesia speakers — mistakenly classified the category of target words, or in other words, different from what is stated in the KBBI V. According to Jewalani (2019: 71-72), this phenomenon demonstrated that the theoretical system of Bahasa Indonesia, especially regarding word class category, is different from its implementation in the Bahasa Indonesia speech community. Thus, the upcoming studies should take this possible phenomenon into consideration.

In addition, as the participants were educated young people aged 20–29, there is a wide scope for future studies to add to these results and develop a database of Bahasa Indonesia lexical associations. Future research could focus on wider age and socio-economic groups in Indonesia and also assess whether there are any word association differences between the many ethnic Indonesian groups, all of which would deepen understanding of the way that Bahasa Indonesia is lexically stored.

Acknowledgements

The authors are very thankful for the financial support given by Direktorat Riset dan Pengabdian kepada Masyarakat (DRPM) Universitas Indonesia through Publikasi Internasional Terindeks untuk Tugas Akhir Mahasiswa (PITTA) grant 2018.

We also would like to express our special thanks to Sansiviera M. Sari for her major contribution to data collection.

References

Dictionaries

- Badan Pengembangan dan Pembinaan Bahasa.** 2016. *Kamus Besar Bahasa Indonesia (KBBI) V* (Version 0.2.1 Beta (21)) [Mobile application software]. Retrieved from <http://play.google.com/store/apps/details?id=yuku.kbbi5>.
- Endarmoko, E.** 2007. *Tesaurus Bahasa Indonesia* [Thesaurus of Bahasa Indonesia]. Jakarta: Gramedia Pustaka Utama.

Other References

- Bahar, M., A.H. Johnstone and R.G. Sutcliffe.** 1999. Investigation of Students' Cognitive Structure in Elementary Genetics through Word Association Tests. *Journal of Biological Education* 33(3): 134-141.
- Barsalou, L.** 2003. Situated Simulation in the Human Conceptual System. *Language and Cognitive Processes* 18(5-6): 513-562.
- Barsalou, L.W., A. Santos, W.K. Simmons and C.D. Wilson.** 2008. Language and Simulation in Conceptual Processing. De Vega, M., A.M. Glenberg and A.C. Graesser (Eds.). 2008. *Symbols and Embodiment: Debates on meaning and cognition*: 345-283. Oxford: Oxford University Press.
- Brysaert, M., I. van Wijnendaele and S. de Deyne.** 2000. Age-of-acquisition Effects in Semantic Processing Tasks. *Acta Psychologica* 104(2): 215-226.
- De Deyne, S. and G. Storms.** 2008. Word Associations: Network and Semantic Properties. *Behavior Research Methods* 40(1): 213-231.
- Dewi, J.** 2013. Wujud Kesalahan Bahasa Guru Bahasa Indonesia SMP Laboratorium Undiksha: Tinjauan Aspek Struktur dan Diksi. [Form of Errors of Bahasa Indonesian Teachers of Undiksha Laboratory Junior High School: Overview of Structural and Dictional Aspects]. *Jurnal Pendidikan Bahasa dan Sastra Indonesia Undiksha* 1(6): 1-15. Retrieved from <https://ejournal.undiksha.ac.id/index.php/JJPBS/article/view/991/859>.
- Geeraerts, D.** 1987. Types of Semantic Information in Dictionaries. Ilson, R. (Ed.). 1987. *A Spectrum of Lexicography: Papers from AILA, Brussels 1984*: 1-10. Amsterdam: John Benjamins.
- Jewalani, A.P.** 2019. *Pengaruh Kelas Kata pada Akses Leksikal Penutur Bahasa Indonesia dalam Kasus Tip-of-the-tongue* [The Influence of Word Categories on Lexical Access of Indonesian Speakers in Tip-of-the-tongue Cases]. Unpublished Thesis. Depok: Universitas Indonesia.

- Pranoto, B.E. and L.K. Afrilita.** 2018. The Organization of Words in Mental Lexicon: Evidence from Word Association Test. *TEKNOSASTIK* 16(1): 26-33.
- Sari, S.M.** 2019. *Defisit Semantik Leksikal pada Anak Autism Spectrum Disorder* [Lexical Semantics Deficit in Children with Autism Spectrum Disorder]. Unpublished Thesis. Depok: Universitas Indonesia.
- Van Sterkenburg, P. (Ed.).** 2003. *A Practical Guide to Lexicography*. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.

Use of Hedges in Definitions: Out of Necessity or Theory-Driven?

Mariusz Piotr Kamiński, *University of Applied Sciences, Nysa,
Poland (mariusz.kaminski@pwsz.nysa.pl)*

Abstract: Language has an inventory of words and expressions (e.g. *especially, sort of, loosely speaking*) used to communicate that what is being said is not exactly precise or complete. Referred to as hedges, they provide support for the conception of prototypically organized categories, developed by Eleanor Rosch in the 1970s and elaborated by her followers in subsequent decades. Given the fact that hedges are extremely useful for lexicographers in defining, this paper examined the frequency and distribution of hedges in major English dictionaries over the past centuries. One of the findings of this research is the fact that, although hedges have been used in English lexicography since Johnson-1785, their use has intensified in certain dictionaries since the rise of prototype theory, suggesting that recent defining practice in these dictionaries must have been influenced by this theory. Other factors determining the use of hedges were explored.

Keywords: HEDGES, DICTIONARIES, LEXICOGRAPHY, DEFINITION, DEFINITION LANGUAGE, PROTOTYPE THEORY, HISTORY, FREQUENCY, DISTRIBUTION, DEFINING STYLE

Opsomming: Die gebruik van vaaghede in definisies: Noodsaaklik of teoriegedrewe? Taal beskik oor 'n lys woorde en uitdrukkings (bv. *veral, 'n soort, oor die algemeen*) wat gebruik word om aan te dui dat wat gesê word nie heeltemal akkuraat of volledig is nie. Hierdie terme, wat vaaghede genoem word, verleen steun aan die voorstelling van prototipies georganiseerde kategorieë wat in die 1970's deur Eleanor Rosch ontwikkel is en deur haar volgelinge in daaropvolgende dekades verfyn is. Aangesien vaaghede uiters nuttig vir leksikograwe is tydens definiëring, is die frekwensie en verspreiding daarvan in die belangrikste Engelse woordeboeke van die afgelope eeue in hierdie artikel ondersoek. Een van die bevindings van hierdie navorsing is dat, alhoewel vaaghede sedert Johnson-1785 in Engelse leksikografie gebruik is, hul gebruik in sekere woordeboeke toegeneem het sedert die opkoms van die prototipeteorie, wat daarop dui dat onlangse definieerpraktyk(e) in hierdie woordeboeke deur dié teorie beïnvloed moes gewees het. Ander faktore wat die gebruik van vaaghede bepaal, is ook verken

Slutelwoorde: VAAGHEDE, WOORDEBOEKE, LEKSIKOGRAFIE, DEFINISIE, DEFINISIE-TAAL, PROTOTIPETEORIE, GESKIEDENIS, FREKWENSIE, VERSPREIDING, DEFINIEERSTYLE

1. Introduction

Language has an inventory of words and expressions used to communicate that what is being said is not exactly precise or complete. George Lakoff has referred to such metalinguistic devices as hedges, explaining that their "mean-

ing implicitly involves fuzziness" and their "job is to make things fuzzier or less fuzzy" (1973: 471). From a syntactic point of view, a great number of hedges are adverbs and sentence adverbials which function in sentences as modifiers, for example: *loosely speaking*, *strictly speaking*, *technically*, *especially*, *chiefly*, *specifically*, *in particular*, *very*. However, hedges are not a grammatically uniform group, as they include other word categories and even punctuation marks.¹ As Ken Hyland (1994 and 2006) remarks, a hedge is any metalinguistic device that marks uncertainty, hesitation, ambiguity, and tentativeness.

Hedges express varying degrees of category membership. As Lakoff demonstrates, a hedge *par excellence* indicates the most central member of a category (as in *A robin is a bird par excellence*), while *sort of* points to a peripheral example of the category (as in *A penguin is sort of a bird*). What is more, *loosely speaking* does not merely exclude the central member but points to "things that would not ordinarily be considered members" (as in *Loosely speaking, a telephone is a piece of furniture*) (Lakoff 1973, Taylor 1995: 77). Such hedges as *loosely speaking* show that not only are category boundaries flexible, but they can also be redefined. These findings provide support for the conception of prototypically organized categories, whereby categories display degrees of typicality and their boundaries are blurred (Rosch 1973; Lewandowska-Tomaszczyk 2007: 145).

What bearings do the above findings have on defining practice? From the point of view of practical lexicography, hedges are convenient tools that make the definer's task much easier. Rather than describing all possible exemplars and features of the category of meaning being defined, which is often doomed to failure, the definer highlights what is typical of this category.² The former situation reflects the classical approach to meaning, which had dominated linguistic thought since Aristotle. This approach rests on the assumption that a meaning is identified through a fixed set of necessary and sufficient features. The same assumption underpins the construction of the traditional definition (*genus proximum plus differentia specifica*),³ which presupposes that the lexicon is organized hierarchically and that lexemes can be defined in discrete rather than continuous terms.⁴ If lexicographers adhere uncritically to the classical principle of defining, they end up with verbose and over-specific definitions⁵ (Atkins and Rundell 2008). In the attempt to search for necessary conditions for the use of a word, lexicographers are likely to obscure central meanings by marginal ones. Hedges do not merely emphasize the prototype but downtone potentially controversial features; for example, as Geeraerts (2006) points out, fruit is usually sweet, not sour like lemons.

Prototype theory is relatively new in the history of linguistics. It was developed by a cognitive psychologist Eleanor Rosch in the 1970s through a series of experiments on the internal structure of categories. Empirical evidence collected by Rosch and her colleagues attracted considerable attention of linguists. Since the early 1980s the theory has exerted a great impact on cognitive linguistics, which is reflected by the growing body of literature with prototypes

in the center of scholarly attention (e.g. Lakoff 1987; Langacker 1987; Taylor 1995; Geeraerts 1997; Lewandowska-Tomaszczyk 2007).

Prototype theory has recently drawn a great deal of attention from dictionary makers and researchers (Hanks 1994, Rey 1990, Swanepoel 1994, Van der Meer 1999 and 2000, Geeraerts 2006, Zgusta 2006). In a discussion of the relationship between prototypicality and lexicography, Geeraerts concludes that cognitive semantics offers "an exciting perspective for the further development of lexicography" (Geeraerts 2006: 363). It is also Zgusta (2006: 115) who notes that "prototype theory promises to be of great usefulness to lexicography". The *New Oxford Dictionary of English* (NODE-1998) is perhaps one of the first dictionaries to refer explicitly to cognitive linguistics as a basis for the description and presentation of meaning (cf. Geeraerts 2006, Hanks 1994). In the Preface, we read:

The *New Oxford Dictionary of English* is [...] informed by currently available evidence and current thinking about language and cognition. [...] Linguists, cognitive scientists, and others have been developing new techniques for analysing usage and meaning, and the *New Oxford Dictionary of English* has taken full advantage of these developments. Foremost among them is an emphasis on identifying what is 'central and typical' [...] The style of definition adopted for the *New Oxford Dictionary of English* aims in part to account for the dynamism, imaginativeness, and flexibility of ordinary usage. [...] The layout and organization of each entry in the dictionary reflect this new approach to meaning. Each entry has at least one core meaning, to which a number of subsenses, logically connected to it, may be attached. (Preface, NODE-1998)

In accordance with the prototypical conception of meaning, the NODE-1998 editor recognizes the fact that the semantic structure of a lexical unit is fuzzy and flexible, with clusters of meanings related to one or more prototypical cores. One may expect that lexicographers who adopt the radial structure model will use special techniques for highlighting prototypical elements and for indicating relationships between meanings. One of these techniques is the use of hedges.

For anyone familiar with the history of dictionaries, it is easy to see that hedges are by no means a novelty for contemporary dictionary writers and that they have been used in definitions long before the rise of prototype theory. However, considering the usefulness of prototype theory for lexicographers, it is likely that hedges are applied more consciously now than before the rise of this theory. It follows that the use of hedges might have intensified over the recent decades. This is the main research question that this paper attempts to address. Other points of interest are whether there are other factors determining the use of hedges (besides the rise of prototype theory), and whether English dictionaries display preferences for particular sets of hedges. This paper aims to examine the frequency and distribution of hedges in definitions in major English dictionaries published over the past centuries.

2. Method

2.1 Materials

The dictionaries under study were published between 1785 and 2011, a period of time covering a large part of the history of English monolingual lexicography:

- Johnson's *Dictionary of the English Language* (henceforth Johnson-1785);
- Webster's *American Dictionary of the English Language* (Webster-1865);
- *The Oxford English Dictionary on Historical Principles* (1888–1928) (OED-1928);
- *Chambers's Twentieth Century Dictionary* (Chambers-1952);
- *The Chambers Dictionary* (Chambers-2011);
- *Webster's Seventh New Collegiate Dictionary* (Collegiate-1963);
- *Merriam-Webster's Collegiate Dictionary* (Collegiate-2004);
- *The New Oxford Dictionary of English* (NODE-1998);
- *The Longman Dictionary of Contemporary English* (LDOCE-2005);
- *The Concise Oxford English Dictionary* (COD-2011)

An advantage of the above selection is that it includes the editions of the same dictionaries: two editions of Chambers and two editions of the Collegiate, published before and after the rise of prototype theory. This material was particularly worth exploring as it ensured the isolation of the effect of publication time on the use of hedges from other variables, notably dictionary type.

Another advantage of the above selection is that it represents the main lexicographic genres that emerged over the past three centuries: general-purpose dictionaries for native speakers, scholarly historical dictionaries for the same readership, and dictionaries for learners. Johnson-1785 was one of the first English dictionaries which aimed at comprehensiveness of vocabulary coverage and thoroughness of treatment. Unlike earlier dictionaries of the 17th century, it aimed to be a scholarly record of the entire language. Along with rare and obsolete words used in English literature, the dictionary covered ordinary everyday words, such as *make*, *do*, *book*, and their meanings. Johnson had a strong influence on subsequent dictionaries both in the United States and Britain, notably Webster-1864 and OED-1928 (Friend 1967; Hanks 2005; Landau 2009), which are also included in this study. Webster aimed to surpass Johnson by covering numerous terms of developing science, art, and technology. His definitions were more exhaustive than Johnson's, with a more elaborate sense division (Landau 2001: 70). OED-1928 is the only historical work in this selection, designed to serve students and scholars as a scholarly resource for research into the history of the English language. NODE-1998, COD-2011, the two editions of the Collegiate (1963 and 2004), and the two editions of Chambers (1952 and 2011), are relatively recent publications for a popular audience. They are single volume works designed to satisfy the needs of a wide range of native speakers especially when they arise while reading. As a dictionary for EFL

learners, LDOCE-2005 represents a relatively new genre in English lexicography. A distinguishing feature of this dictionary is that it uses in its definitions a restricted set of defining vocabulary of around 2000 words. Although the above selection is modest, it covers the dictionaries that have been extremely influential in both English (Johnson-1785, OED-1928) and American lexicography (Webster-1865). The dictionaries under study either shaped English lexicography or grew directly out of the long tradition of dictionary making. They vary considerably one from another with respect to the target audience, the publication date, the amount and type of information provided. All these factors may have an effect on the use of hedges.

2.2 Data preparation

Approximately 30 to 60 pages were drawn randomly from each dictionary.⁶ The page images were converted into text files using OCR software.⁷ The data were further processed by extracting definitions, and removing other parts of entries. The text was proofread and checked. To facilitate the identification of hedges, abbreviated forms such as *esp* and *usu*, which were conventionally used in some of the dictionaries, were expanded to their full forms. For the same purpose, some abbreviations were converted to their variant forms; for example *&c* into *etc.*, and *e.g.* into *eg.* The samples ranged in size from over 16,000 word tokens in the OED-1928, Johnson-1785, and LDOCE-2005 to 37,000 in the Collegiate-1963 (see Table 1), but in order to minimize the effect of different sample lengths, the hedge frequencies were normalized by 100,000. The comparison of two editions of the Collegiate seems to have been least affected by the size differences, as the samples were roughly comparable in length (37664 word tokens in Collegiate-1963 and 35019 in Collegiate-2004). A similar situation obtained in Chambers (23550 in Chambers-1952 and 28153 in Chambers-2011).

83 hedges were selected for the study, with more than half of them (48) being drawn from Lakoff (1973: 472).⁸ Due to the fact that Lakoff's list excludes a number of items that typically occur in definitions (e.g. *etc.*, *e.g.*, *also*, *or*, *such as*, *usually*), while including those that are unlikely to be found in this type of text (e.g. *in a manner of speaking*, *loosely speaking*, *mutatis mutandis*), the selection was expanded by 35 additional hedges. These were identified by perusing dictionaries, including those under consideration. Given the fact that it was often impossible to make arbitrary decisions regarding whether a hedge is representative of the defining language, the majority of Lakoff's hedges were retained,⁹ regardless of whether they were characteristic of definition style or not. The selection under study is by no means complete (see their grammatical variation in the Introduction), but it includes a large number of hedges that typically occur in definitions. The list is as follows:

"a real", "a regular", "a true", "actually", "all but", "all but a", "almost", "also", "anything but a", "approximately", "as", "as if", "as it were", "assumed", "basi-

cally", "believed", "broadly", "can be viewed as", "chiefly", "considered", "definitely", "eg", "especially", "essentially", "etc", "exceptionally", "for example", "for the most part", "generally", "in a manner of speaking", "in a real sense", "in a sense", "in a way", "in essence", "in one sense", "in particular", "kind of", "largely", "like", "likely", "literally", "look like", "looked upon", "looks like", "loosely speaking", "more or less", "mostly", "mutatis mutandis", "nominally", "occasionally", "often", "or", "par excellence", "particularly", "perhaps", "practically", "presumably", "pretty much", "principally", "quintessential", "rather", "really", "regarded", "relatively", "roughly", "seem", "seems", "so-called", "so to say", "sometimes", "somewhat", "sort of", "specifically", "strictly speaking", "such as", "technically", "tend", "tends", "typical", "typically", "usually", "very", "virtually"

After identification of the hedges in each dictionary sample, their frequencies were computed and normalized per 100,000 with the aid of a specially designed computer program.¹⁰

2.3 Data analysis

The hedges were analyzed with respect to the overall frequency of the tokens and the distribution of the types across the dictionaries. The results of the former analysis are presented in Figure 1. As for the hedge types, their distribution is presented in Table 2. For easier identification of the correlation between the dictionaries and the hedges, the data were visualized using a correspondence analysis (CA) plot. CA produces a pictorial representation of the variables (here dictionaries and the hedges) on the same set of axes (Figure 2). The plot reveals associations between different data points on the basis of their co-occurrence (Glynn 2014: 133). The association is greater for the points located in proximity to one another, and away from the center of the plot. CA was performed on the hedges with the highest overall frequency of occurrence; to that end, the hedges ranked 1–30 were taken into consideration (see Table 2).

The CA plot indicates the most distinctive associations between the hedges and the dictionaries, but without showing whether the associations are statistically significant. For a more fine-grained view of these relations, a one-sided binomial test was performed on the same data, that is the first 30 rows of Table 2. The goal of the test was to verify the hypothesis that the probability of obtaining the observed value of hedge frequency in each cell is significantly higher than chance probability.¹¹ The resulting p-values for this hypothesis are displayed in Table 3. The values lower than the significance level 0.05, which are highlighted in grey, indicate that the frequencies are statistically higher than those expected by chance.

Figure 1: Overall frequencies of hedge tokens in the samples (normalized per 100,000)

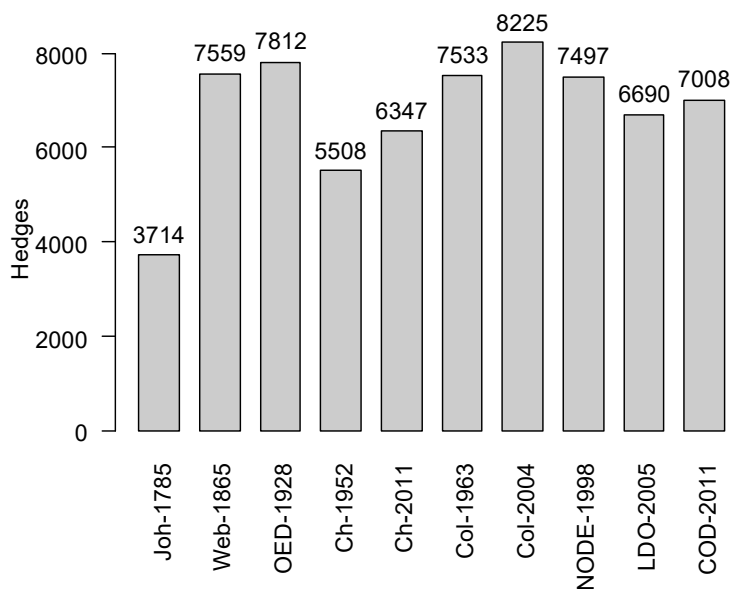


Table 1: The size of the samples and percentage of hedges

	Johnson-1785	Webster-1865	OED-1928	Chambers-1952	Chambers-2011	Collegiate-1963	Collegiate-2004	NODE-1998	LDOCE-2005	COD-2011
Sample size (in word tokens)	16276	32685	16134	23550	28153	37664	35019	29444	16737	20530
Percentage of hedges	3.7	7.5	7.8	5.5	6.3	7.5	8.2	7.5	6.6	7.0

Table 2: Normalized frequencies of hedge types in the samples

		Joh-1785	Web-1865	OED-1928	Ch-1952	Ch-2011	Col-1963	Col-2004	NODE-1998	LDO-2005	COD-2011
1	or	2255	4497	5014	3210	3694	4649	4532	5149	3866	4944
2	as	725	1787	862	769	586	1062	1471	672	341	653
3	especially	6	132	310	310	469	605	791	567	418	438
4	etc	0	128	359	327	668	0	0	31	633	117
5	also	49	162	359	59	60	244	371	48	0	49

6	like	55	285	217	268	170	45	23	54	102	97
7	usually	6	76	130	72	89	287	411	34	102	15
8	very	18	37	25	81	60	45	49	95	544	83
9	often	43	58	136	42	67	114	183	54	78	44
10	typically	6	0	0	4	7	27	46	282	0	156
11	sometimes	129	70	56	68	25	24	51	20	6	0
12	kind of	80	104	62	64	36	5	3	17	18	10
13	such as	37	3	12	8	60	3	0	71	114	73
14	eg	0	0	68	38	167	0	0	20	0	58
15	chiefly	6	31	74	13	4	48	51	54	0	63
16	specifically	0	21	0	0	4	138	77	0	0	0
17	generally	55	18	37	42	14	11	9	3	6	0
18	rather	37	3	0	8	21	16	17	24	42	24
19	considered	0	18	12	13	18	8	11	34	42	5
20	regarded	0	12	6	0	11	8	6	34	6	44
21	in particular	0	0	0	0	0	0	0	109	0	10
22	likely	0	0	0	0	4	21	6	17	42	0
23	for example	0	0	0	0	0	0	0	14	66	5
24	in a way	0	6	0	0	0	3	3	7	60	5
25	somewhat	6	15	25	8	4	19	0	3	0	0
26	seems	55	3	0	0	4	3	0	0	6	5
27	typical	0	0	6	13	4	8	9	3	18	15
28	perhaps	43	6	0	0	14	8	0	0	0	0
29	almost	18	6	0	4	7	8	0	3	6	15
30	broadly	0	0	0	4	0	42	20	0	0	0
31	really	0	0	0	8	4	0	0	0	48	0
32	particularly	18	15	0	0	4	0	0	17	0	5
33	relatively	0	6	0	0	0	5	26	10	0	10
34	actually	0	3	12	0	4	3	6	10	12	5
35	believed	0	0	0	13	0	5	0	3	18	15
36	more or less	6	3	12	13	0	5	3	7	0	0
37	seem	0	0	0	0	0	3	0	3	42	0
38	assumed	0	12	6	4	0	8	14	3	0	0
39	sort of	31	12	0	0	4	0	0	0	0	0

74	looked upon	0	0	0	0	0	0	0	0	0	0
75	loosely speaking	0	0	0	0	0	0	0	0	0	0
76	mutatis mutandis	0	0	0	0	0	0	0	0	0	0
77	nominally	0	0	0	0	0	0	0	0	0	0
78	par excellence	0	0	0	0	0	0	0	0	0	0
79	presumably	0	0	0	0	0	0	0	0	0	0
80	pretty much	0	0	0	0	0	0	0	0	0	0
81	quintessential	0	0	0	0	0	0	0	0	0	0
82	so to say	0	0	0	0	0	0	0	0	0	0
83	strictly speaking	0	0	0	0	0	0	0	0	0	0

Figure 2: Correspondence Analysis: Associations between dictionaries and hedges

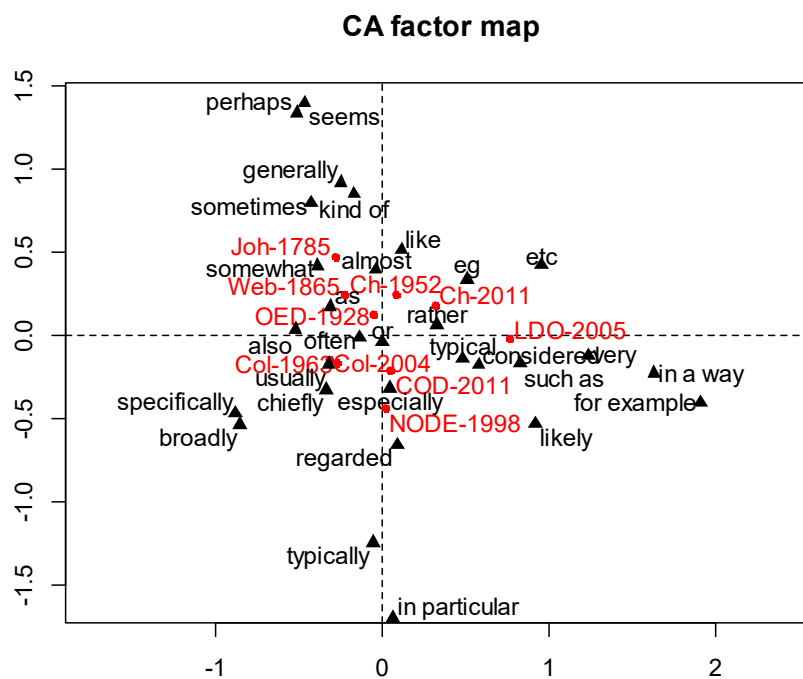


Table 3: p-values for the binomial test calculated on the 30 most frequent hedges¹²

		Joh-1785	Web-1828	OED-1928	Ch-1952	Ch-2011	Col-1963	Col-2004	NODE-1998	LDO-2005	COD-2011
1	or	0.58	1	0.01	1	1	0.5	1	0	1	0
2	as	0	0	1	0.04	1	0.01	0	1	1	1
3	especially	1	1	1	0.85	0	0	0	0	0.1	0.16
4	etc	1	1	0	0	0	1	1	1	0	1
5	also	1	0.33	0	1	1	0	0	1	1	1
6	like	0.98	0	0	0	0	1	1	1	0.99	1
7	usually	1	1	0.86	1	1	0	0	1	0.96	1
8	very	1	1	1	0.65	1	1	1	0.98	0	1
9	often	0.6	1	0	1	0.89	0.01	0	1	0.59	1
10	typically	1	1	1	1	1	1	1	0	1	0
11	sometimes	0	0	0.3	0	1	1	0.72	1	1	1
12	kind of	0	0	0.01	0	0.61	1	1	1	1	1
13	such as	0	1	1	1	0	1	1	0	0	0
14	eg	1	1	0	0.04	0	1	1	1	1	0
15	chiefly	1	0.91	0	1	1	0.06	0.08	0.01	1	0
16	specifically	1	0.9	1	1	1	0	0	1	1	1
17	generally	0	0.83	0	0	0.88	1	1	1	1	1
18	rather	0	1	1	0.99	0.26	0.91	0.94	0.3	0	0.19
19	considered	1	0.53	0.97	0.55	0.25	1	0.99	0	0	1
20	regarded	1	0.77	1	1	0.65	0.98	1	0	0.99	0
21	in particular	1	1	1	1	1	1	1	0	1	0.8
22	likely	1	1	1	1	0.97	0	0.97	0.02	0	1
23	for example	1	1	1	1	1	1	1	0.08	0	0.95
24	in a way	1	0.92	1	1	1	1	1	0.83	0	0.94
25	somewhat	0.27	0.03	0	0.32	0.95	0	1	0.99	1	1
26	seems	0	0.99	1	1	0.93	0.99	1	1	0.76	0.9
27	typical	1	1	0.89	0.01	0.93	0.62	0.58	0.99	0	0.01
28	perhaps	0	0.82	1	1	0.01	0.54	1	1	1	1
29	almost	0	0.77	1	0.8	0.44	0.47	1	0.98	0.64	0
30	broadly	1	1	1	0.79	1	0	0	1	1	1

3. Results and discussion

The first point to make is that the hedges occur in the definitions in a relatively high proportion, which ranges from 3.7% in Johnson-1785 to 8.2% in Collegiate-2004 (see Table 1). As shown in Figure 1, they are used most frequently in the Collegiate (both editions), the OED, and Webster, and least frequently in Johnson and Chambers. The second point is that they occur statistically more frequently in the recent edition of the Collegiate than in the one from 1963 (8225 word tokens in Collegiate-2004 vs. 7533 in Collegiate-1963).¹³ A similar finding emerges from the comparison of the two editions of Chambers (6347 word tokens in Chambers-2011 and 5508 in Chambers-1952).¹⁴

Analyzing the data in Tables 2 and 3, one finds that there is a great deal of variation in the use of hedges across the dictionaries. Although they are distributed rather unevenly in the dictionaries, the texts display preferences for particular sets of hedges. Johnson relatively rarely used hedges (3.7%), but it does not mean that he dispensed with them altogether. He arguably preferred *sometimes*, *kind of*, *sort of*, *such as*, *generally*, *rather*, *seems/seem*, *perhaps*, *almost*. They are characteristic of Johnson's defining style, as they occur significantly more often than one would expect by chance (see Table 3). Interestingly, Johnson appears to have a tendency to use hedges that point to less typical examples of a category (*sometimes*) or those extending the category boundary (*rather*, *almost*). For example, *sometimes* introduces secondary senses and less typical grammatical and semantic properties of words:

- farewell** The parting complement; adieu ... It is sometimes used only as an expression of separation without kindness.
- hight** ... 4. It is sometimes used as a participle passive: called; named.
- none** ... 4. *None of* sometimes signifies only emphatically *nothing*.
- virtue** ... 5. Efficacy ; power. Before *virtue* is used sometimes *by* and sometimes *in*

Looking at Table 2, one notices that *especially*, *typically* / *typical*, *usually*, which have become a standard feature of the metalanguage of many contemporary dictionaries, are rarely used in Johnson. According to Hanks (2005, 265), the lack of "and the like", "any of various", and "etc." should be attributed to Johnson's ability to "write pointed, no-nonsense definitions, seizing again and again on the central point."

In quantitative terms, Webster stands in contrast to Johnson, as the former has twice as many hedges as the latter (7559 and 3714, respectively) (Figure 1). The reasons lie in Webster's earlier edition of 1828, which had borrowed heavily from Johnson, using the latter as a "working base" for constructing the definitions (Friend 1967, 44-5). In this edition, a number of definitions were copied from Johnson verbatim or revised by adding more details on the basis of Johnson's examples. Webster's attention to details manifests itself by a more fine-grained division of senses¹⁵ and provision of typical contextual uses of a word

being defined. Roughly speaking, what Johnson provides in illustrative quotations Webster (1828) includes in definitions following the hedge *as*:

Johnson-1785: **nice** ... 6. Requiring scrupulous exactness.

Supposing an injury done, it is a nice point to proportion the reparation ...

L'Estrange

Webster-1828: **nice** ... 4. Requiring scrupulous exactness; as a *nice* point.

The 1864 edition of Webster, which is under study¹⁶, was revised substantially, but the use of *as* introducing typical examples survived:

Webster (1865): **nice** ... 3. ... to be scrupulously and exactly handled; delicate; refined; as, a *nice* distinction of point.

Worth noting is that *as* is distinctively associated with both Webster and its recent descendant, the Collegiate (see Table 2).

The dictionary that comes second in the ranking (Figure 1) is the OED. One of the reasons for this dictionary being rich in hedges is certainly the fact that it is a historical dictionary. The methodology and the purpose of compiling historical dictionaries make hedges inevitable. Historical lexicographers examine corpus evidence and peruse countless uses of a lexical unit in order to determine one or more prototypical core senses and then possible lines of development of secondary and peripheral senses (see also Kay 2000). The aim is to give a comprehensive account of the history of lexical units by uncovering the full path of their semantic change. A fully thorough treatment of meaning is impossible to accomplish when the description is to be rendered in terms of necessary and sufficient conditions. For one thing, indeterminacy of meaning causes demarcation problems, which can only be solved by focusing on what is typical in the semantic structure. It is more convenient to use hedges than to attempt to enumerate the entire range of category members or features. This strategy can be seen in the definition of **head** below, which mentions *ale* and *beer* as typical examples of liquor that bears a *head*, while omitting *palm-wine* which is nevertheless recorded in one of the citations (see below).

head 10. A collection of foam or froth on the top of liquor, esp. ale or beer.

... 1760–72 ... *Palm-wine* .. bears a greater head than beer, and is of a very inebriating quality (OED)

The frequent use of hedges is a distinctive characteristic of the definition style, which is especially noticeable in the dictionaries that aim to serve as a historical record of the language. They aim to describe the entire semantic structure of lexical units by providing dominant and peripheral senses, along with less frequent and more specialized readings. Hedges along with an elaborate sense division (e.g. I, II, III, 1, 2, 3, a, b, c), albeit a linear one, are lexicographic means of capturing the prototype (Geeraerts 2006, 357; Zgusta 2006, 114).

While Johnson uses relatively few hedges, Collegiate-2004 resorts to them

most often. The fact that it has the largest total number of hedges is arguably due to the developed system of sense division, with hedges being used as sense dividers. Along with Arabic numerals, letters, parenthesized numerals, and colons, which indicate various levels of the hierarchy of senses, the dictionary uses *esp(pecially)*, *specif(ically)*, *also*, and *broadly* to indicate "a particular semantic relationship" between (sub-)senses (The Collegiate 2003, 20a). In addition, it employs a hedge *as* to indicate "that the following subsenses are typical or significant examples" (*ibid.*), for example:

- billow** ... 1 : WAVE; *esp* : a great wave or surge of water 2 : a rolling mass (as of flame or smoke) that resembles a high wave
- billy club** ... a heavy usu. wooden club; *specif* : a police officer's club
- binary** ... something made of or based on two things or parts: as **a** : binary star **b** : a binary number system
- outgas** 1 : to remove occluded gases from usu. by heating; *broadly* : to remove gases from
- putting green** a smooth grassy area at the end of a golf fairway containing the hole; *also* : a similar area usu. with many holes that is used for practice

As Table 3 demonstrates, the above hedges occur in both editions of the Collegiate significantly more often than chance would predict. They are distinctively associated with this dictionary. In addition to the hedges indicating relations between particular senses, the Collegiate (in both editions) displays a significant preference for *usu(ally)* and *often* as indicators of typicality within definitions (e.g. **aspersion** "a sprinkling with water *esp.* in religious ceremonies" Collegiate-2004). Neither of the above uses is original in lexicography, as they can be traced back to earlier dictionaries, notably Webster, of which the Collegiate is a remote descendant.¹⁷ Nevertheless, as far as sense-division is concerned, the Collegiate has exploited hedges to a remarkable extent.

Of all the popular dictionaries, Chambers uses the least number of hedges. One of the reasons for this is a marked tendency to use brief definitions, a strategy resulting from the policy of all-inclusiveness. Brief definitions, often in the form of a string of near-synonyms, which are a distinctive feature of Chambers, leave little room for prototypical information. The comprehensive coverage of meanings, which Chambers aims at, stems from the desire to make the dictionary useful in reading a vast range of texts. Chambers' aim of comprehensiveness is pursued by the coverage of literary words and senses used by Shakespeare, Spenser, and Milton, as well as the vocabulary pertaining to non-English varieties of English (e.g. Scottish English). The need to provide the user with peripheral, rare, and less predictable meanings outweighs the need to give the details of prototypical meanings, which the native user is familiar with (cf. Geeraerts 2006: 330). Yet the two editions of Chambers have their own distinctive set of hedges: *etc*, *like*, and *e.g.* occur significantly more frequently in this dictionary than one may expect by chance (see Table 3). Unlike the Collegiate, Chambers

does not use hedges as markers of the hierarchy of senses.¹⁸

The other popular dictionaries, COD-2011 and NODE-1998, share a few characteristic hedges: *or, typically, such as, chiefly, regarded* (see Table 3). These are arguably traces of the same methodology of preparing definitions; note that both dictionaries were compiled by the same editor and that the former dictionary drew heavily on the latter (Preface to COD-2011). However, NODE-1998 has a few other hedges that are distinctively associated with this dictionary: *especially, considered, likely, and in particular*. The last expression is noteworthy, as it is used systematically to introduce specialized cases of a core sense, as in:

- crossbar** ... a horizontal bar fixed across another bar or between two upright bars, in particular:
- the bar between the two upright posts of a football goal.
 - the horizontal metal bar between the handlebars and saddle on a man's or boy's bicycle.

It is the only hedge used in NODE-1998 consistently to specify the type of relationship that holds between a core sense and one or more sub-senses.

As the only dictionary for foreign learners, LDOCE-2005 has its own distinctive set of hedges, which occupy the most extreme positions in the right section of Figure 2, relatively close to this dictionary. It is clear that the dictionary displays a marked preference for *etc., very, such as, rather, considered, likely, for example, in a way, typical, seem, definitely* (see also Tables 3 and 2). These words and expressions belong to the restricted defining vocabulary, which is intended to make the definitions accessible to the learner. They were selected largely according to the criterion of frequency and range in the language, in order to ensure that their form and meaning is straightforward and leaves little room for misinterpretation. This policy differs radically from that employed in some native-speaker dictionaries which permit words in peripheral meanings, for instance, while LDOCE frequently uses *for example* to introduce prototypical examples, Johnson, Webster, and the Collegiate consistently use *as* to this end.

Finally, it is worth noting that 25% of the hedges (i.e. 21 hedge types) under consideration are not used in any definition (see Table 2). This can be explained by the fact that some of them, such as *mutatis mutandis, par excellence, quintessential, in a manner of speaking* are too formal to be admitted into definitions in a systematic way. In turn, *pretty much* and *so to say* are excluded due to their informality.

4. Conclusions

The foregoing analysis suggests several conclusions. Firstly, the fact that hedges are used more extensively in the recent editions of the popular dictionaries (i.e. Chambers-2011 and the Collegiate-2004) than in the editions published prior to the 1970s suggests that contemporary defining practices in these

dictionaries must have been inspired by prototype theory. Being part of the defining metalanguage, they are imposed on lexicographers by editors through dictionary style manuals. Their use is more conscious and controlled now than in the past, with the aid of dictionary writing systems, which help the editor manage the production of a dictionary and execute consistently the dictionary policy (Abel 2012). Besides being used as indicators of prototypical elements of meaning within a definition, hedges are applied to highlight sense relationships. While the former use has been widespread in the English dictionaries, the latter is not.

Prototype theory, however, is not the only factor determining the use of hedges. The analysis of the sample shows that while the intensification of the use of hedges in the recent popular dictionaries should be attributed to the influence of this theory on dictionary editors, the high rate of hedges in early dictionaries (notably the OED-1928 and Webster-1865) suggests that other factors must be at play. A few aspects relevant to dictionary function (Tarp 2008) emerge as key factors: the depth of treatment, dictionary type, and defining policy. Hedges are particularly useful in large-scale dictionaries compiled on historical principles (such as the OED), as the full account of the history of word meanings would be impossible to achieve without them. They are also useful in desk-size dictionaries that aim at thoroughness of treatment (cf. Geeraerts 1997). Their frequent use in Webster-1865 results partly from the systematic strategy of combining Johnson's definitions and examples by means of "as". On the other hand, they are less common in dictionaries that rely heavily on brief and synonymous definitions (such as Chambers). Other factors such as individual lexicographers' preferences and the stylistic guidelines as specified in dictionary style manuals cause that different dictionaries show marked preferences for different sets of hedges.

Finally, it is hardly surprising that hedges have been used more or less extensively in the English dictionaries since Johnson-1785. Arguably, the well-established use of hedges in English lexicography results from their usefulness, if not necessity: they enable the definer to achieve what otherwise would be difficult to achieve: economy of expression, as well as reliability and thoroughness of description.

Endnotes

1. John Taylor mentions examples of adjectives (*true, real*), suffixes (*-ish*), conjunctions (*in that, and, or*), certain lexical and auxiliary verbs (*seem, appear, may, can*), and even inverted commas (1995: 76).
2. However, the overuse of hedges in definitions has been criticized by Anna Wierzbicka, who considers them as "visible signs of indecision and analytical failure" (1996: 269).
3. The former specifies the class to which an object being defined belongs, and the latter, the features distinguishing the object from other members of the class (Svensén 1993).

4. While the classical definition is effective in defining large sections of the lexicon, it is clearly not suitable for those areas of lexis which display taxonomic gaps (Adamska-Salaciak 2012).
5. A much-quoted definition of **door** in *Webster's Third International* (1961) occupies 11 lines of text (Atkins and Rundell 2008).
6. The sampling was aided with an R function called *sample*, which allowed for a random selection of page numbers. Initially, 30 pages from each dictionary were drawn. However, due to the different density of the dictionaries texts, further pages were selected from the dictionaries with the lower text density in order to reduce differences in sample sizes. The error caused by the different sample sizes was further minimized by the normalization of hedge frequencies.
7. The images were processed with ABBYY FineReader 11.
8. Lakoff's list (1973: 472) has 67 items collected under the heading "Some hedges and related phenomena".
9. Among the hedges excluded from the study were those that were computationally intractable (e.g. "he's another (Caruso/Lincoln/ Babe Ruth/ ...)") and those likely to produce false hits (e.g. a suffix "-ish", which is often used in the nationality sense, as in *English*).
10. The program was written in R, an open source programming language (R Development Core Team 2013).
11. The binomial test was conducted with the *binom.test* function in R, with the *alternative* argument set to "greater" for a one-sided test. The test calculated for each cell the probability of the observed value given the expected probability. The latter was computed by multiplying the column frequency total by the row frequency total and dividing the result by the total frequencies in the table (for the rows 1-30) (see also Gries and Stefanowitsch 2006; Janda 2013; Stefanowitsch and Gries 2003).
12. The values equal 0 or lower than the significance level 0.05 indicate that the frequencies are statistically higher than those expected by chance.
13. The difference is statistically significant, as a chi-square test gives $p < 0.05$, $\chi^2 = 30.389$, $df = 1$.
14. A chi-square test gives $p < 0.05$, $\chi^2 = 59.378$, $df = 1$.
15. For example, while Johnson provides one sense at **hospitable**, Webster distinguishes three. At **nice**, there are 9 and 14 senses, respectively.
16. Although the current study was conducted on the 1865 version of Webster, it is the same edition as that published in 1864.
17. The Collegiate series derives from the unabridged *Webster's International Dictionary* (1890) (and its revisions), which in turn is a direct descendant of Webster's famous revision of 1864.
18. In Chambers-1952 sense arrangement generally adheres to the historical principle, whereby the first definition describes the "original" or early meaning, and the following ones cover a series of secondary meanings "branching out and diverging from" the original one (Chambers-1952: vii). Although the linear presentation of senses, separated by colons, is a practical space-saving strategy, it creates a simplified picture of sense relations, implying that secondary meanings derive directly from one and the same etymological root. A more sophisticated presentation of senses, arguably inspired by prototype theory, is found in COD-2011. The dictionary organizes senses according to typicality, providing more than one core sense and subsenses deriving from the core.

References

- Abel, A. 2012. Dictionary Writing Systems and Beyond. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 83-106. Oxford: Oxford University Press.
- Adamska-Sałaciak, A. 2012. Dictionary Definitions: Problems and Solutions. *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 129(4): 7-23.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Friend, J.H. 1967. *The Development of American Lexicography 1798–1864*. The Hague: Mouton.
- Geddie, W. 1952. *Chambers's Twentieth Century Dictionary*. Second edition. Edinburgh: W. & R. Chambers.
- Geeraerts, D. 1997. *Diachronic Prototype Semantics: A Contribution to Historical Lexicology*. Oxford: Clarendon Press.
- Geeraerts, D. 2006. The Definitional Practice of Dictionaries and the Cognitive Semantic Conception of Polysemy. Geeraerts, D. (Ed.). 2006. *Words and other Wonders. Papers on Lexical and Semantic Topics*: 345-363. Berlin: Mouton de Gruyter.
- 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.
- Gove, P. (Ed.). 1963. *Webster's Seventh New Collegiate Dictionary*. Seventh edition. Springfield, Mass.: G. & C. Merriam. (Collegiate-1963)
- Gries, S.Th. and A. Stefanowitsch (Eds.). 2006. *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin: Walter de Gruyter.
- Hanks, P. 1994. Lexicography: Theory and Practice. *Dictionaries* 14(1992/3): 97-112.
- Hanks, P. 2005. Johnson and Modern Lexicography. *International Journal of Lexicography* 18(2): 243-266. <https://doi.org/10.1093/ijl/eci024>.
- Hyland, K. 1994. Hedging in Academic Writing and EAP Textbooks. *English for Specific Purposes* 13(3): 239-256. [https://doi.org/10.1016/0889-4906\(94\)90004-3](https://doi.org/10.1016/0889-4906(94)90004-3).
- Hyland, K. 2006. Medical Discourse: Hedges. Brown, K. (Ed.). 2006. *Encyclopedia of Language and Linguistics*: 694-697. Second edition. Oxford: Elsevier.
- Janda, L.A. 2013. Quantitative Methods in Cognitive Linguistics: An Introduction. Janda, L.A. (Ed.). 2013. *Cognitive Linguistics: The Quantitative Turn*. Berlin: De Gruyter Mouton.
- Johnson, S. 1785. *A Dictionary of the English Language*. [Vol. 1 and 2]. London: J.F and C. Rivington. (Johnson-1785).
- Kay, J.C. 2000. Historical Semantics and Historical Lexicography: Will the Twain Ever Meet? Coleman, J. and C.J. Kay (Eds.). 2000. *Lexicology, Semantics and Lexicography*: 53-68. Amsterdam: John Benjamins.
- Lakoff, G. 1973. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2(4): 485-508.
- Lakoff, G. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Landau, S.I. 2001. *Dictionaries: the Art and Craft of Lexicography*. Second edition. New York/ Cambridge: Cambridge University Press.
- Landau, S.I. 2009. Major American Dictionaries. Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography. Vol. 1*: 182-229. Oxford: Oxford University Press.

-
- Langacker, R.W.** 1987. *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Lewandowska-Tomaszczyk, B.** 2007. Polysemy, Prototypes, and Radial Categories. Geeraerts, D. and H. Cuyckens (Eds.). 2007. *The Oxford Handbook of Cognitive Linguistics*: 139-169. Oxford: Oxford University Press.
- Marr, V. (Ed.)**. 2011. *The Chambers Dictionary*. Twelfth edition. London: Chambers Harrap. (Chambers-2011)
- Mish, F. (Ed.)**. 2004. *Merriam-Webster's Collegiate Dictionary*. Eleventh edition. Springfield, Mass.: Merriam-Webster. (Collegiate-2004)
- Murray, J.A.H. (Ed.)**. 1888–1928. *The Oxford English Dictionary on Historical Principles*. Vol. 1–10. [Edited by James Murray, Henry Bradley, William A. Craigie, and Charles. T. Onions]. Oxford: Clarendon Press. (OED-1928)
- Pearsall, J. (Ed.)**. 1998. *The New Oxford Dictionary of English*. Oxford: Clarendon Press. (NODE-1998)
- R Development Core Team**. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rey, A.** 1990. Definitional Semantics: Its Evolution in French Lexicography. Tomaszczyk, J. and B. Lewandowska-Tomaszczyk (Eds.). 1990. *Meaning and Lexicography*: 43-55. Amsterdam/Philadelphia: John Benjamins.
- Rosch, E.** 1973. Natural Categories. *Cognitive Psychology* 4: 328-350.
- Rosch, E.** 1975. Universals and Cultural Specifics in Human Categorisation. Brislin, R.W., S. Bochner and W.J. Lonner (Eds.). 1975. *Cross-cultural Perspectives on Learning*: 177-206. New York: Halstead/Wiley.
- Stefanowitsch, A. and S.Th. Gries**. 2003. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.
- Stevenson, A. and M. Waite (Eds.)**. 2011. *Concise Oxford English Dictionary*. Twelfth edition. Oxford: Oxford University Press. (COD-2011)
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Swanepoel, P.** 1994. Problems, Theories and Methodologies in Current Lexicographic Semantic Research. Martin, W. et al. (Eds.). 1994. *Euralex 1994. Proceedings (Beiheft)*: 11-26. Amsterdam: Vrije Universiteit.
- Summers, D. et al. (Eds.)**. 2005. *Longman Dictionary of Contemporary English*. Fourth Edition with Writing Assistant. Harlow: Pearson Education. (LDOCE-2005)
- 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.
- Taylor, J.R.** 1995. *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Clarendon Press.
- Van der Meer, G.** 1999. Metaphors and Dictionaries: The Morass of Meaning, or How to Get Two Ideas for One. *International Journal of Lexicography* 12(3): 195-208. <https://doi.org/10.1093/ijl/12.3.195>.
- Van der Meer, G.** 2000. Core, Subsense, and the *New Oxford Dictionary of English* (NODE). On How Meanings Hang Together, and Not Separately. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 419-431. Stuttgart: Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.

Webster, N. 1828. *An American Dictionary of the English Language*. New York: S. Converse. (Webster-1828)

Webster, N. 1865 [1864]. *An American Dictionary of the English Language*. [Revised by C.A. Goodrich and N. Porter.] Springfield, Mass.: Merriam. (Webster-1865)

Wierzbicka, A. 1996. *Semantics. Primes and Universals*. Oxford: Oxford University Press.

Zgusta, L. 2006. *Lexicography Then and Now. Selected Essays*. [Edited by Fredric S.F. Dolezal and Thomas B.I. Creamer]. Tübingen: Max Niemeyer.

The Lexicographic and Lexicological Aspects of a Web- Based Chrestomathy of Gothic and Anglo-Saxon Written Records

Tinatin Margalitadze, *Lexicographic Centre, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia* (tinatin.margalitadze@tsu.ge)

and

George Meladze, *Lexicographic Centre, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia* (meladzegeorge@yahoo.com)

Abstract: There is a general lack of web-based tools for morphologically complex dead/old languages. Reading texts in such languages even with dictionaries is quite challenging. It is difficult to identify the lemma of a word form occurring in texts, which one could look up in a dictionary. The need for additional grammatical information about a word (classes of declension, conjugation, etc.) poses another problem.

The Lexicographic Centre at Ivanè Javakhishvili Tbilisi State University (TSU) has embarked on creating a fully digitalized, web-based chrestomathy of Gothic and Anglo-Saxon texts with dictionaries and grammatical paradigms integrated in it, which would facilitate the study of these linguistically important languages. Each word of the digital versions of Gothic and Anglo-Saxon texts is hyperlinked to the corresponding headword from the dictionary. The dictionary entry itself, in addition to the meaning of the word, provides via another hyperlink all necessary information concerning the morphological class and inflectional patterns of the word in question.

The paper describes the structure of the Chrestomathy and its *modus operandi*; analyses the dictionary component of the online resource and some lexicographic solutions; discusses lexicological and technical aspects of the online resource, etc.

The method applied in the Chrestomathy can be successfully used in developing similar resources for extant, morphologically complex languages characterized with the abundance of inflectional and suppletive forms, such as Hungarian, Turkish, Russian, German, Georgian and many others.

Keywords: DIGITAL HUMANITIES, GOTHIC, ANGLO-SAXON, COMPARATIVE LINGUISTICS, SYNCHRONY AND DIACHRONY, SEMANTIC EQUIVALENCE

Opsomming: Die leksikografiese en leksikologiese aspekte van 'n webgebaseerde chrestomatie van Gotiese en Angel-Saksiese geskrewe rekords. Daar is 'n algemene tekort aan webgebaseerde hulpmiddels vir morfologies komplekse dooie/antieke tale. Selfs met behulp van woordeboeke is dit redelik uitdagend om tekste in hierdie tale te lees. Dit is moeilik om die lemma van 'n woordvorm wat in die tekste voorkom, te identifiseer sodat dit in 'n woordeboek nageslaan kan word. Die behoefte aan addisionele grammatikale inligting van 'n

woord (tipes verbuiging, vervoeging, ens.) skep weer 'n ander probleem.

Die Leksikografiese Sentrum by Ivanè Javakhishvili Tbilisi Staatsuniversiteit (TSU) het begin met die skep van 'n volledig gedigitaliseerde, webgebaseerde chrestomatie of studiehulp van Gotiese en Angel-Saksiese tekste met woordeboeke en grammatikale paradigmas daarin geïntegreer, wat die studie van hierdie taalkundig belangrike tale sal vergemaklik. Elke woord van die digitale weergawes van Gotiese en Angel-Saksiese tekste is deur hiperskakels verbind met die ooreenstemmende trefwoorde in die woordeboek. Buiten die betekenis van die woord verskaf die woordeboekinskrywing self via 'n ander hiperskakel al die noodsaaklike inligting rakende die morfologiese klas en verbuigingspatrone van die tersaaklike woord.

Hierdie artikel beskryf die struktuur van die chrestomatie en die *modus operandi* daarvan; analiseer die woordeboekkomponent van die aanlynbron en sommige leksikografiese oplossings; bespreek leksikologiese en tegniese aspekte rakende die aanlynbronne, ens.

Die metode wat toegepas word in die chrestomatie kan suksesvol aangewend word in die ontwikkeling van soortgelyke bronne vir lewende, morfologies komplekse tale, wat gekenmerk word deur talle verbuigings- en suppletiewe vorme, soos Hongaars, Turks, Russies, Duits en nog vele ander.

Sleutelwoorde: DIGITALE GEESTESWETENSKAPPE, GOTIES, ANGEL-SAKSIES, VERGELYKENDE LINGUISTIEK, SINCHRONIE EN DIACHRONIE, SEMANTIESE EWIVALENSIE

1. Introduction

The article provides a researched-based and theoretically-engaging presentation of the online Chrestomathy which was executed at the Lexicographic Centre of TSU with the funding of Shota Rustaveli National Science Foundation of Georgia (Grant No. FR17_87).

The Chrestomathy is conceived as a web-based resource for studying the Gothic and Anglo-Saxon (Old English) languages and represents the rapidly developing field of Digital Humanities (DH).

The Chrestomathy is intended as a complex digital resource which can considerably facilitate the study of aforesaid languages, so important for general linguistics, German, English and Scandinavian studies. The electronic online resource comprises dictionaries, as well as Gothic and Anglo-Saxon (Old English) texts represented in their electronic, programmatically processed format.

The texts included in the online Chrestomathy are programmatically integrated with Gothic-Georgian/Gothic-English and Anglo-Saxon-Georgian/Anglo-Saxon-(Modern) English dictionaries and with the morphological paradigms of Gothic and Anglo-Saxon words, i.e. electronic grammatical tables demonstrating the patterns of inflection (declension, conjugation, etc.) from the aforesaid two languages. The entire resource has its Georgian and English versions, making them accessible to the individuals interested in philology, linguistics and, especially, in Germanic studies in general, or, in Indo-European

studies and comparative linguistics in particular both in Georgia and (thanks to the international status of English) in the entire world.

This is the brief introduction to the online Reader and its aims. A more detailed analysis of its composition, as well as its lexicographic, lexicological and technical aspects will follow.¹

2. Why Old/Dead Languages and Why Specifically Gothic and Anglo-Saxon?

Gothic is a so-called dead language from the East Germanic subgroup of languages. The language is preserved in the form of a few surviving manuscripts (mainly fragments from Gothic gospels). Gothic is not directly ancestral to any extant language, but due to various circumstances, it has unique importance both for the diachronic study of modern Germanic and Indo-European languages and comparative or general linguistics.

By a number of its morphological and lexical features, Gothic is a very archaic language, in some aspects even more archaic than runic Norse (e.g. the gradation of root-forming vowels: **u/-au-** [sunus *vs* sunaus, sunau] or **-i/-ai-** [*mahtis, mahtim *vs* mahtais, mahtai] in Gothic -u- stem and -i- stem nouns, absent in runic Norse; Krause 1951). This very archaism makes Gothic immensely valuable to linguists. Gothic, as it seems, is very similar to Common Germanic (Proto-Germanic) parent language. One could state that Gothic may be regarded as one of the surviving (in the form of manuscripts, of course) dialects of presumed Proto-Germanic *Ursprache*. To certain degree, the relation of Gothic to modern Germanic languages can be compared to the relation of scantily (unlike Gothic) preserved Umbrian or Oscan to modern Romance languages.

The aforesaid archaic nature of Gothic makes its linguistic affinity (and, consequently, that of other Germanic languages) especially conspicuous with Latin, Lithuanian, Sanskrit and other Indo-European languages. Mastering Gothic language incredibly broadens the linguistic thinking of a scholar. Knowing Gothic and other old languages is scientifically much more advantageous than acquiring general information about linguistic affinity in scientific papers or etymological dictionaries. Immediate familiarity with this unique and archaic language whose written records are still available, gives a scholar some very clear idea of the peculiarities of the functioning and development of language as such both on synchronic and diachronic levels. We dare say that the knowledge of Gothic and other ancient Indo-European languages (such as Greek, Latin, Lithuanian, Sanskrit, etc.) is as necessary for a linguist, as the knowledge of human anatomy is for a physician or a surgeon. Without such knowledge one cannot form an adequate understanding of the processes defining how language functions and develops.

This is why Gothic is taught at the institutions of higher education of many countries, such as in Leiden University (the Netherlands), University of Copenhagen (Denmark) and other universities. There is an important centre for

the research on Gothic at Uppsala University (Sweden), where *Codex Argenteus*, the most valuable surviving record of the Gothic language is kept and where the computerized database of Gothic texts was created. In the Soviet era, Gothic was taught to linguists during their postgraduate studies (including at Tbilisi State University). Nowadays, Gothic is taught to the students of Tbilisi State University Bachelor's Degree Programmes and Master's Degree Programmes within the framework of theoretical courses: "Introduction to Germanic Philology" (Bachelor's Degree Programme) and "Comparative Grammar of Germanic Languages" (Master's Degree Programme). "Introduction to Germanic philology" is also taught at other Georgian universities, e.g. at Batumi Shota Rustaveli State University.

Such an attitude towards Gothic indicates that, in academic circles, Gothic philology by its status is practically equated with Classical philology.

From the philological and linguistic points of view, the Anglo-Saxon language is equally important. Especially taking into consideration the fact that it constitutes an early stage of development of the English language, which is so important and widely used today.

Beowulf, one of the most important works of English literature, a rare product of the synthesis of Pagan and Christian cultures, is written in Anglo-Saxon/Old English. This epic poem written in alliterative verse can be fully appreciated only in the original, which necessarily requires the knowledge of the Anglo-Saxon language. Literary works composed in Anglo-Saxon/Old English are considered to be an important part of English literature. Accordingly, quotations from *Beowulf* and other pieces of literature written in Old English are included in the articles of the *Oxford English Dictionary* as illustrative citations.

The knowledge of Old English and Middle English (which is a kind of transitional stage between the English of the Anglo-Saxon period and the Modern English language) helps one clearly understand why spelling and pronunciation in Modern English differ so substantially. Such knowledge also facilitates the attainment of proficiency in English spelling and orthography. Without at least some basic knowledge of Middle English and Early New English (the early stage of the Modern English language) it is practically impossible to read and adequately comprehend the works of important and influential English authors such as Chaucer or Shakespeare. This is why Anglo-Saxon or Old English is taught in many institutions of higher education having respective departments of English philology.

3. Background: Types of Available Printed and Electronic Resources for Old Languages

Qualitatively speaking, there are two types of resources for studying old languages, viz. **printed** and **electronic**.

In the Soviet Union and post-Soviet countries (where the concerned pro-

ject originates from) the users interested in old languages could avail themselves of several different **printed resources** (in Russian). The majority of these resources consisted of language manuals (including the description of the phonetics, morphology and syntax of the language in question), as well as the paradigms/grammatical tables of word declension and conjugation, texts to read for the language acquisition and necessary glossaries. Such printed resources included for instance a Gothic manual + reader + glossary by Mirra Gukhman (Gukhman 1958). There was a similar book by S. Chemodanov, which included texts in Old High German, Old Low Saxon and Middle High German (Chemodanov 1978). The book was appended with Gothic material: similar morphological tables for Gothic, Gothic glossary and some excerpts from *Codex Argenteus*.

The book by the Russian-Soviet scholar Alexandr Smirnitsky (Smirnitsky 1953) included Old English/Anglo-Saxon, Middle English and Early New English texts, as well as brief description of the historical development of the English grammar and the glossary of Old English, Middle English and Early New English words.

The book by the outstanding Russian specialist in Scandinavian studies, Mikhail Steblin-Kamensky "*Old Icelandic Language*" (Steblyn-Kamensky 1955) shared essentially the same structure (i.e. a review of grammar, morphological paradigms, Old Icelandic texts and the glossary).

The use of this type of educational material for the acquisition of old languages is too tiresome and time-and-energy consuming to be sufficiently effective for the education of our modern, highly "digitalized" youth, who are accustomed to the routine use of computerized devices like smartphones and tablets, to say nothing of computers themselves.

The root of the problem lies in the fact that the languages in question (Gothic and Anglo-Saxon) belong to the so-called synthetic type and are marked with considerable morphological diversity. It means that each word, whether it be a noun, adjective or verb, occurs in text in many different inflectional forms, depending on its case, number, gender or person. For example, *qamt* in Gothic means 'you came' (or, more precisely, 'thou camest'), *qēmum* means 'we came' and *qēmun* – 'they came', while the initial form, the infinitive of this verb which a reader/student should look up in a dictionary, is *qiman* – 'to come'. A similar pattern is observed with nouns and adjectives. Nouns and adjectives are inflected in Gothic in accordance with their number, gender and case, so that it is not always easy to determine the respective lemma (which must be looked up in a dictionary) for each of these word forms.

There is a very similar picture in Anglo-Saxon, which is marked with the nearly identical morphological complexity and diversity of inflectional forms, whether it be in the declension of nouns and adjectives, or in the conjugation of verbs.

As a result, reading a Gothic or Anglo-Saxon text in a printed chrestomathy, a learner of language has to constantly refer to the grammar tables (para-

digms) showing the regularities of the inflection of nouns, adjectives and verbs, in order to determine their lemmas. Only then can he/she look up the word in a dictionary and determine its meaning. It is easy to imagine how laborious and cumbersome this process may be using printed books. It may take a vast amount of time and energy to read, morphologically analyze and understand just one sentence using such an obsolete method, no matter how comprehensive dictionaries or reference books may be available to the person in question.

The second type of educational tools for studying old languages includes **electronic databases/chrestomathies (with dictionaries/glossaries)**. Some examples include the Gothic language database (available online) administered by Uppsala University Library (*The Gothic language*); Gothic Online — an online reader of Gothic texts compiled by the Linguistic Research Center of the University of Texas at Austin (*Gothic Online*);² Wulfila Project — a small digital library dedicated to the study of Gothic and Old Germanic languages, hosted by the University of Antwerp, Belgium (*Project Wulfila*) (<http://www.wulfila.be/>), contains almost all known Gothic texts, thoroughly analyzed on word-by-word basis both morphologically and semantically;³ the online version of the well-known Bosworth-Toller Anglo-Saxon Dictionary (*Anglo-Saxon Dictionary*); an online dictionary of Old High German hosted by the Saxon Academy of Sciences in Leipzig, Germany (*Althochdeutsches Wörterbuch*) and some others.

It should be noted that the online electronic resources providing exhaustive information on the meaning and morphological status of the texts they feature, leave very little to students' creativity, curiosity and diligence and make the reading process too easy. Experience shows that such 'simplification' is not really helpful for students in memorizing morpho-grammatical regularities and semantics of the old language(s) they wish to master.

4. **Web-Based Chrestomathy — Reasons for its Composition and *Modus Operandi***

The problems and difficulties described in the previous section led to the development of an electronic resource which, in the authors' opinion, would significantly facilitate the mastery of these two philologically important languages.

4.1 **The Initial Reader**

As a first step, we tentatively developed a web-based educational Reader (<http://gothic.margaliti.com/siteatta.htm>) which includes the Gothic version of the Lord's Prayer (*Atta unsar* which in Gothic means "Our Father", see Figure 1), the Lord's Prayer in Anglo-Saxon, as well as some other fragments of Gothic and Anglo-Saxon texts together with certain auxiliary and/or supporting materials.

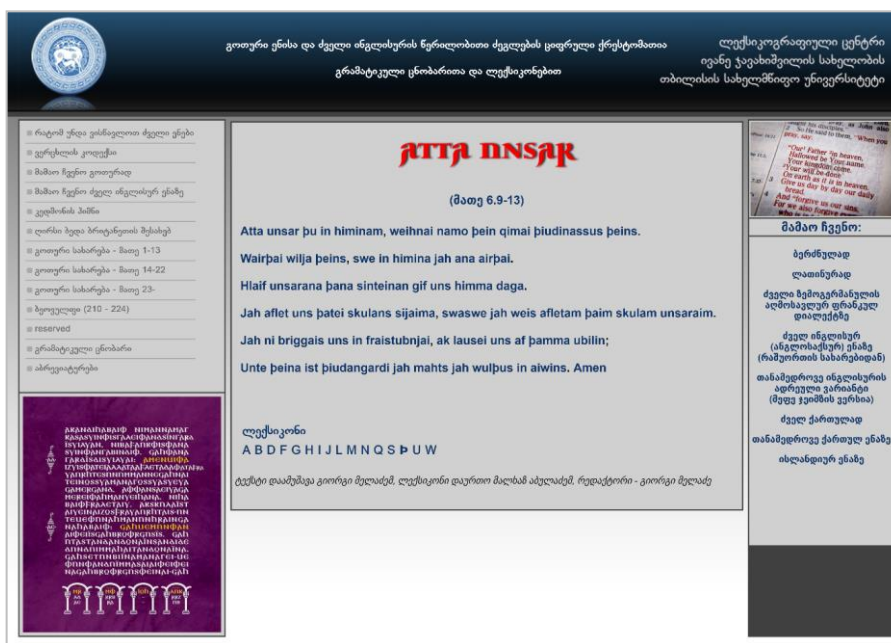


Figure 1

In the Gothic text of the Lord's Prayer every word is hyperlinked and pop-up windows supply complete morpho-semantic analysis of words. Proceeding from the considerations discussed in the previous section, this method was rejected and other texts are not hyperlinked. Instead of hyperlinks, glossaries, provided for each text, contain not only lemmas as headwords, but also all word forms of a given text which are cross-referenced to corresponding lemmas. For example: in the glossary of the Old English version of the Lord's Prayer, the word form *heovonum* is cross-referenced to the lemma *heovon* 'heaven'; the word form *urne* is directed to *ure*, 'our' and so on.

The development of the Reader was dictated by the need of teaching Gothic and Anglo-Saxon to the students of Tbilisi State University. The introduction of the Reader to students made the teaching process more enjoyable and considerably increased their interest in these languages, in etymologies of words, in affinity of languages and comparative linguistics in general. Encouraged by this initial success, we further elaborated our idea and the decision was made to develop an integrated, complex, more sophisticated electronic web-based resource for reading Gothic and Anglo-Saxon texts, based on an adequate piece of computer software. We also decided to supply this upgraded version of the resource with an English interface, in order to make it accessible to English-speakers/foreigners. Such an electronic resource, as we hope, will enable Georgians as well as English-speaking foreigners to read and analyze old texts in these linguistically very important languages by means of their

native Georgian or English. The importance of computer software is emphasized because without an adequate programmatic support, any collection of electronic texts will be exactly as (or even more) unmanageable and hardly navigable, as the printed versions of such texts.

4.2 The New Chrestomathy

In order to give the readers of the present article some general idea of the composition and organization of the Chrestomathy, how it works and what the exact method of its operation is, we would like to make a brief demonstration using computer screenshots.

The Chrestomathy (<http://germanic.ge/en/got>) includes 6 Gothic texts from the New Testament⁴ and 7 Anglo-Saxon texts,⁵ all of which are lemmatized, that is, all words of these texts have their lemmas identified and each of them is hyperlinked to the respective dictionary headwords.

This is illustrated below. If we choose a Gothic text from the Chrestomathy (Chapter 8 from Matthew's gospel for instance <http://germanic.ge/en/got/record/matthew-8/>) and move our mouse pointer over some word (say, *fairgunja* in the very first line of the text, see Figure 2), we shall see a hover box with the essential information about the word: lemma/dictionary headword *fairguni*; Part of speech (PoS) tag *noun* and the definition of the word 'mountain, mountainous area'.

The screenshot displays the website interface for the Chrestomathy of Gothic and Anglo-Saxon Written Records. At the top, there is a logo and the title 'CHRESTOMATHY OF GOTHIC AND ANGLO-SAXON WRITTEN RECORDS'. Below this is a search bar and a navigation menu with 'GOTHIC' and 'ANGLO-SAXON' tabs. A horizontal menu lists the letters of the Gothic alphabet: A B G D E Q Z H Þ I K L M N J U P R S T W F X H U O. Below the menu, there are sections for 'WRITTEN RECORDS' and 'REFERENCES'. Under 'WRITTEN RECORDS', a list of records is shown, with 'Matthew 8' selected. The 'REFERENCES' section includes 'Noun paradigms'. The main content area shows the text of Matthew 8, with a hover box over the word 'fairgunja' in the first verse. The hover box displays the lemma 'fairguni', the part of speech 'noun', and the definition 'mountain; mountainous area'.

Figure 2

Double-clicking on the word *fairgunja* will direct us to the full dictionary entry containing more detailed information concerning the word in question (see Figure 3).

This "shorter" default version of the article provides the following information: lemma/headword (*faírguni*); PoS tag (*noun*); Gender (*neuter*); Stem type (*-ja- stem*); and Definition (*'mountain, mountainous area'*).



Figure 3

The extended version of the same article with Etymology, Concordance and Paradigm boxes toggled on conveys respectively the following information: (i) etymology (*the word's supposed origin and its purported correspondences in cognate Germanic and Indo-European languages*); (ii) concordance (*inflectional forms of a word as they occur in various places of Gothic texts, with their exact occurrences precisely indicated*); and (iii) the generalized paradigm of -ja- stem nouns, shown on the example of the Gothic word *kuni* ('kin, family') (see Figure 4; for more detailed information about the dictionary component of the Chrestomathy, see section 4.3).

Similarly, in their extended versions, entries defining verbs contain (in addition to the PoS tag) information on the type of each verb in question (*strong, weak, irregular*), its class (*strong verb class 3, weak verb class 2, etc.*), and, due to the wide and frequent use of participial constructions in Gothic, present participle (in cases where the present participle form is attested for the particular Gothic verb).

With the additional boxes toggled on, each verb entry provides the information on the word's etymology and concordance/occurrences (as in case with other parts of speech) and also the paradigm of the conjugation of this type of verb.

The screenshot shows the website interface for the Chrestomathy of Gothic and Anglo-Saxon Written Records. The main heading is "CHRESTOMATHY OF GOTHIC AND ANGLO-SAXON WRITTEN RECORDS". There are tabs for "GOTHIC" and "ANGLO-SAXON". A search bar is present, and below it is a navigation bar with letters A through O and their corresponding Gothic characters. The main content area is titled "faírguni".

WRITTEN RECORDS

- Matthew 8
- Matthew 27
- Mark 15
- Luke 15
- John 18
- Corinthians I, 15

REFERENCES

- Noun paradigms

faírguni

Part of speech: **noun** Gender: **neuter** Stem: **-ja-**

mountain; mountainous area

Etymology

[< *Prot-Germ* *fergunjan < *fergu- (*etymology unclear*) + *-nja *suffix*; cf *OE* fyrzen, firzen "mountain, mountain-woodland"; Presumed to be: (1) borrowed from Proto-Celtic; (2) derived from *Indo-Europ* *perkw- "oak"; or (3) related to the name of the god of thunder and lightning in Baltic languages cf *O Prus* Perkūns]

Concordance

faírguni - *Nom, Acc, sing* - Mrk. III, 13; IX, 2; {abbr}Luk. IV, 5; VI, 12; IX, 28; Jhn. VI, 3; VI, 15; Galat. IV, 25
 faírgunjis - *Gen, sing* - Luk. IV, 29; XIX, 37
 faírgunja - *Dat, sing; Nom, pl* - Matth. VIII, 1; Mrk. V, 11; IX, 9; XI, 1; XI, 23; Luk. VIII, 32; IX, 37; XIX, 29 etc.
 faírgunje - *Gen, pl* - Luk. III, 5
 faírgunjam - *Dat, pl* - Mrk. V, 5

Paradigm

1.1.1. (c)

(-a- stem type subtype)

Nouns, -ja- stem, neuter gender		
kuni (kin, race)		
	Singular	Plural
Nominative	kuni	kunja
Genitive	kunjis	kunjē
Dative	kunja	kunjam
Accusative	kuni	kunja
Vocative	kuni	

Goto [Noun paradigms](#)

Figure 4

In addition, taking into consideration the morphological specificities and importance of present participles in Gothic, "extended" versions of verb entries are hyperlinked to the generalized declensional paradigm of present participles (shown on the example of **qībands** – 'saying, speaking'). This paradigm is supplied with a short explanatory text (which can also be toggled on/off) characterizing the use and functions of present participles in Gothic.

The web-based Chrestomathy also contains the paradigms of strong and weak declensions of Gothic (and Anglo-Saxon) adjectives, of personal, possessive, relative, etc. pronouns and so on.

The Anglo-Saxon version of the web-based reader operates quite similarly.

4.3 The Dictionary Component of the Chrestomathy

A dictionary entry of Gothic words consists of three parts: 1. Description of meaning; 2. Concordance; 3. Etymology.

4.3.1 Meaning

Description of meaning of Gothic words in a bilingual Gothic-Georgian Dictionary would be somewhat different from the approach adopted by us for the present Chrestomathy. In an ordinary bilingual dictionary, one of the challenges of lexicographers is the problem of equivalence, which is to be addressed properly (Adamska-Salaciak 2010; Margalitadze and Meladze 2016).

Comparison of the translation of one and the same Gothic word in the Georgian translation of Gospels, reveals that in many cases, a Gothic word has several contextual equivalents in Georgian. For example: meaning of the Gothic verb *gafulljan* is 'to fill', in Georgian *ავსება* 'avseba'. In Luke 1:15 'jah **ahmins weihs gafulljada** nauhþan in wambai aiþeins seinazos' ('and he shall be filled with the Holy Ghost, even from his mother's womb'), the Gothic verb *gafulljan* is translated by the Georgian verb *აღვსება* *aghvseba* (lit. 'to overflow'); in Mark 15:36 'þragjands þan ains jah **gafulljands swam akeitis ...**' ('And one ran and filled a sponge full of vinegar'), the Gothic verb *gafulljan* is translated by a Georgian verb *გაჟღენტა* 'gazhghenta' (lit. 'to soak, to saturate').

In a bilingual Gothic-Georgian Dictionary entry of the Gothic verb *gafulljan*, its meaning would be described by the Georgian equivalent *ავსება* 'avseba' (lit. 'to fill'), followed by several illustrative citations from Gospels, where the Gothic verb in question is translated by different Georgian contextual equivalents ('to overflow', 'to soak'; see examples above).

As mentioned above, a different approach was adopted for the dictionaries integrated in the present Chrestomathy. Dictionary entries do not contain illustrative quotations as the Chrestomathy contains texts from Gospels where Gothic words are attested. As to description of meaning, alongside general Georgian equivalents of Gothic words, contextual equivalents may also be provided in some instances, when these contextual equivalents are used in the texts of the Chrestomathy. In the case of the Gothic verb *gafulljan*, discussed above, its dictionary entry in the Chrestomathy contains two meanings: 1. *ავსება* 'to fill'; 2. *გაჟღენტა* 'to soak, to saturate'. The latter is a contextual equivalent of the Gothic verb *gafulljan* (see discussion above), but it is added to the entry as a separate meaning, because this contextual equivalent is attested in Mark 15, one of the texts of the Chrestomathy. Some dictionary entries of the Chrestomathy contain Gothic collocations, also attested in the texts of the Chrestomathy.

Thus an entry structure of a dictionary incorporated in the Online Chrestomathy differs from an entry structure of an independent bilingual Gothic-Georgian dictionary. Such lexicographic decisions were dictated by didactic

considerations and, from our point of view, will facilitate the understanding of Gospel texts.

4.3.2 Concordance

Articles in the Gothic dictionary are also supplied with *Concordance* information, showing exact textual occurrences of each Gothic word (chapters and paragraphs of Gospels, Epistles, etc.). Textual occurrences of Gothic words are sorted out according to grammatical forms. For example:

gafulljada — III person, singular, mediopassive (voice), indicative (mood) — Luke I, 15;
gafullidedun — III person, plural, past, indicative (mood) — Luke V, 7; John VI, 13;
Skeireins VII, 8
gafulljands — present participle — Mark XV, 36.

The subject of concordances of Gothic words and the linguistic importance we attach to the identification and analysis of contextual occurrences of Gothic lexical units will be discussed in more detail in the final section of this article.

4.3.3 Etymology

Each dictionary entry includes *Etymologies*: i.e. cognate words from related Old Germanic (Old English, Old High German, Old Saxon, Old Frisian, Old Norse, Old Icelandic) or other Indo-European languages (Ancient Greek, Latin, Sanskrit, Hittite, Slavic languages, etc.) etymologically connected with the particular Gothic word/dictionary headword. Thus, the resource gives the user a clear idea of the etymological background of a word, showing at the same time the pattern of development, which the meaning of its root has followed in particular related languages.

Experience shows that etymologies are one of the most interesting issues in the study of these languages for students. They find it very interesting to discover that an English word may have cognate forms in Gothic, Sanskrit, Latin, Old Church Slavonic and other languages. Consequently, this part of the dictionary entry is treated with due attention. *Etymologia proxima* (immediate etymology) is provided for all Gothic words, i.e. each entry contains cognate words from Old Germanic languages, as well as reconstructed Proto-Germanic words. For example: etymological part of the entry of the Gothic word *dags* 'day' includes the following information: [Proto-Germanic **dagaz*; Old English *dæg* (Modern English *day*); Old Frisian *dei, dē*; Old Saxon *dag*; Old High German *tag, tac* (Modern German *Tag*); Old Icelandic *dagr* (Modern Icelandic *dagur*)].

Despite the wide-spread opinion that providing *etymologia remota* (remote etymology) should be avoided in order to make less mistakes in etymologies (Buchi 2006; Durkin 2009), the decision was made still to include cognate words from other Indo-European languages. This decision was based on stu-

dents' interest in such information, on the other hand, the dictionaries in question are not etymological ones, which fact justifies occasional inclusion of even conjectural information. For example: the etymology of the Gothic word *himins* 'heaven' includes the following information: [Proto-Germanic **hemina-*, **hemna-*; Old Saxon *hevan*, *heban*; *himil*; Old Icelandic *himinn*; Old Frisian *himel*, *himul*; Old High German *himil*; ← Indo-European **akmen-*; Sanskrit *ásmān* "stone, rock; heaven, firmament"; Avestan *asman-* "stone; heaven, firmament"; Greek *ἄκμων* "anvil"; Lithuanian *akmuõ* "stone"; Old Slavic *камь* (Russian *камень* "stone")]. Students find such etymologies very interesting. They realize that the English word *heaven* may be cognate with a word denoting 'stone' in another related language, because ancient peoples believed that the firmament was of solid matter. Semantic differentiation between cognate words in different related languages is another issue which sparks students' interest, therefore such cases are provided not only with cognate forms, but also their meanings, as is seen from the example above with *himins*.

The dialog boxes containing the information concerning word etymology and concordance can be toggled on or off by means of on-screen buttons.

4.3.4 Grammatical Information in Entries

As already mentioned above, each dictionary entry provides all necessary information concerning the morphological class and inflectional patterns of the word in question. Furthermore, morphological paradigms of main parts of speech: noun, verb, adjective and pronoun are created electronically and incorporated in the Chrestomathy. Figure 4 shows the toggled-on generalized paradigm of -ja- stem nouns, presented on the example of the Gothic word *kuni* ('kin, family').

Thus the dictionary article is connected via another hyperlink with the grammatical paradigm showing the declensional pattern of this class (-ja- stem) of Gothic neuter nouns. The forms given in the paradigm will help the user quickly and clearly understand that the word he/she has looked up in the dictionary (*fairgunja*) is a form of dative case, singular of the entry headword *fairguni*.

The treatment of Anglo-Saxon texts and corresponding morphological information is essentially the same insofar as the Anglo-Saxon/Old English language shows considerable morphological similarity with Gothic. Thus, the ways of the presentation in the Chrestomathy of lexico-semantic and morphological data concerning Anglo-Saxon are practically identical with those employed with respect to Gothic. The only difference is that Anglo-Saxon dictionary articles are not supplied with concordance information, as far as the great abundance of Old English texts makes it unnecessary and practically impossible to indicate all manuscripts, texts or contexts where this or that Anglo-Saxon word occurs.

4.4 The Method

As one can see, instead of leafing to and fro through the volumes of printed readers, the Georgian-speaking or English-speaking student of Gothic and Anglo-Saxon having access to our Chrestomathy, can comfortably sit in front of a computer screen with all the important lexical and grammatical information, necessary for the proper understanding of a text written many centuries ago, being just a mouse click away.

On the other hand, this method was also developed because we do not approve of the approach involving word-for-word translation and the presentation of the full grammatical information concerning the word under analysis inside the pop-up or drop-down windows. Our deliberate purpose was to avoid unnecessary simplification of the process of analysis of the surviving old texts for language learners. We preferred to encourage the application of the dictionary integrated in the electronic reader with respective morphological paradigms connected with dictionary articles. This will, hopefully, on the one hand, encourage the development of independent and creative linguistic thinking in old language learners and, on the other hand, facilitate stable memorization of semantic and morphological peculiarities of these languages. We also hope that such approach to learning old languages will give students better insight in understanding the peculiarities of the functioning and development of language as such both on synchronic and diachronic levels, understanding of underlying processes which are at play in language.

5. Some Linguistic, Lexicographic and Other Difficulties Associated with the Chrestomathy

In this section there will be discussed some difficulties, encountered at various stages of the development of the online resource.

When the work on the Gothic-Georgian and Gothic-English dictionaries started, it was expected that old Georgian and English translations of Gospels would substantially facilitate the task of finding exact Georgian and English equivalents for Gothic words. Such expectations in this respect were based on the fact that, as a rule, older translations of Gospels very closely follow their Greek original, but in fact this was not always the case. This happened because despite the presence of considerable contentual proximity of respective Gothic, Georgian and English Gospel texts, there were some significant differences in the concrete means of grammatical conveyance of the same meaning. For instance, one difficulty encountered in the process of defining Gothic verbs was caused by the issue of the verb transitivity-intransitivity, namely, many Georgian and English **intransitive** verbs corresponded to their Gothic counterparts which were in fact **transitive**, but stood in (medio)passive voice. Relying upon Georgian/English rendition of such Gothic verbs and interpreting them in the dictionary as intransitive would be erroneous and inadmissible. Thanks to

background knowledge as concerns the etymology of Gothic and other Germanic words, possible mistakes in addressing this particular issue were avoided.

The disambiguation of homonyms also posed significant linguistic and technical problems. It should be noted that each and every specific homonym in the Chrestomathy is individually hyperlinked to its corresponding lemma in the dictionary, eliminating any chance of mistake or misinterpretation.

For instance, the Gothic word *im*, wherever it means 'them' (*personal pronoun, dative case, plural of all three genders*) is hyperlinked to the dictionary lemma *eis* 'they' (*personal pronoun, nominative case, plural, masculine*), but any homonymous Gothic word *im* encountered anywhere in the texts with the meaning '(I) am' is hyperlinked to the lemma/dictionary entry *wisan* 'to be'.

Likewise, the Anglo-Saxon word *for* when used in the sense of the preposition, is linked to the lemma/entry *for* (meaning essentially the same as the modern English preposition 'for'); however, when the homonymous word *for* represents the past tense (*singular, 1, 3 persons*) of the strong verb *faran* ('to fare', 'to go', 'to proceed'), in such cases a mouse-click on *for* will direct the user to the dictionary headword *faran* – verb 'to fare'. Also, the occurrences of the words *hund* whenever they may be encountered in the text, will direct us either to the entry where *hund* means 'hound', 'dog', or to the entry where *hund* means 'hundred', depending on the meaning of the word in its concrete context.

Another problem is a purely Anglo-Saxon issue, which also had to be addressed while implementing the project. The point is that many Old English words are attested in their different variant forms. This is due to the fact that Old English as a language and its literature developed over several centuries, so it is only natural that Anglo-Saxon texts reflect the diachronic evolution of Old English in its morphology and vocabulary. Besides, these texts are written in various regional varieties of Anglo-Saxon (traditionally referred to as dialects), such as West Saxon, Anglian (Northumbrian and Mercian), and Kentish. This fact also adds to the multiplicity of spelling variants of Old English words. For instance, Anglo-Saxon noun *æcer* (meaning "field" or "acre") may occur in different texts (or even in one and the same text) in its variant spelling forms like *æcyr* or *acer*. Pronoun *ælc* (Modern English "each") is often spelt also as *elc*, *ealc* or *ylk*. Noun *frip* (which means "peace; truce") is attested in a variety of spelling forms such as *frið*, *fryþ*, *fryð*, *frioþu*, *frioðu*, *freoþu* and so on. This very abundance of spelling variants caused certain difficulties which had to be addressed somehow. The said difficulties were associated both with the disambiguation of homonyms and with the lemmatization of tokens, since it was necessary to determine initial forms for various words occurring throughout seven Anglo-Saxon texts in many different spelling varieties. These initial forms coincide with the word forms/lemmas, chosen for the headwords of respective dictionary entries of the Chrestomathy.

6. Prospects for the Development and Sustainability of the Chrestomathy

With the ever increasing availability of the Internet, computers, digitalized devices, electronic databases and mobile technologies, the application of digitalized, web-based, computationally engaged research, scientific and teaching resources logically becomes advisable and even necessary.

Working in the field of Germanic languages and comparative linguistics, we could not oversee this necessity of the use of digital technologies in the pursuit of our scholarly and academic activities.

The advantages (as we see them) of the application of digital tools in the process of the acquisition of Gothic and Anglo-Saxon languages were demonstrated above. The above-described method allows a reader to quickly process and "digest" substantial amounts of morphological and lexico-semantic information using multiple blocks of grammatical data integrated into the single digital resource. This method allows avoiding the waste of time and energy on the mind-numbing "navigation" through the seas of pages of printed volumes, as it was inevitable with the use of older non-digital textual resources. This is especially important in case of old/dead languages, which by definition do not allow of the use of any form of immersion methods or interaction with native speakers. Reading daily more pages of the texts written in old languages guarantees the easier memorization and better mastery of the languages available exclusively in the form of written records.

In the future, Lexicographic Centre at TSU plans to work on the composition of similar web-based readers/chrestomathies for the texts in other linguistically important languages such as Old Icelandic, Old High German and Old Frisian. To this end, we are holding some preliminary talks with certain Icelandic and Frisian academic and research institutions, as well as with individual scholars, about the possible cooperation in this field.

We also do not rule out the possibility of composing, in the not so distant future, similar web-based educational resources also with respect to linguistically and scientifically important Indo-European languages like Lithuanian, Avestan and Sanskrit.

In this regard it should be mentioned that Tbilisi State University has a longstanding tradition of teaching ancient Indo-European languages: back in the 1920s, Avestan and Sanskrit were taught at TSU on the initiative of a prominent Georgian linguist Giorgi Akhvlediani, who had composed the Avestan and Sanskrit textbooks. It is our ambition to try to revive the tradition of studying and teaching these important languages at TSU.

There are also other plans with regard to the Gothic and Anglo-Saxon Chrestomathy discussed in the present article, as well as to our prospective web-based resources involving other languages. We expect that the English interface, making the web-based educational resources accessible to an international scholarly audience and general public, will facilitate and encourage their use by foreign students and can help Georgia become a regional centre for

the study and teaching of these languages. The substantial improvement of the level of teaching Gothic, Anglo-Saxon and other linguistically significant languages, as well as the establishment of an entire school of trained specialists in the field of old Indo-European languages and comparative linguistics will be another substantial benefit brought about by the successful implementation of the project(s).

Finally, another possible application of the digital readers/chrestomathies is that this same method can be equally successfully used for the acquisition of proficiency in contemporary, "living" languages. Our native Georgian for example, at least in some of its aspects, is known to be quite difficult to learn. Foreigners find the structure and morphology of Georgian verbs particularly difficult to adequately study and master (cf. Shanidze 1973; Holisky 1979; Gippert 2016). We believe that digitalized, integrated readers comprising texts in Georgian or other notoriously difficult languages like Hungarian or Finnish, for instance, could greatly facilitate the acquisition of such languages. Georgian texts with all words hyperlinked to their respective dictionary entries, providing in their turn additional information on the peculiarities of morphology of the words in question could greatly help foreigners consolidate their theoretical knowledge of the Georgian grammar and transform it into practical awareness of the laws and regularities of the contemporary Georgian language. In particular, the use of digitalized Georgian readers would make the task of mastering the conjugation of extremely versatile and multiform Georgian verbs much easier and much more feasible for foreign learners of Georgian.

7. Some Lexicological Aspects of the Project

The Chrestomathy discussed above, combines two main aspects of scholarly activities of the authors of the present article: Germanic philology and Lexicography. One more sphere of our interest is the issue of semantic equivalence and anisomorphism between the synonyms from different languages. Many prominent specialists in the field of lexicography and linguistics have expressed their views on the subject of equivalence of synonyms in various languages (cf. Adamska-Salaciak 2010; Gouws and Prinsloo 2008; Hartmann 2007; Zgusta 1971, 2006). Without the adequate realization of the fact that synonyms are not always fully equivalent, it is practically impossible to find the best translation and give proper definition to a word from the source language within a dictionary word-entry, as it has often been discussed in the papers of renowned lexicologists (cf. Geeraerts 2010; Hausmann 1985; Katz and Postal 1964).

Working on the Gothic component of the Chrestomathy, editing and perfecting Gothic-Georgian dictionary and analyzing the text of Gothic Gospels, we could not fail to note how linguistically interesting it would be to find out how much equivalent the synonymous words used in the Gothic, Greek, Latin, Old Georgian and English versions of the New Testament really were. This

vision of the Gothic texts motivated us to supplement the entries of the Gothic-Georgian and Gothic-English dictionaries with the special section, the aforementioned *Concordance* feature, containing information about occurrences of this or that Gothic word throughout the surviving Gothic texts. This information is meant to help us in the future in "tracking" Gothic words in specific contextual surroundings in order to identify their matches in Old Georgian, Greek or Latin translations of the New Testament and help us determine the degree of their equivalence.

Using the information on the contextual occurrences of the Gothic lexical units, it is planned to dedicate a special research to the analysis of the issues of semantic equivalence which the Gothic, Georgian or other translators of the New Testament must have had to address. Our experience with various (Gothic first and foremost) translations of the Scripture has convinced us that such an analysis will yield interesting results from the general linguistic, lexicological, lexicographical, cultural and other points of view.

One interesting example could be the comparison of a well-known passage from Matthew 6:28, where Gothic *gakunnaiþ blomans haiþjos* (literally "perceive the flowers of the heath/field") corresponds to the Latin "*considerate lilia agri*" in Clementine Vulgate, Greek *καταμάθετε τὰ κρίνα τοῦ ἀγροῦ* ("learn thoroughly/examine/consider the lilies of the field") and the Old Georgian *განიცადენით ზროშანნი ველისანი* (literally "perceive/experience the lilies of the field/valley"). It is especially interesting to note the parallelism between Gothic *gakunnan* and Old Georgian *განცდა* (both verbs meaning approximately "to perceive", "to get to know" or "to experience"), remembering that the majority of modern translations of this excerpt have something like "*see/observe/look at the lilies of the field*". No less interesting is the semantic anisomorphism between the Gothic *blomans* ("flowers") on the one hand and Latin, Greek and Old Georgian *lilia*, *κρίνα*, *ზროშანნი* ("lilies") on the other hand.

Similar and even more interesting results can be obtained after the deep analysis of the usage of Gothic words in the surviving chapters of Gospels and other Gothic texts, which analysis will be conducted based on the *Concordance* feature of the web-based Chrestomathy.

8. Conclusion

This is in brief what we wanted to share with colleagues from lexicographical, lexicological and linguistic community about the digitalized Chrestomathy. We thought that the information about this educational and research resource was worth making public for various reasons: (1) the Chrestomathy will increase the interest of students in studying Gothic and Anglo-Saxon; its English interface makes this online resource globally available for those who are interested in these two linguistically important old languages and, generally, in language diachrony and comparative linguistics; (2) as far as old/"dead" languages are concerned, the Reader will enable a user easily to consolidate his/her knowl-

edge of the languages which are no longer in use; (3) the Chrestomathy allows to scientifically analyze the vast blocks of data pertaining to old languages (Gothic, Anglo-Saxon), since all primary and auxiliary texts/data included therein are represented in electronically searchable format; (4) furthermore, the Resource has the potential to facilitate the composition of similar resources with respect to extant/"living" languages. Chrestomathies based upon the method applied in the online resource described above can be successfully used for mastering morphologically complex (mostly synthetic) languages characterized with the abundance of inflectional and suppletive forms, such as Hungarian, Turkish, Russian, German, Georgian and many others. Web-based chrestomathies similar to the one described in the present article, allowing the instant identification of the lemmas of each allomorph occurring throughout the texts, could be very useful and reliable digital tools for the mastery of the languages of the said type; (5) especially with respect to Anglo-Saxon (which is the earlier form of modern English language) the Chrestomathy facilitates the tracking and analysis of the processes of linguistic change and allows one to arrive at certain generalized conclusions concerning language diachrony; and finally (6) the *Concordance* feature, provided for the Gothic section of the Chrestomathy allows researchers to study and analyze the problem of semantic equivalence, which is universally relevant and important already on the synchronic level of language research and analysis. All these features, as we think, make the online resource interesting and noteworthy both from theoretical and practical points of view.

Acknowledgements

The authors would like to thank their colleagues, participants of the project presented above, who have contributed to its successful implementation. These are (in alphabetical order) Malkhaz Abuladze (the compiler of Gothic-Georgian and Anglo-Saxon-Georgian dictionaries included in the Chrestomathy), Maia Davlianidze (technical assistant) and George Kerechashvili (programmer, web developer, software specialist).

Endnotes

1. As an explanatory note, it should be mentioned here that the Chrestomathy is due to be completed by the end of 2019 when all of its programmatic features will be fully functional.
2. This online resource includes about ten Gothic texts/fragments from Codex Argenteus and Skeireins; texts are morphologically and semantically analyzed on sentence-by-sentence basis; some basics of Gothic grammar are also supplied.
3. It is worth mentioning however, that homonyms in many cases are not disambiguated. This is, most likely, due to the reliance on the text analyzer using relatively general algorithms.
4. **Matthew 8; Matthew 27; Mark 15; Luke 15; John 18; Corinthians I, 15**

5. **Ælfric's Genesis XXVII, 1-29; Bede on Britain** (*Geographical description of Britain by Bede the Venerable*); **Cædmon — Story and Hymn** (*Bede's narrative about Cædmon and his famous hymn*); **Béowulf** (*several excerpts from the poem*); **The Coming Of The English** (*Bede's description of the coming of the English to Britain*); **Ohthere's Account** (*Account of a 9th century Norwegian seafarer Ohthere about his voyages which he gave to King Alfred the Great of Wessex*); **An Excerpt From Old English Chronicles** (*from Parker Manuscript*)

References

- Adamska-Salaciak, A.** 2010. Examining Equivalence. *International Journal of Lexicography* 23(4): 387-409. *Althochdeutsches Wörterbuch*. Sächsische Akademie der Wissenschaften zu Leipzig. Accessed at: <http://awb.saw-leipzig.de> [28/01/2019].
- Anglo-Saxon Dictionary (digital edition).** Bosworth-Toller *Anglo-Saxon Dictionary*. Accessed at: <http://bosworth.ff.cuni.cz/> [26/01/2019].
- Buchi, É.** 2006. Etymological Dictionaries. Durkin, Ph. (Ed.). 2006. *The Oxford Handbook of Lexicography*: 338-349. Oxford: Oxford University Press.
- Chemodanov, N.S.** 1978. *Chrestomathy in the History of the German Language*. Moscow: Vysshaya Shkola. (In Russian)
- Durkin, Ph.** 2009. *The Oxford Guide to Etymology*. Oxford: Oxford University Press.
- Geeraerts, D.** 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Gippert, J.** 2016. Complex Morphology and its Impact on Lexicology: The Kartvelian Case. Margalitadze, T. and G. Meladze (Eds.). 2016. *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity, 6–10 September, 2016, Tbilisi, Georgia*: 16-36. Tbilisi: Ivane Javakhsishvili Tbilisi State University. <http://euralex.org/category/publications/euralex-2016/>.
- Gothic Online.** The Linguistics Research Center, The University of Texas at Austin College of Liberal Arts. Accessed at: (<https://lrc.la.utexas.edu/eieol/gotol/10>) [31/01/2019].
- Gouws, R.H. and D.J. Prinsloo.** 2008. What to Say about "mañana", "totems" and "dragons" in a Bilingual Dictionary? The Case of Surrogate Equivalence. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 869-877. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Gukhman, M.M.** 1958. *Gothic Language*. Moscow: Izdatelstvo literatury na inostrannykh yazykakh. (In Russian)
- Hartmann, R.R.K.** 2007. *Interlingual Lexicography: Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: Max Niemeyer.
- Hausmann, F.J.** 1985. Lexikographie. Schwarze, Christoph and Dieter Wunderlich (Eds.). 1985. *Handbuch der Lexikologie*: 367-411. Königstein/Ts.: Athenäum.
- Holisky, D.A.** 1979. On Lexical Aspect and Verb Classes in Georgian. Clyne, Paul R., Hanks, William F. and Carol L. Hofbauer (Eds.). 1979. *The Elements: A Parasession on Linguistic Units and Levels Including Papers from the Conference on Non Slavic Languages of the USSR*: 390-401. Chicago: Chicago Linguistic Society.
- Katz, J.J. and P.M. Postal.** 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.

- Krause, W.** 1951. *Die Sprache der urnordischen Runeninschriften*. Heidelberg: Carl Winter Universitätsverlag.
- Margalitadze, T. and G. Meladze.** 2016. Importance of the Issue of Partial Equivalence for Bilingual Lexicography and Language Teaching. Margalitadze, T. and G. Meladze (Eds.). 2016. *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity, 6–10 September, 2016, Tbilisi, Georgia: 787-797*. Tbilisi: Ivane Javakhishvili Tbilisi State University. <http://euralex.org/category/publications/euralex-2016/>.
- Project Wulfila* (a small digital library dedicated to the study of Gothic and Old Germanic languages, hosted by the University of Antwerp, Belgium). Accessed at: <http://www.wulfila.be/> [30/01/2019].
- Shanidze, A.** 1973. *Foundations of Georgian Grammar*. Tbilisi: Tbilisi University Press.
- Smirnitsky, A.I.** 1953. *Chrestomathy in the History of the English Language from VII to XVII Centuries*. Moscow: Izdatelstvo literatury na inostrannykh yazykakh. (In Russian)
- Steblyn-Kamensky, M.I.** 1955. *Old Icelandic Language*. Moscow: Izdatelstvo literatury na inostrannykh yazykakh. (In Russian)
- The Gothic Language*. Uppsala University Library. Accessed at: (<http://www.ub.uu.se/about-the-library/exhibitions/codex-argenteus/the-gothic-language/>) [31/01/2019].
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague: Mouton.
- Zgusta, L.** 2006. *Lexicography Then and Now: Selected Essays*. Edited by Fredric S.F. Dolezal and Thomas B.I. Creamer. Tübingen: Max Niemeyer.

Theoretical and Practical Reflections on Specialized Lexicography in African Languages*

Dion Nkomo, *School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda (Grahamstown), South Africa (d.nkomo@ru.ac.za)*

Abstract: In this article, reflections are made on some specialized lexicographical/terminographical resources being produced in African languages. The resources are produced in order to contribute towards the intellectualization of those languages for expanded functional usage. The article focuses on lemma selection, provision of data/information for included lemmata and structural aspects of the surveyed resources. With regard to the first area of focus, the article identifies the lack of a systematic approach to lemma selection, which undermines the potential of the resources as communicative and cognitive tools in specialized subject fields and disciplines. Secondly, regarding the provision of data categories, instances of insufficient information and cases of inclusion of irrelevant information are identified, both of which have implications for the functional value of the resources within specialized domains. Finally, reflections on aspects of dictionary structure indicate sub-standard structural designs which affect the user-friendliness of the resources, but some innovative structural designs are also identified. Overall, the article argues for a stronger lexicographic orientation in terms of the theoretical underpinnings guiding the production of specialized lexicographical/terminographical resources in African languages.

Keywords: SPECIALIZED LEXICOGRAPHY, TERMINOGRAPHY, AFRICAN LANGUAGES, INTELLECTUALIZATION OF AFRICAN LANGUAGES, LEMMA SELECTION, AFRICAN LEXICOGRAPHY, TERMINOLOGY, TERMINOGRAPHY

Opsomming: Teoretiese en praktiese gedagtes oor die gespesialiseerde leksikografie in Afrikatale. In hierdie artikel word gedagtes oor sommige gespesialiseerde leksikografiese/terminografiese hulpbronne wat in Afrikatale saamgestel word, weergegee. Dié hulpbronne word geskep om 'n bydrae te lewer tot die intellektualisering van hierdie tale vir uitgebreide funksionele gebruik. Daar word gefokus op lemmaseleksie, op die verskaffing van data/inligting vir die lemmas wat ingesluit is en op strukturele aspekte van die hulpbronne wat ondersoek is. In die artikel word daar met betrekking tot die eerste fokusarea 'n gebrek aan 'n sistematiese benadering tot lemmaseleksie geïdentifiseer wat die moontlikhede van die hulpbronne as kommunikatiewe en

* This article is a revised version of a paper that was presented at the 24th Annual International Conference of the African Association for Lexicography (AFRILEX), hosted by the Department of Language and Literature Studies, University of Namibia, Windhoek, Namibia, 26–29 June 2019. The author acknowledges the constructive feedback obtained from conference delegates as well as the two anonymous adjudicators.

kognitiewe hulpmiddels in gespesialiseerde onderwerpsvelde en dissiplines beperk. Tweedens word, met verwysing na die aanduiding van datakategorieë, gevalle van onvoldoende inligting en gevalle van insluiting van irrelevante inligting geïdentifiseer wat albei implikasies vir die funksionele waarde van die hulpbronne binne gespesialiseerde domeine inhou. Laastens dui gedagtes oor aspekte van woordeboekstruktuur op substandaard strukturele ontwerpe wat die gebruikersvriendelikheid van die hulpbronne beïnvloed, maar sommige innoverende strukturele ontwerpe word ook geïdentifiseer. In die geheel beskou, word in hierdie artikel gepleit vir 'n strenger leksikografiese oriëntering in terme van die teoretiese basis wat die skep van gespesialiseerde leksikografiese/terminografiese hulpbronne in Afrikatale rig.

Slutelwoorde: GESPELIALISEERDE LEKSIKOGRAFIE, TERMINOGRAFIE, AFRIKATALE, INTELLEKTUALISERING VAN AFRIKATALE, LEMMASELEKSIE, AFRIKA-LEKSIKOGRAFIE, TERMINOLOGIE, TERMINOGRAFIE

1. Introduction

Language policies and language planning efforts seeking to develop and promote the continent's indigenous African languages that were marginalized during the colonial era and the apartheid era in South Africa have culminated in the proliferation of lexicographical resources focusing on specialized academic and professional disciplines. Examples include specialized dictionaries in the fields of biology, linguistics, literature, medicine and music produced under the auspices of the African Languages Research Institute (ALRI) in Zimbabwe. In South Africa, the Department of Arts and Culture (DAC), the National Lexicography Units (NLUs), commercial publishers, institutions of higher education and other organisations have made significant contributions. Appearing in different sizes, formats and mediums, whether all those resources are fit to be called dictionaries is a debate that emerges from their compilers and users alike. For example, in the first volume of *Understanding Concepts in Mathematics and Science — A Multilingual Learning and Teaching Resource Book in English, IsiXhosa, IsiZulu and Afrikaans*, henceforth *Understanding Concepts in Mathematics and Science*, Young et al. (2005: viii) make a strong disclaimer that:

This is a resource book, not a textbook. *It is also not a dictionary, nor a teaching method book. It should thus be always used together with approved classroom teaching and learning materials (present author emphasis).*

There are numerous resources whose compilers appear to be in a dilemma of both attraction and repulsion by the term *dictionary*, in a manner that is reminiscent of what Landau (1984: 6) said about the power of the term *dictionary*. On the one hand, such power tends to elevate the status of some reference works that are called dictionaries, thereby making them successful and popular among users. On the other hand, it causes other resources to be judged harshly when they are called dictionaries but then fail to satisfy certain criteria commonly associated with dictionaries. Consequently, some compilers play it safe

by calling their products by several other names instead of dictionaries. What is remarkably common among the studied resources is that, through their titles and/or subtitles, front matter or blurb texts, they identify specific cognitive and communicative lexicographic functions that they seek to achieve, and their overarching endeavour to contribute to the intellectualization of African languages. Broadly speaking, they can comfortably be accounted for in Bergenholtz and Nielsen's (2006) inclusive disciplinary conceptualization of specialized lexicography, even though some of their compilers would prefer to call them otherwise, e.g. terminology products. It is for this reason that this article adopts a seemingly indecisive use of *specialized lexicographical/terminographical resources*. Tarp (2000: 214) is quite critical of scholars who consider specialized lexicography as something different from lexicography in general (which is the case with some terminographers). Łukasik (2016) even prefers to use the term 'pedagogical terminographers' to refer to compilers of specialized pedagogical dictionaries.

This article offers a theoretical engagement with an assortment of products whose production constitutes an integral aspect of the intellectualization of African languages for purposes of expanding their functional spaces (Kaschula and Nkomo 2019). The article illustrates how different compilers seek to contribute to this endeavour by identifying and analyzing the specific functions of some of the products. The development, documentation and description of terminology in these resources are regarded as central to the intellectualization of African languages, and the significance and potential of such works cannot be overemphasized. Hence the need to subject them to critical reflections.

While reflecting on the surveyed resources, this article argues for a stronger lexicographic orientation in terms of the theoretical underpinnings guiding their production. Weak theoretical foundations compromise the utility value of some of the products. This is particularly the case regarding lemma selection, as illustrated in Section 4. The article observes that in their lemma selection, some compilers lack systematic approaches that are appropriate for specialized lexicography. Beyond lemma selection, the analysis is extended to the provision of data categories in some of the studied lexicographic resources, identifying instances of insufficient information and cases of inclusion of irrelevant information. This is done in Section 5. Finally, reflections are made on some aspects of dictionary structure in Section 6, focusing on sub-standard structural designs and innovative structural designs. However, before getting into those core sections, the article provides some context in Section 2 by giving a brief overview of lexicography in African languages, following by a quick survey of specialized lexicography in Section 3. Concluding remarks are made in Section 7.

2. An overview of lexicography in African languages

Situating specialized lexicography in African languages within a broader historical context of African lexicography enables an appreciation of the vital role

that specialized dictionaries could play in solving communication and cognitive problems facing African societies and the challenges confronting this branch of lexicography (Gouws 2013; Nkomo 2010). Gouws (2013: 52) observes that:

Too many metalexigraphers were only concerned with general language dictionaries ... This lack of concern with LSP dictionaries led in far too many cases to LSP dictionaries not really qualifying as dictionaries but merely playing an inferior role as word lists or other restricted (and often handicapped) reference products.

As general language dictionaries have claimed most of the attention of historical dictionary research, the needs of diverse user groups regarding specialized subject knowledge and languages used in different disciplines and professions have been neglected. Efforts of addressing such needs have quite often lacked theoretical insights, resulting in what Tarp (2012) aptly describes as a slow-motion development of this branch of lexicography, especially in terms of the quality of dictionaries produced.

At least four types of dictionaries, albeit not mutually exclusive, may be identified in the lexicography of African languages from a historical perspective. Firstly, lexicography in African languages began with bilingual dictionaries pairing African languages with more powerful European languages like English produced by early missionaries to support their learning of African languages for evangelization purposes (Gouws 2007; Nkomo 2018). Dictionaries such as Kropf's (1899) *Kafir-English Dictionary* in isiXhosa and Hannan's (1959) *Standard Shona Dictionary* contributed to the codification and standardization of those African languages, thereby laying foundations for further intellectualization and use of the languages in standardized forms (Gouws 2007).

Secondly, monolingual dictionaries would follow to consolidate the position of African languages as either official or national languages. This is typically seen in post-apartheid South Africa where National Lexicography Units were established with their main function being the production of comprehensive general purpose dictionaries in the country's eleven official languages. These types of dictionaries are generally expected to further contribute to the standardization of the languages, allowing them to function authoritatively in their own right, without reference to the more powerful languages such as English. Simango (2009) refers to this as 'weaning Africa from the Europe', which suggests an ideological endeavor of linguistic decolonization.

Thirdly, especially in South Africa, the language-in-education policy which seeks to cultivate multilingualism in education incentivized the production of bilingual and multilingual school dictionaries mainly by commercial publishers such as Oxford University Press — Southern Africa, Maskew-Miller Longman and Pharos. For example, Oxford University Press — Southern Africa has over recent years produced a series of bilingual school dictionaries pairing English with languages such as isiZulu, isiXhosa and Sesotho sa Leboa.

The NLUs have also produced school dictionaries including a recent series of foundation phase picture dictionaries (see Taljard and Prinsloo 2019).

Lastly, although not a distinct category, as some of them fall under school dictionaries, specialized dictionaries have also gained significant attention linked to the endeavour of using indigenous official languages in specialized academic and professional disciplines. Just like general-purpose monolingual dictionaries in African languages, specialized dictionaries also seek to enhance and affirm the potency of languages in the post-colonial dispensation of the continent. The remainder of this article reflects on this category of dictionaries in African languages.

3. Specialized lexicographical and terminographical resources in African languages

In line with Gouws (2007) and Nkomo (2018) who underscore the interface between lexicography and language policy, this section offers a brief survey of specialized lexicographical and terminographical resources within the language policy environments of South Africa and Zimbabwe. South Africa arguably stands out in terms of an explicit lexicographical/terminographical and language policy interface. The post-apartheid legislative framework, central to which is the constitutional proclamation of the nine indigenous languages as official languages, as well as the imperative that they must be treated with parity alongside Afrikaans and English, had a catalytic effect on lexicography and terminology in African languages. At the level of basic education, the *Language-in-Education Policy* (LiEP), adopted in 1997, acknowledges "the cognitive benefits ... of teaching through one's medium (home language)". The cognitive challenges associated with the use of a language that learners are not proficient in are intricately linked to communicative problems, hence the LiEP's commitment to "the development of official languages" in order "to counter disadvantages resulting from different kinds of mismatches between home languages and languages of learning and teaching". It is in the context of such a legislative and policy framework that the majority of specialized lexicographical products have been developed in the African languages of South Africa. Most of them have been targeted at learners. This is consistent with Łukasik's (2016: 211) assertion that "one of the most important function of specialised dictionaries is the pedagogical (didactic) function". This is also captured in Fuertes-Olivera's (2010) edited volume entitled *Specialised Pedagogical Lexicography for Learners*. As one of the key role-players in the field of terminography in South Africa, the DAC has devoted significant attention to the "Schools Project" which is dedicated to the "documentation of existing terminology, and facilitation of the development of terminology in the African languages for new concepts that appear in the teaching materials for Grades 1 to 6" (DAC 2013a: v). This has resulted in the publication of the following series of multilingual resources:

- *Multilingual Financial Terminology List*
- *Multilingual Human, Social, Economic and Management Sciences Terminology List*
- *Multilingual Natural Sciences and Technology Term List (SeSotho)*
- *Multilingual Natural Sciences and Technology Term List (Tshivenda–Xitshonga)*
- *Multilingual Natural Sciences and Technology Term List (Nguni)*
- *Multilingual Mathematics Dictionary: Grade R–6*
- *Multilingual HIV/Aids Terminology*
- *Multilingual Soccer Terminology List*

The needs facing the education sector have also inspired other projects such as the Concept Literacy Project, leading to the publication of two concept literacy resource books for Mathematics and Science by Young et al. (2005; 2009). Organizations such as the Project for the Study of Alternative Education in South Africa (PRAESA) and the Human Sciences Research Council (HRSC), among others have also made notable contributions in terms of specialized lexicographical/terminographical work. Even the National Lexicography Units (NLUs) find themselves compelled to move beyond their primary function of producing comprehensive general-purpose monolingual dictionaries to produce dictionaries such as *Isichazi-magama seMathematika neNzululwazi*, an isiXhosa Mathematics and Science dictionary for primary schools produced by the IsiXhosa National Lexicography Unit. What is perhaps more remarkable is that some of these resources have been published by commercial publishers who were previously too cautious if not skeptical to publish material in African languages due to limited market demand. Maskew Miller Longman's *Longman Multilingual Maths Dictionary for South African Schools: English, isiXhosa, Afrikaans* and Cambridge University Press's *Isichazi-magama seziBalo Sezikolo saseCambridge* are examples that illustrate an attitude change by commercial publishers towards African languages dictionaries.

Universities have also responded to pedagogical challenges connected to the language barrier in South African higher education by developing, documenting and describing terminology in African languages (Nkomo and Madiba 2011). The *Language Policy for Higher Education* and the *Ministerial Report on the Use of Indigenous Languages in Higher Education* expanded the potential functional space for African languages and the need for their development in higher education. The latter explicitly mandated each institution to identify specific African language(s) that they would promote and develop for academic purposes. Terminology work focusing on isiXhosa at Stellenbosch University and isiZulu at the University of KwaZulu-Natal illustrate commitment to such a mandate.

Łukasik (2016: 212) aptly states that "the role of ... [specialized lexicographical/terminographical works] as carriers of specialized knowledge renders them indispensable tools in maintaining the flow of professional information". As such, terminographical works have also been produced for other types of users, who include professionals in various subjects and translators.

The *Use of Official Languages Act* compels government departments and state entities to choose and use at least three official languages with parity, which means that their services should be available in all the chosen languages. This requires the departments to translate their key policies, plans and reports into African languages. It is against this backdrop that the DAC's *Multilingual Financial Terminology* (DAC 2013b), *Multilingual HIV Terminology* (DAC 2013c) as well as the *Multilingual Parliamentary/Political Terminology* (DAC 2013d) lists become relevant. The legislative framework makes it imperative that terminographical tools and resources are developed to address communicative problems associated with the underdevelopment of the languages in South Africa.

The Zimbabwean legislative environment that inspired lexicographic activities from the 1990s has been slightly different from that of South Africa in that Zimbabwe did not have an overt language policy document until 2013. However, similar ideological aspirations regarding lexicography and the intellectualization of African languages have been obvious. For example, at the University of Zimbabwe, ALRI was established with a clear mission: To research and document the Zimbabwean languages in order to expand their use in all spheres of life. The expansion of the use of Zimbabwean languages *in all spheres of life* was envisaged to be attainable through mainly specialized lexicography. In the master plan of the African Languages Lexical (ALLEX) Project, which would be institutionalized in ALRI, several specialized dictionaries were listed as target outputs of the project. By the time the ALLEX Project ended, the following specialized dictionaries had been published:

- *Duramazwi reDudziramutauro neUvharanomwe* (a Shona dictionary of linguistic and literary terms)
- *Durazwi remiMhanzi* (a Shona dictionary of music terms)
- *Isichazamazwi SesoMculo* (a Ndebele dictionary of music terms)
- *Duramazwi reUrapi neUtano* (a Shona dictionary of biomedical terms)

These dictionaries were meant to support the infusion of Shona and Ndebele, the country's major indigenous languages, into the specialized domains that were mainly reserved for English. For example, English is the dominant language of teaching and learning in Zimbabwe across all educational levels. Advanced linguistic and literary studies in Shona and Ndebele continue to be conducted mainly through the medium of English since the early studies by colonial scholars. Hence the need for the specialized *Duramazwi* series in Shona and similar dictionaries in Ndebele, although progress could not be made in the latter beyond *Isichazamazwi SesoMculo*. Targeting student doctors and nurses, *Duramazwi reUrapi neUtano* notes on its blurb and introductory texts that because of their English-dominant academic and professional training, health-care providers cannot easily communicate with Zimbabweans who are not competent in English. It is such communication gaps and challenges between medical experts and their non-expert clientele that specialized lexicography in African languages can potentially address.

4. Lemma selection issues

Lemma selection decisions determine the typological distinction of specialized lexicographical resources from general-purpose dictionaries. While the latter may include lexical items with specialized disciplinary designations (Gouws 2013), they often label or mark them to indicate that they are not the default members of the word-list structure. This is clearly illustrated in Figure 1, a screenshot of an article stretch **leucine** – **leucoplast** from the *South African Concise Oxford Dictionary* (SACOD) below:

leucine /'lu:si:n/ ■ n. Biochemistry a hydrophobic amino acid which is an essential nutrient in the diet of vertebrates.
– ORIGIN C19: coined in French from Greek *leukos* 'white'.
leuco- (also **leuko-**) ■ comb. form 1 white: *leucoma*.
2 representing **LEUCOCYTE**.
– ORIGIN from Greek *leukos* 'white'.
leucocyte /'lu:kə(ʊ)saɪt/ (also **leukocyte**) ■ n. Physiology a colourless cell which circulates in the blood and body fluids and is involved in counteracting foreign substances and disease; a white (blood) cell.
– DERIVATIVES **leucocytic** adj
leucocytosis /,lu:kə(ʊ)saɪ'təʊsɪs/ (also **leukocytosis**) ■ n. Medicine an increase in the number of white cells in the blood, especially during an infection.
– DERIVATIVES **leucocytotic** adj
leucoderma /,lu:kə(ʊ)'dɜ:mə/ (also **leukoderma**) ■ n. another term for **VITILIGO**.
leucopenia /,lu:kə(ʊ)'pi:nɪə/ (also **leukopenia**) ■ n. Medicine a reduction in the number of white cells in the blood.
– DERIVATIVES **leucopenic** adj
– ORIGIN C19: from Greek *leukos* 'white' + *penia* 'poverty'.
leucoplast /'lu:kə(ʊ)plɑ:st, -plɑ:st/ ■ n. Botany a colourless organelle found in plant cells, used for the storage of starch or oil.
leucosis /lu:'kəʊsɪs/ (also **leukosis**) ■ n. a leukaemic disease of animals, especially one of a group of malignant viral diseases of poultry or cattle.
– DERIVATIVES **leucotic** adj
leucospermum /l(j)ukəʊ'spɜ:məm/ ■ n. a southern African shrub of a genus of the protea family that comprises the pincushions. [Genus *Leucospermum*.]
leukaemia /lu:'ki:mɪə/ (US **leukemia**) ■ n. a malignant progressive disease in which the bone marrow and other blood-forming organs produce increased numbers of

Figure 1: An article stretch from SACOD

Users of dictionaries such as SACOD are not always guaranteed to get assistance regarding specialized terminology or concepts from general-purpose dictionaries. On the other hand, specialized dictionaries are primarily consulted for information regarding language and concepts with specialized academic or professional designations. A law student or a legal practitioner will be more disappointed if they fail to find a specific legal term in a law dictionary than in a general-purpose dictionary.

As they were conceived to assist specific users regarding specific academic or professional disciplines and subject fields, lexicographical/terminological

resources surveyed in the previous section are expected to describe the relevant terminology and concepts included as lemmata. The South African Mathematics and Science dictionaries compiled for learners and teachers clearly delineate their broad subjects (Mathematics and Science) as well as the specific sub-fields within those disciplines. *Isichazi-magama seMathematika neNzululwazi* indicates that its mathematics section explains concepts falling under algebra and geometry while the science section covers fauna, flora, matter, energy, change, earth and planets. The *Longman Multilingual Maths Dictionary for South African Schools: English, isiXhosa, Afrikaans* commits itself to responding to the national Mathematics curriculum.

According to Nielsen (1995), a systematic presentation of the subject fields covered by dictionaries is crucial for lemma selection. A detailed conceptual mapping of the subject-field and its sub-fields is necessary to ensure that relevant lexical items within the delineated scope are systematically included. Supposing that the compilers of the studied Mathematics and Science dictionaries judiciously included lexical items within the clearly delineated scopes, the dictionaries would be expected to complement other official curriculum materials like textbooks and workbooks to a greater effect. However, subject-field delineation and its systematic presentation may not be sufficient. Subsequent procedures of lexicographical description, including how the lemmata are defined, are equally important for the resources to be useful products of specialized lexicography.

Some resources developed to assist users with specialized linguistic and conceptual data are less helpful because they omit key terminology. While Łukasik (2016: 214) identifies the major shortcoming of specialized dictionaries to be a consequence of having "been constructed for 'everybody'", it is worth noting that far too often, they are also 'constructed by everybody'. This includes subject experts or lexicographers who over-rely on the former for advice on lemma candidates. While the input of subject experts remains indispensable in specialized lexicography, it should be solicited and utilized within clearly conceived lexicographical frameworks established by qualified lexicographers rather than over-relying on experts for lexicographical matters. Experts may be oblivious of the needs of non-expert users and take for granted some terms and concepts assuming that users know them, or even impose complex definitions and defining language to the detriment of users who may either fail to find some terms or struggle to comprehend their meanings. A survey conducted by Mawonga et al. (2014) to test a Political Philosophy Terminology resource whose list was compiled by a lecturer established that students could not find most of the terms that challenged them in their Political Science module offered at their university. Students experienced difficulties with more terms and concepts beyond what the lecturer considered to be difficult terms.

Compilers of some specialized lexicographical resources in African languages have adopted a commendable procedure of using specialized texts to circumvent the problem highlighted above. Specialized dictionaries for schools typically draw lemmata from the language used in curriculum documents,

textbooks and assessment material. The *Multilingual Parliamentary/Political Terminology List* published by the DAC also illustrates this practice, having drawn its lemmata from Hansard reports (parliamentary debates); parliamentary proceedings (speeches, motions, notices of motions); parliamentary papers (order papers, minutes, announcements, tablings and committee reports) and legislation (Acts, Bills, government notices, proclamations and Gazettes) (DAC: 2013d: iii). With such diverse specialized texts dealing with the relevant specialized domains, it is possible to capture the key concepts and terms for inclusion. Using specialized texts as a form special corpora is therefore recommended (Bowker and Pearson 2002). Corpus query tools such as WordSmith Tools or Sketch Engine can enable lexicographers to identify lemma candidates and their keyness values with more efficiency and precision. However, special corpora may not be used exclusively given the limitations of special corpora when it comes to matters of size, representativeness and balance (Bowker and Pearson 2002). For example, despite the commendable use of Hansard reports (parliamentary debates); parliamentary proceedings (speeches, motions, notices of motions); parliamentary papers (order papers, minutes, announcements, tablings and committee reports) and legislation (Acts, Bills, government notices, proclamations and Gazettes) as a form of a special corpus to identify relevant parliamentary and political terminology, the DAC (2013d) resource does not include terms related to different types of democracy. While the user may get meaning information, including translation equivalents in ten languages for the English terms *democracy*, *conciliatory democracy* and *consociational democracy*, he/she may get stuck regarding terms such as *authoritarian democracy*, *direct democracy*, *presidential democracy*, *parliamentary democracy*, *participatory democracy* and *representative democracy*. While such terms might not have appeared in any of the collected texts, an exploration of the conceptual structure of political science as a subject would have helped the compilers, working closely with the subject expert, to foresee that translators and language practitioners for whom the resource is developed may need those terms in the future. It is also possible that the terms could have been omitted due to manual or traditional term extraction from texts, which is inefficient.

Writing on theoretical challenges to practical specialized lexicography almost twenty years ago, Tarp (2000) ably demonstrated how a discussion of lemma inclusion or exclusion might become irrelevant. However, such discussions remain necessary as they stimulate theoretical reflections on lexicographic practice. The Shona music terms dictionary, *Duramazwi RemiMhanzi*, has included, described and illustrated diagrammatically as lemmata the following words: *gamburabota* (thumb), *mungedzapenzi* (index finger), *mungedzazvose* (index finger), *munongedzo* (index finger), *mudapakati* (middle finger), *nhembayemwana* (ring finger), *kasiyanwa* (little finger). The inclusion and lexicographical description of these finger names, three of which are synonyms for the index finger, needs to be interrogated, especially in view of the target users. A close analysis of the dictionary reveals that these lexical items also appear frequently in the explanations of lemmata referring to music instruments, especially how

instruments are played or handled. For example, reference is made to middle fingers (*minongedzo*) in the explanation of the lemma *madhebhe*, a type of *mbira*. Therefore, what emerges to have been adopted is the principle that definitions should not use words that are not explained in the same dictionary. However, the relevance of such a principle is questionable in this dictionary, which is targeting mother-tongue speakers of Shona studying and teaching music at schools, colleges and universities. Given that the premise of compiling this dictionary is making music knowledge accessible to the target users in their mother tongue, the dictionary does not offer the relevant help as the definitions of such lexical items do not make explicit references to how the fingers referred to are useful in playing or handling specific instruments.

At the beginning of this section it was asserted that lemma selection decisions will, to a large extent determine whether the final product would be a specialized dictionary or general-purpose dictionary. As in the case of the music terms dictionary discussed above, one may indeed wonder about the inclusion of the twenty-five lemmata identified from the Shona biomedical terms dictionary listed in Table 1 below:

Shona lemma	English gloss
bhabharazi	hangover
bhandeji	bandage
bhonzoz	bone
bipito	shoulder
bofu	blind person
book	hand
bupununu	halfwit/imbecile
burunzeve	deaf person
bvi	knee
bwaira	blink
chanza	palm
chichi	religion/denomination/cathedral
chidhumbu	corpse
chidodoma	heel
chidya	thigh
chidya	food
chifuva	chest
chigumwe	toe
chikumwe	small finger
chimedzo	throat
chirevu	chin
chirevhu	doctor
chironda	wound
chirwere	illness/ailment/disease
chitemo	headache

Table 1: Some lemmata from *Duramazwi reUrapi neUtano*

The twenty-five lemmata were identified from the first fifteen pages of the dictionary. Twelve of them refer to basic body parts like shoulder, hand, chin, palm, chest, heel, thing, etc. Some of the lemmata have multiple synonyms which are also entered and as separate lemmata. Seven lemmata have been entered with individual comprehensive treatment in the dictionary which however, includes general rather than specialized definitions. The target users of this dictionary are mother-tongue Shona-speaking medical student doctors and trainee nurses. Shona is taught as a compulsory language subject for Shona-speaking learners from the first grade to the General Certificate of Education: Ordinary Level after which students can train as nurses or future medical students may specialize in science subjects. The point is that the target users are mother-tongue speakers of the language who would have studied it for eleven years. One would wonder if they would not know such words for them to search for their general definitions, which may be looked up in a general dictionary, if needed, and whether there were no more unfamiliar lexical items from the biology and medical fields that could be accommodated.

While similar questions may be raised regarding some lemmata in the DAC's *Multilingual Mathematics Dictionary: Grade R-6*, the compilers seem to have thought deeply about such lexical items. The compilers write:

Mathematics is generally referred to as Numeracy Skills in Grades 1 to 3. In these grades a number of general terms such as *match, choose, fill in, light, heavy* etc. are included. To teach the learners about space and position many prepositions such as *like, behind, on, under*, etc. are included. Learners have to learn about measurements, capacity, height, weight, length, shapes, and patterns. In the context of Mathematics terms such as *long, tall, wide, full, half-full* and even *cup* (measurement: 250 ml) have a mathematical meaning, and are thus included in the list, although it might be argued that they are general words in other contexts.

In order to read the time on a clock the learners need to know that *hand* may be used to indicate the hand of a clock (long hand, short hand) and they learn that even a clock has a *face*. Learners also have to learn how to use a calculator. It is sometimes difficult to decide on the status (general words or subject specific terms) of lexical items in school texts, and that is why many terms used in teaching Mathematics at primary level are regarded to be ordinary words, but are nevertheless included in the glossary (DAC 2013a: v).

Such a theoretically-motivated lemma selection procedure needs to be considered in the compilation of specialized lexicographical/terminographical resources in African languages in order to improve their reliability as sources of specialized terminological and conceptual information. Furthermore, an appropriate lexicographical treatment of lemmata needs to accompany such a procedure in order to provide users with assistance that they would barely get from general dictionaries. This is discussed in the next section.

5. Lexicographical treatment of lemmata: Insufficient versus irrelevant information

While a well-conceived and diligently constructed lemma list will provide a

solid foundation for a specialized dictionary, it will not automatically translate into an effective and user-friendly product. After locating the lexical item that prompts dictionary consultation, the user must find the relevant data type(s) from which the needed information may be retrieved. Thus the lexicographical treatment of included lemmata is another equally important theoretical decision with practical import. The retrievable information must be relevant, sufficient and accurate in order to address the question(s) that prompted the dictionary consultation procedure. Questions of relevance, adequacy and accuracy are therefore crucial in the analysis of the surveyed dictionaries.

5.1 Insufficient information

The problem of insufficient treatment of included lemmata is one notable weakness of African language lexicographical/terminographical resources that prevails more when compilers distance themselves, their practice and products from lexicography as a discipline. In some of the resources, the compilers avoid the term *dictionary* in their titles, opting instead for *glossary*, *terminology* or *terminology list*. As noted earlier, the term remarkably emerges in blurbs, forewords and other introductory texts, as illustrated by the resources produced by the Unit for IsiXhosa at Stellenbosch University's Language Centre, e.g. *Isigama Somthetho / Law Terminology / Regsterterminologie*. Although the DAC products are clearly identified as dictionaries by their compilers, they offer the typically bare minimal treatment of the included terms as illustrated by the following screenshot (DAC 2013e) and can barely be described as more than terminology lists.



decoder	
Afrikaans	dekodeerder
IsiZulu	isiguquli
IsiZulu	idikhoda
IsiXhosa	isiguquli maz'omoya
IsiXhosa	idikhowuda
Siswati	sihlungi
Siswati	idekhoda
IsiNdebele	ivulafihlo
Setswana	seranodi
Sepedi	sefetoši sa maphoto
Sesotho	dikhoda
Sesotho	sefetoledi
Tshivenda	dikhoda
Xitsonga	mpaluxana

Figure 2: A screenshot of the lemma decoder from the *Multilingual Terminology for Information Communication Technology*

The screenshot illustrates the treatment of the term **decoder** in the *Multilingual Terminology for Information Communication Technology*, which is saved as a *Multilingual ICT Dictionary*. From this resource, the data that the user may access constitutes the translation equivalents in the other ten South African official languages. For some of the languages, two equivalents, a coinage and a transliterated borrowing, are provided. On the one hand, while the coinages may capture some properties of a decoder, they may not be easily understood by a mother-tongue speaker in terms of their specialized ICT designation, which lies somewhere between the three senses captured by the online *Collins English Dictionary* (CED) as follows:

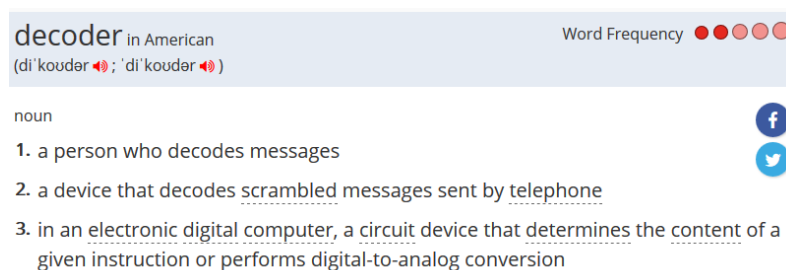


Figure 3: An article for the lemma *decoder* in the online CED

The problem with some of the coinages is that despite appearing to be transparent and self-explanatory, they may not be unambiguous enough for one to easily determine which of the three special designations captured in the online CED they refer to. For example, the coined isiZulu term *isiguquli* may literally translate into 'that which changes or transforms something'. This could be a transformer, an electric device, which is not necessarily the same as a decoder, or it could be a person who transforms something. With the term not appearing in other isiZulu dictionaries, lack of additional information implies that the mother-tongue isiZulu speaker who is challenged because of limited or a lack of English competence does not necessarily benefit from this resource in his/her quest for unambiguous communication and comprehension of the term. Newly coined terms in African languages will not be known to the majority members of the speech communities despite morphological appropriateness and what may appear to be semantic transparency. Therefore, while the preface of the *Multilingual Terminology for Information Communication Technology* states that "[T]o promote effective communication in these domains it is essential that terminology should be available for all the languages in these fields of knowledge" (DAC 2013e: v), it is clear that equivalent terminology alone is insufficient and that newly coined terms are not sufficiently explicit to independently facilitate unambiguous communication within the specialized disciplinary and subject fields.

In order to communicate effectively, terminology users in African languages

will need additional information which includes terminological definitions in both English and the target language to provide explanations of meaning. For those who are competent in English, English explanations will serve a disambiguating purpose as they search for target language terminology for mother-tongue text production purposes. Target language explanations may assist those with limited English competence to understand the meaning of the terms first in their own language and then in English if necessary. Failure to provide such information makes DAC products inferior to many others such as the bilingual glossary series of the University of KwaZulu-Natal, see for example Figure 4 illustrating an article from *A Glossary of Law Terms: English–isiZulu*.

ANIMO VICINO NOCENDI. An enquiry into the conduct of a neighbour so as to establish whether it constitutes a nuisance or not. It involves an investigation into the mental state of the neighbour at the time of the commission of the said act. The investigation would seek to establish whether there was an intention to create a nuisance, such intention being termed *animo vicino nocendi*. Where the act in question was solely to create a nuisance to the neighbour, such an act would be deemed unreasonable.

Uphenyo ngokuhlukumezana komakhelwane: I-Animo vicino nocendi, uphenyo lokuthola inhloso yamakhelwane, ngesenzo esibanga isicefe komunye umakhelwane. Uphenyo lubheka ukuthi wayecabangani umakhelwane ngesikhathi enza lesisenzo esibaluliwe. Uma isenzo samakhelwane sicubungulwa kutholakale ukuthi okuyona nhloso yangempela yalesosenzo ukubangela omunye umakhelwane isicefe, kuthiwa isenzo senzwe nge-animo vicino nocendi.

Figure 4: An article for the lemma from *A Glossary of Law Terms: English–isiZulu*

The detailed explanation of *Animo vicino nocendi* in both English and isiZulu has great potential to assist novice law students who speak isiZulu to understand the concept to which the term refers, especially as the explanations use simplified language. However, etymological information for the Latin dominated law terminology would be useful for to provide further assistance for cognitive purposes.

Furthermore, pictorial and diagrammatical illustrations prove to be useful in conveying meaning information, especially in the scientific and practical disciplines. The *Illustrated Glossary of Southern African Architectural Terms: English–isiZulu* and the *Longman Multilingual Maths Dictionary for South African Schools:*

English, isiXhosa, Afrikaans have been well-conceived in this respect, as they try to provide as many pictures and diagrams as possible. However, others such as the *Illustrated Science and Technology Dictionary / Isichazi-magama sezeNzululwazi neTeknoloji* appear to be economic and arbitrary in illustrating concepts. Not only does arbitrariness result in some resources providing insufficient information, it may also lead to the inclusion of irrelevant information as shown in the next subsection.

5.2 Irrelevant information

In order for lexicographers to provide relevant information types in their works, it is instructive to recall Tarp's (2000: 198) assertion that "the only way to reach a scientific conclusion of what should be included in a dictionary is to base this conclusion on an analysis of the user, the user characteristics, the user situations, the user needs and the corresponding lexicographic functions". Although specialized lexicography in African languages has thus far been discussed within the general framework of language intellectualization, it is important to engage in a more nuanced analysis of the products according to the function theory of lexicography (Tarp 2008) before identifying data that offers irrelevant information from some of the resources.

The cognitive lexicographic function (Tarp 2008) of specialized lexicographical/terminographical products recognizes their "pedagogical (didactic) function" (Łukasik 2016: 211; see also Fuertes-Olivera 2010; Tarp 2005) "as carriers of specialised knowledge" (Łukasik 2016: 212). This function, or at least the intention to serve it, prevails in most specialized lexicographical/terminographical products in African languages. For example, the cognitive function finds expression when Young et al. (2005: 9) articulate that the aim of the resource *Understanding Concepts in Mathematics and Science 2* is "to provide teachers and learners with accredited specialists' expert knowledge, understandings and descriptions of ... key concepts" in "Mathematics, Physics, Chemistry, Geography and Life Sciences". The cognitive function similarly motivated the compilation of the *Illustrated Glossary of Southern African Architectural Terms: English–isiZulu*, as Frescura and Myeza (2016: xv) reiterate its intention to support students with "ready access to a specialized lexicon", part of which carries "an inherent symbolism associated with the social values and cultural heritage of traditional rural society". *A Glossary of Law Terms: English–isiZulu* was developed to "improve cognitive development, critical thinking and epistemic access to complex pedagogies" (Khumalo 2018: x). As such, specialized lexicography/terminography in African languages is not simply about languages *per se* but more importantly about knowledge (cf Antia and Ianna 2016: 78), which underscores the cognitive function of the lexicographical products.

While cognitive lexicographic functions are distinguished from communicative functions (Tarp 2008), it is noteworthy that they are not mutually exclusive. Comprising of mainly text reception and text production, communicative functions are integral components of knowledge acquisition, knowledge pro-

duction, knowledge reproduction and knowledge dissemination. The functions are equally critical for specialized lexicography/terminography in African languages. The use of English specialist terms in various professional fields and disciplines impedes comprehension of texts, both in oral and written forms. In order to address the problem of text comprehension that is impeded by English, mainly meaning information is provided in the mother tongue either in the form of translation equivalents or explanations of meaning in the form of terminological definitions. Once English terms have been understood via the mother tongue, they may then be used better for text production in English, the dominant official language. However, with specialized lexicography/terminography being conceived as part of the ultimate intellectualization of African languages for use in high status domains, the produced resources are also meant to assist users with terms and information that support text production in the target languages. As an example, *A Glossary of Law Terms: English–isiZulu* was conceived to launch "the process of cultivating and advancing isiZulu to be an appropriate tool for legal education and education practice" (Zondi 2018: xiv), i.e. facilitating legal communication within the academy and the profession.

Apart from insufficient information in some of the studied resources, as shown in the previous subsection, the opposite — inclusion of irrelevant data, may be noted in some specialized lexicographical/terminographical products in African languages. The problem may not be as prevalent as that of insufficient information but it is significant in terms of theoretical insights into specialized lexicography. One illustration of this issue is part-of-speech (PoS) information in *Isichazi-magama seMathematika neNzululwazi* and the *Illustrated Science and Technology Dictionary / Isichazi-magama sezeNzululwazi neTeknoloji*. In the former, all lemmata, which are English lexical items accompanied by isiXhosa translation equivalents, bilingual PoS labels such as n/b (noun/*isibizo*), v/nz (verb/*isenzi*) and adj/bl (adjective/*isibaluli*) are provided, with nouns being the majority. In the latter, most but not all lemmata are accompanied by monolingual PoS labels, namely *n.* (noun), *v.* (verb), *pref.* (prefix) and *adj.* (adjective), with nouns dominating again. Apart from the unexplained and inconsistent provision of PoS information in the latter dictionary, the decision to provide this information is not motivated anywhere in any of the two dictionaries which seem to focus on the cognitive function. With a more holistic curriculum approach which does not isolate language teaching from content subjects, it could be argued, even without the lexicographers clearly identifying this possibility, that the provision of such information could develop learners' knowledge of grammar. Furthermore, it could be argued that the provision of this information could develop users' awareness of the value of this type of lexicographic information. However, the lexicographers' utter silence about it suggests that they are equally oblivious of its inclusion. Compare this with the provision of PoS in the *Multilingual Parliamentary Dictionary* where this type of information is only purposefully included for disambiguation in cases where an included term belongs to more than one word class, e.g. *audit* as a noun and also as a verb. The motivation appears to be clear in the latter case.

Another type of lexicographic information of which the inclusion may be questioned is tone marking in Shona music terms dictionary *Duramazwi reMimhanzi*. This dictionary is largely cognitive in its functional dimension as it seeks to facilitate Shona mother-tongue speakers' access to specialist and ethnomusicological knowledge through Shona terminology and encyclopedic explanations. With the tonal aspects of language pertaining to oral text production (pronunciation) and, to some extent, oral text reception (auditory perception), this information could possibly be relevant if the dictionary was targeted at non-mother-tongue speakers of Shona. Tonal information is also supplied in *Duramazwi ReDudziramutauro neUvaranomwe*, the Shona dictionary of linguistic and literary terms, where it could possibly be justified in that tone would be an important topic of which each lemma could serve as a practical example for students. Without a clear motivation, it could be postulated that this type of information was included in both dictionaries taking after the general-purpose dictionaries, which had just been published as trendsetters in Shona monolingual lexicography.

The foregoing analysis presents the provision of PoS information and tonal marking in the discussed specialized dictionaries as grey areas. The value added to those dictionaries by the inclusion of such information remains doubtful. While insufficient information would clearly reduce the functional value of dictionaries, users may not always feel the negative impact of irrelevant information. However, from an academic point of view, inclusion of irrelevant information illustrates insufficient theoretical guidance in the planning and execution of practical lexicographic tasks. There may also be practical implications in terms of wasted dictionary space and time that is dedicated to the inclusion of irrelevant information. This may unnecessarily delay the publication of the dictionary. Without a convincing motivation of including such information for mother-tongue Ndebele-speaking music students and practitioners, the editors of the Ndebele music terms dictionary, *Isichazamazwi SezoMculo*, opted against its inclusion in light of the financial cost of hiring a linguist specializing in tone. The sentiment remains that some lexicographers included the data following trends in general-purpose lexicography rather than catering for the needs of their users in specific usage situations.

6. Design and structural issues

Improvement in lemma selection decisions and lexicographical treatment of selected lemmata in specialized lexicographical/terminographical products in African languages will need to be complemented by effective design and structural frameworks. Taking design and structural issues for granted may undermine a product that provides the necessary data at both macro- and microstructural levels. Accessibility of the necessary data is as important as its availability. Against this principle, the lexicographical/terminographical resources surveyed in this study may be divided into three categories, i.e. those that adopt fairly standard structural designs without much creativity in their

data presentation, those that are sub-standard in their approach because of little if any inspiration from theoretical lexicography and those that adopt innovative approaches. The next two subsections will exemplify the latter two, i.e. sub-standardly structured resources and innovatively structured resources in 6.1 and 6.2 respectively, noting that there is not much to learn from the first category. It suffices to note that fairly standard dictionary designs are generally traditional in terms of their structural presentation of included lexicographical data even though they may not be as bad as those described as sub-standard in 6.1 below. However, the innovative practices noted in 6.2 need to be considered in the improvement of not only the sub-standardly structured resources but also those that display standard macro- and microstructures as there is always a room for improvement.

6.1 Sub-standard structural designs and presentation

This category pertains to those resources that are mainly called glossaries or terminology lists. Compilers tend to list translation equivalents of selected English terms in target languages in a manner that indicates limited attention to structural and design issues. Consider Figure 5 below (DAC 2013d):

appropriation		arbitration	
Afrikaans	bewilliging	Afrikaans	arbitrasie
IsiZulu	isabelo	IsiZulu	ukulamula
IsiXhosa	uhlahlo-mali	IsiZulu	ukwahlulela
IsiXhosa	uhlahlo lwabiwo-mali	IsiZulu	ukuthatha isinqumo
Siswati	sabelo	IsiXhosa	ulamlo
Siswati	liphakelo	Siswati	sehlulelo
IsiNdebele	isabelo seemali	IsiNdebele	ukulamulisa
IsiNdebele	ukwabiwa kweemali	Setswana	tsereganyo
Setswana	kabo	Sepedi	bolamodi
Sepedi	tekanyetšo	Sepedi	bonamodi
Sesotho	kabo	Sesotho	bonamodi
Tshivenda	mukovho	Sesotho	bolamodi
Xitsonga	nkavelo	Tshivenda	vhulamukanyi
		Xitsonga	vulamuri
appropriation act		arbitrator	
Afrikaans	begrotingswet	Afrikaans	arbiter
IsiZulu	umthetho wesabelo	IsiZulu	umlamuli
IsiXhosa	umthetho wolwabiwo- mali	IsiZulu	umahluleli
Siswati	umtsetfo wemaphakelo	IsiZulu	umthathi sinqumo
Siswati	umtsetfo wesabelo	IsiZulu	umthathi sinqumo
IsiNdebele	umthetho wokwabiwa kweemali	IsiXhosa	umlamli
IsiNdebele	umthetho wesabelo seemali	Siswati	umehluleli

Figure 5: An extract from the *Multilingual Parliamentary Dictionary*

The extract from the *Multilingual Parliamentary Dictionary* is typical of DAC products which simply provide lists of translation equivalents in the other ten South African official languages for English terms. As they appear in bold print, English lemmata are admittedly easy to identify. So are the translation equivalents, which are also in bold and accessible via a vertical list of language names in regular print. However, in cases where more than one translation

equivalent is provided in a particular language, e.g. three equivalents in isiZulu for **arbitration**, the language name is listed three times to indicate each of the equivalents. This results in some articles being longer than others, although each article provides translation equivalents for ten other official languages for English terms, which could be avoided by listing the equivalents horizontally for each English term. More importantly, the user will never be certain about the sense relations of multiple translation equivalents. At a cursory level, the translation equivalents may be regarded as synonyms, given that the English term is given without any explanation, which suggests multiplicity of senses. However, it will only take a competent isiZulu speaker to know that the isiZulu equivalents *ukulamula* (stopping a conflict), *ukwahlulela* (judging) and *ukuthatha isinqumo* (making a decision) are not absolute synonyms. Translators working under pressure may miss the nuanced differences and pick any of the translation equivalents, if not the first one, given that the criteria and order for listing multiple equivalents is not explained anywhere in the DAC dictionaries. Thus, as simple as the structure may appear to the eye, making sense of it may be more complex in a real user situation. Such a challenge may not be addressed by only thinking about the implications of the presentation but also by making use of outer texts to describe the structure of the resources, including providing user guidelines.

The seemingly simple but sub-standard structural designs and presentation of lexicographical data is not exclusive to DAC terminological dictionaries but characterize terminological products compiled for use in institutions of higher learning. Consider the Figure 6 below.

English	isiXhosa	Afrikaans
Ab initio	Ab initio	Ab initio
From the beginning / the start.	Ukusuka ekuqaleni / isiqalo.	Van die begin af / die begin.

Figure 6: A screenshot from *Isigama Somthetho / Law Terminology / Regsterminologie*

The screenshot is an article from *Isigama Somthetho / Law Terminology / Regsterminologie*, which aims to "support Xhosa-speaking students who ... struggle with difficult Law terms" and "widen the scope of ... [their] understanding, so that these students are afforded the opportunity to learn and understand these Law terms through their mother tongue" (Sibula 2007: iii). Having adopted a similar structure, the editors of the *Illustrated Multilingual Science and Technology Dictionary / Isichazi-magama sezeNzululwazi neTeknoloji Ngeelwimi Ezininzi* aver that "[t]he column format makes it simple to move from one language to another" (Mbude-Shale, Wababa and Welman 2008: vi). However, the brevity of the English explanations and their translations, as well as the lack of discipline-specific contextual examples of usage make it inconceivable how first-year law students struggling with legal terminology would benefit from such an article. The structural presentation shows that these simplistic structures accompany the provision of insufficient data as a major limitation of some spe-

cialized lexicographical/terminographical products in African languages. A tentative approach in the conceptualization of the majority of the products results in compilers neglecting design and structural issues, focusing on translating English terms.

6.2 Innovative design and structural presentation

Notwithstanding the neglect of structural and design considerations in the compilation of lexicographical/terminographical resources highlighted in the previous sub-section, there are other gratifying and inspiring products in African languages which demonstrate lexicographers' meticulous consideration for the accessibility and user-friendliness of their products. Consider the examples in Figure 7a and 7b below. The examples illustrate key design features of two Mathematics dictionaries, namely *IsiChazi-magama sezoBalo* (Figure 7a) and the *Longman Multilingual Maths Dictionary for South African Schools: English, isiXhosa, Afrikaans* (Figure 7b). Apart from being specialized dictionaries, these dictionaries target junior school learners, i.e. children. Their use of colorful illustrations is a typical feature of children's dictionaries that increases the accessibility of the specialized mathematical content in the dictionaries. Not only are the colorful illustrations able to demonstrate complex mathematical operations for the target users, they also serve to attract the users and instill in them love for a generally intimidating and challenging subject, thereby serving a pedagogic function of specialized lexicography.

<p>Evaluate - to find the value of. <i>Hlola, phicotha - ukukhangela ixabiso lento ufumana umyinge.</i></p> <p style="text-align: center;">$14 \times 2 = 28$</p> <p>Even - equally balanced, equal in number or amount. <i>Lingana - ziyalingana, ziyalingana ngokwamanani okanye ixabiso.</i></p>  <p>Exchange - when we go shopping, we exchange money for goods. <i>Utshintshiselwano - xa styokuthenga senza utshintshiselwano ngemali neempahla.</i></p>  <p>Expand - write in full, if given a shortened expression write in its fullest form such as opening up brackets and not using exponential forms. The word can be taken as an antonym of simplify. <i>Ukunabisa - ukubhala ngokuzeleyo.</i></p> <p style="text-align: center;">Expand 3</p> 	<p>Expanded notation - a way of writing numerals or algebraic expression. <i>Ukucazulula inani ngokubhala - indlela yokubhala itimboli zamanani.</i></p> <p style="text-align: center;">$249 = 200 + 40 + 9$ $= (2 \times 100) + (4 \times 10) + (9 \times 1)$</p> <p>Exponent - the number of times that a number multiplies itself by 10. <i>Inani lobungakanani bokuphindaphindwa-inani elixela liphindaphindwa kangaphi inani elo. Eli nani libhalwa libe phezulu kuneli liphindaphindwayo, ngale ndlela kubhalwa ngayo idigri.</i></p> <p style="text-align: center;">$2 \times 2 = 2^2$ $10 \times 10 = 10^2$</p> <p>Exterior -The outside of something. <i>Umphandle - umphandle wento.</i></p>  <p>The exterior of the hut has two colours and a thatched roof. <i>Umphandle woronta unemibala emibini kunye nomnqwazi elupahlenti.</i></p>
--	--

Figure 7a: Design features of *IsiChazi-magama sezoBalo*

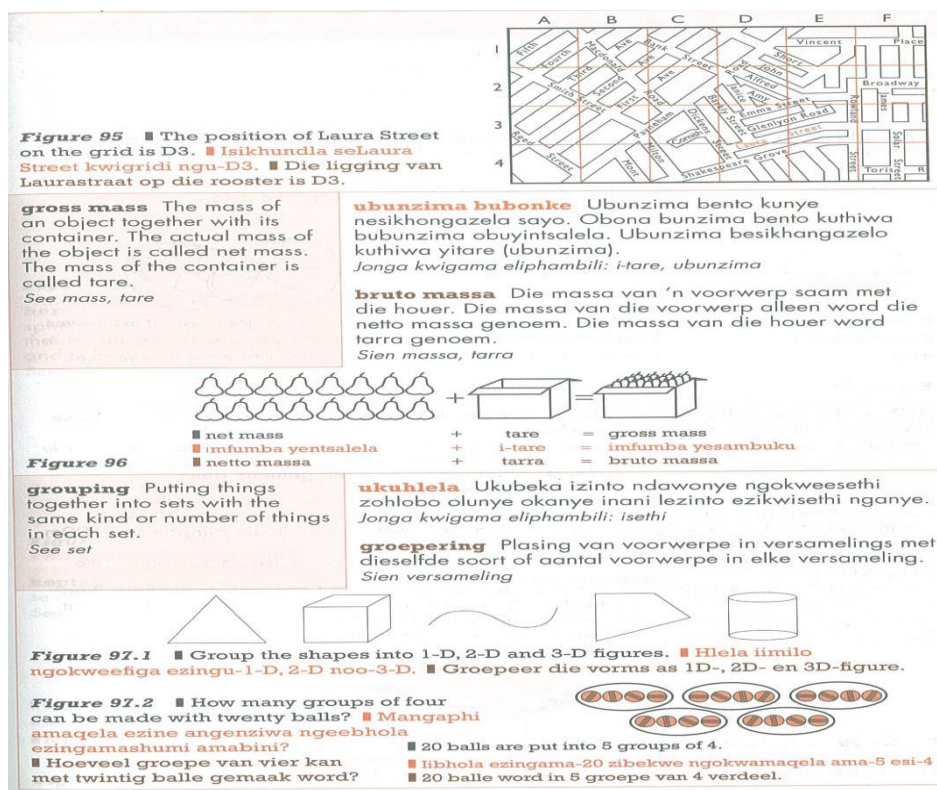


Figure 7b: Design features of *Longman Multilingual Maths Dictionary for South African Schools: English, isiXhosa, Afrikaans*

Young et al. (2005; 2009) also provide innovative examples of a functional approach to the design and structure of specialized lexicographical products. It is probably in view of the adopted structural design that the authors avoid calling their works dictionaries. The main text of *Understanding Concepts in Mathematics and Science Vol. 2* covers sixty-eight broad topics that are considered key concepts within five Grade 10–12 subject areas under the old National Curriculum Statement, namely Mathematics, Physics, Chemistry, Geography and Life Sciences. A thematic macrostructure is then adopted whereby those topics are grouped under specific units, as shown in Figure 8 below for Chemistry, Geography and Life Sciences.

Chemistry

Unit 34: Atomic models	229
Unit 35: Chemical bonding	237
Unit 36: Radioactivity	247
Unit 37: Quantitative analysis	252
Unit 38: Types of reactions	259
Unit 39: Energy and chemical change	268
Unit 40: Rates of reaction	272
Unit 41: Chemical equilibrium	276
Unit 42: Organic chemistry	284
Unit 43: Chemical systems	293
Unit 44: The kinetic theory of matter	295

Geography

Unit 45: The hydrosphere	301
Unit 46: Rivers and fluvial activity	305
Unit 47: Water management	314
Unit 48: Slopes and mass movement	319
Unit 49: Landforms	323
Unit 50: Climate	327
Unit 51: Weather systems	335
Unit 52: Atmospheric and water pollution	343
Unit 53: Energy and the environment	353

Life Sciences

Unit 54: Biological compounds	361
Unit 55: Proteins and enzymes	363
Unit 56: DNA	367
Unit 57: Genes and alleles	371
Unit 58: Chromosomes	373
Unit 59: Sexual reproduction, meiosis and fertilisation	375
Unit 60: Inheritance and Mendelian genetics	377
Unit 61: Cellular respiration	380
Unit 62: Gaseous exchange	386
Unit 63: Thermoregulation	388
Unit 64: Osmoregulation	390
Unit 65: The skeleton	392
Unit 66: Chemical and nervous coordination in humans	394
Unit 67: Survival and evolution	402
Unit 68: Population and population growth	405
Physical Sciences: Reference material	411
Index	420
Websites	421
References	422

Figure 8: A screenshot of an extract from the table of contents of Volume 2 of *Understanding Concepts in Mathematics and Science*

On a microstructural level, comprehensive treatment and clarity are prioritized ahead of space economy, with the multilingual and encyclopedic treatment of each concept taking several pages. Consider the treatment of the concept of probability shown in Figure 9 below. The treatment includes the presentation of etymological data, isiXhosa (X), Afrikaans (A) and isiZulu (Z) translation equivalents for each English (E) term followed by explanations in all four languages. Thereafter, more information follows about each topic, supported by examples and illustrative diagrams where necessary.

Unit 14: Probability

Etymology

Probability: L. *probare*, to show, prove, demonstrate

E English

Probability

X isiXhosa

Uqikelelo-
mathuba

A Afrikaans

Waarskynlikheid

Z isiZulu

Okuno-
kwenzeka

Explanation of probability

See *Understanding Concepts in Mathematics and Science, Volume 1, Unit 16.*

E **Probability** Probability is the measure of the **likelihood** of an event occurring. Probability is expressed using numbers from zero to one. Events that cannot happen (impossible events) have a probability of zero, while events that will definitely happen (certain events) have a probability of one. If an event has a probability of $\frac{1}{2}$ (0,5 or 50%), then it is equally likely to occur as not to occur.

X **Uqikelelo-mathuba** Uqikelelo-mathuba lubalo lokuqikelela amathuba okwenzeka kwento. Uqikelelo-mathuba luchazwa ngokusebenzisa amanani aphakathi kwe-0 nesi-1. Iziganeko ezinginakwehla zinoqikelelo-mathuba ali-0, ukanti uqikelelo-mathuba oluqinisekileyo lunoqikelelo-mathuba olusi-1. Ukuba isiganeko sinoqikelelo-mathuba olusisi-0,5 loo nto ithetha ukuthi amathuba okwehla kweso siganeko nawokungehli ayalingana.

A **Waarskynlikheid** Waarskynlikheid is die maatstaf van die **kans** dat 'n gebeurtenis sal plaasvind. Waarskynlikheid word met behulp van getalle van nul tot een uitgedruk. Gebeurtenisse wat nie kan plaasvind nie (onmoontlike gebeurtenisse), het 'n waarskynlikheid van nul, terwyl gebeurtenisse wat beslis sal plaasvind, 'n waarskynlikheid van een het. As 'n gebeurtenis 'n waarskynlikheid van $\frac{1}{2}$ (0,5 of 50%) het, is dit ewe waarskynlik dat dit sal plaasvind of nie sal plaasvind nie.

Z **Okunokwenzeka** Okunokwenzeka kuyisilinganiso **sokubanokwenzeka** kwesehlakalo. Kuchazwa ngokusebenzisa izinombolo eziphakathi kuka 0 no 1. Izehlakalo ezinganakwenzeka zinokuba zenzeke okungu 0, bese kuthi lezo eziqinisekile ukuthi zingenzeka zinokuba zenzeke okungu 1. Uma isehlakalo sinokuba senzeke okungu $\frac{1}{2}$ (0,5 noma 50%), kusho ukuthi ukuba kwenzeka nokungenzi kuyalingana.

Figure 9: The treatment of *probability* in Volume 2 of *Understanding Concepts in Mathematics and Science*

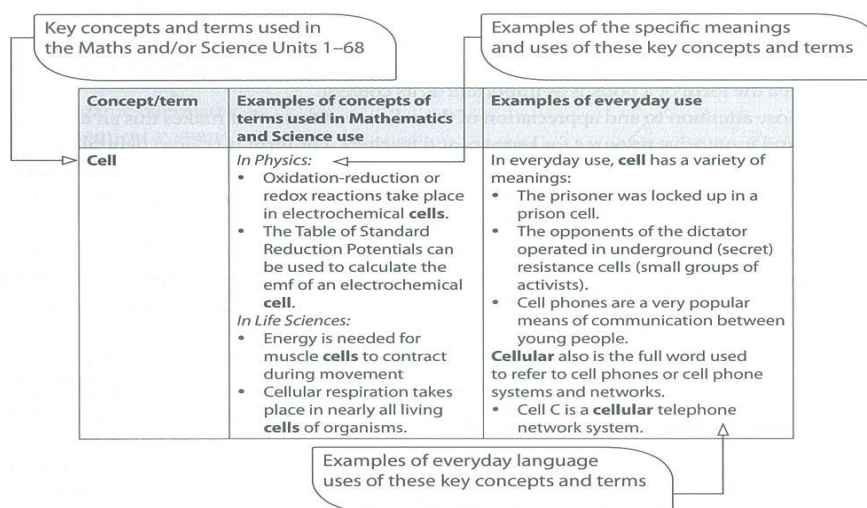
Besides the main text, there are useful outer-texts. Front matter texts include the following:

- How to use this book, a well-structured user guide that may help users as they familiarize themselves with the resource (see Figure 10 below)
- Multilingual contents, a table of contents listing unit titles in English and giving their translation equivalents in the four languages with their relevant pages

- Language matters: Concepts and terms, which provide scientific and everyday explanations of specific terms used within the broad topics covered in the main texts. The explanations are in English
- Language Matters: Task words, a text which provides isiXhosa, Afrikaans and isiZulu translation equivalents of instruction words used in teaching and assessment of the five subject areas, together with explanations in all the four languages.

How to use this book

1. Language matters: Specific concepts and terms and their everyday uses (pages 14–55)



2. Language matters: Task words used by educators/teachers, in learning materials and assessment tasks, in English (E), isiXhosa (X), Afrikaans (A) and isiZulu (Z) (pages 56–71)

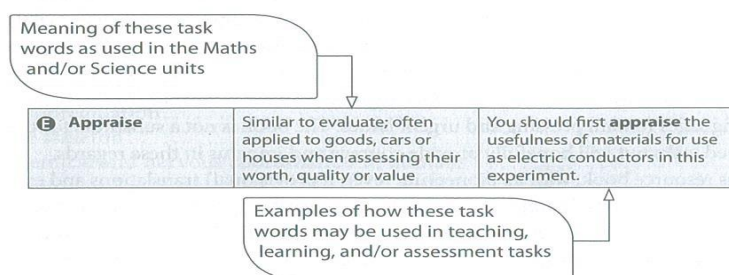


Figure 10: The user-guide of *Understanding Concepts in Mathematics and Science*

The main text is followed by back-matter texts, two of them being:

- Physical Sciences: Reference material, a text dealing with examination guidelines and providing tables of common quantities symbols and SI units used in the five subject areas
- Index, a text which provides quick access to the specific pages in which specific terms and concepts are discussed.

The other two provide a list of websites and references to other publications that were used in the entire project. Based on the foregoing, it may be seen that not only rich are the *Understanding Concepts in Mathematics and Science* volumes in terms of their provision of subject-specific content and language, they also facilitate access to this rich material through their diligent data distribution structure. They are poly-accessible through their different access routes and may be used together with textbooks or to some extent as textbooks. Once mastered by the users, their innovative and hybrid structural features could be critical in establishing a dictionary culture not only for specialized lexicography within the framework outlined in Gouws (2013) but for the utilization of a variety of lexicographical products.

7. Conclusion

From the preceding sections of this article, it can be seen that there is a proliferation of specialized lexicographical/terminographical products in African languages. It was noted that the production of such products is inspired by the language intellectualization endeavor that is aligned to post-colonial language policies which seek to eradicate the colonial and apartheid (in South Africa) legacy of marginalization of African languages. The main focus of the article was reflecting on lemma selection, which determines to a large extent the cognitive and communicative assistance that users may get from these products within the context of specialized disciplinary fields and subjects, especially at a terminological level, the provision of additional data to support the understanding and usage of terminology, as well as aspects of design and structures. On the first issue, the article noted problems with lemma selection, with some dictionaries including items that are of little if any value in the respective subject fields or disciplines, while compilers of other products do not seem to use criteria that facilitate the inclusion of important terms and avoid glaring gaps. Problems of insufficient information and irrelevant information were also noted, with the former being typical of oversimplified products that mainly provide only translation equivalents while the latter appears to be an uncritical application of principles of general-purpose dictionaries. Similarly, problems were noted regarding structural aspects of some resources, but positives were also noted in this respect. Overall, it is apparent that there is more room for improvement, with the main problem being that the production of such resources lacks a strong orientation from lexicographic theories. Most of the

surveyed resources were compiled largely by translators or at best terminologists, working in collaboration with educationists and disciplinary experts in the relevant fields, all of them united by enthusiasm or sympathy towards multilingualism and African languages. Thus the relevant lexicographic expertise for lemma selection, inclusion of data categories and presentation of data is lacking. Antia and Ianna (2016) express a similar concern regarding terminology work in South Africa, which they regard as mere translation exercise without engagement with ontological issues that inspire terminology as a discipline. The compilers of the studied resources and African language speech communities appear to be content about the mere availability of the resources in their languages, even though most of the products may not be effective and user friendly in usage real situations.

Acknowledgements

Financial support by the National Research Foundation of South Africa (Grant specific unique reference numbers UID108981) is hereby acknowledged. The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the author, and that the sponsor accepts no liability whatsoever in this regard.

References

- Antia, B. and B. Ianna. 2016. Theorising Terminology Development: Frames from Language Acquisition and the Philosophy of Science. *Language Matters* 47(1): 61-83.
- Bergenholtz, H. and S. Nielsen. 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.
- Bowker, L. and J. Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.
- Chimhundu, H. and E. Chabata. 2006. *Duramazwi reDudziramutauro neUvaranomwe*. Gweru: Mambo Press.
- Collins English Dictionary Online*. <https://www.collinsdictionary.com/submission/5703/Collins+English+Dictionary+online>. Date of Access: 22 August 2019.
- Department of Arts and Culture. 2013a. *Multilingual Mathematics Dictionary: Grade R–6*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture. 2013b. *Multilingual Financial Terminology List*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture. 2013c. *Multilingual HIV/Aids Terminology*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture. 2013d. *Multilingual Parliamentary/Political Terminology*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture. 2013e. *Multilingual Terminology for Information Communication Technology*. Pretoria: Department of Arts and Culture.

- Department of Arts and Culture.** 2013f. *Multilingual Human, Social, Economic and Management Sciences Terminology List*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture.** 2013g. *Multilingual Natural Sciences and Technology Term List (Nguni)*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture.** 2013h. *Multilingual Natural Sciences and Technology Term List (SeSotho)*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture.** 2013i. *Multilingual Natural Sciences and Technology Term List (Tshivenda–Xitshonga)*. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture.** 2013j. *Multilingual Soccer Terminology List*. Pretoria: Department of Arts and Culture.
- Department of Education.** 1997. *Language-in-Education Policy*. Pretoria: Department of Education.
- Department of Higher Education and Training.** 2002. *Language Policy for Higher Education*. Pretoria: Department of Higher Education and Training.
- Deyi, S., G. Minshall and T. Tokwe.** 2008. *Longman Multilingual Maths Dictionary for South African Schools: English, isiXhosa, Afrikaans*. Cape Town: Maskew Miller Longman.
- Feza, N. et al.** 2016. *IsiChazi-magama sezoBalo*. Pretoria: Human Sciences Research Council.
- Frescura, F. and J. Myeza.** 2016. *Illustrated Glossary of Southern African Architectural Terms: English–isiZulu*. Durban: University of KwaZulu-Natal Press.
- Fuertes-Olivera, P.A. (Ed.).** 2010. *Specialised Dictionaries for Learners*. Berlin: De Gruyter.
- 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.** 2013. Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, V. (Ed.). 2013. *Specialised Lexicography. Print and Digital, Specialised Dictionaries, Databases*: 51-62. Lexicographica Series Maior 144. Berlin/Boston: Walter de Gruyter.
- Hannan, M.** 1959. *Standard Shona Dictionary*. Salisbury: The Rhodesia Literature Bureau.
- IsiXhosa National Lexicography Unit.** 2013. *Isichazi-magama seMathematika neNzululwazi*. Cape Town: South African Heritage Publishers.
- Kaschula, R.H. and D. Nkomo.** 2019. Intellectualisation of African Languages: Past, Present and Future. Wolff, H.E. (Ed.). 2019. *The Cambridge Handbook of African Linguistics*: 601-622. Cambridge: Cambridge University Press.
- Khumalo, L.** 2018. Introduction. Zondi, K. 2018. *A Glossary of Law Terms: English–isiZulu*: ix-x. Durban: University of KwaZulu Natal Press.
- Kropf, A.** 1899. *Kafir–English Dictionary*. Alice: Lovedale Mission Press.
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography*. New York: Scribner Press.
- Łukasik, M.** 2016. Specialised Pedagogical Lexicography: A work in Progress. *Polilog: Studia Neofilologiczne* 6: 211-226.
- Mawonga, S., P. Maseko and D. Nkomo.** 2014. The Centrality of Translation in the Development of African Languages for Use in South African Higher Education Institutions: A Case Study of a Political Science English–isiXhosa Glossary in a South African University. *Alternation* 13: 55-79.
- Mbude-Shale, N., Z. Wababa and K. Welman.** 2008. *Illustrated Multilingual Science and Technology Dictionary / Isichazi-magama sezeNzululwazi neTeknoloji Ngeelwimi Ezininzi*. Cape Town: New Africa Books.
- Mheta, G. (Ed.).** 2005. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
- Ministerial Committee appointed by the Ministry of Education.** 2003. *Ministerial Report on the Use of Indigenous Languages in Higher Education*. Pretoria: Department of Higher Education Training.

- Mpofu, N., H. Chimhundu, E. Mangoya and E. Chabata (Eds.).** 2004. *Duramazwi reUrapu neUttano*. Gweru: Mambo Press.
- Nielsen, S.** 1995. Dictionary Components. Bergenholtz, Henning and Sven Tarp (Eds.). 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries: 167-187*. Amsterdam/Philadelphia: John Benjamins.
- Nkomo, D.** 2010. Affirming a Role for Specialised Dictionaries in Indigenous African Languages. *Lexikos* 20: 371-389.
- Nkomo, D.** 2018. Dictionaries and Language Policy. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography: 152-165*. New York: Routledge.
- Nkomo, D. and M. Madiba.** 2011. The Compilation of Multilingual Concept Literacy Glossaries at the University of Cape Town: A Lexicographical Function Theoretical Approach. *Lexikos* 21: 144-168.
- Nkomo, D. and N. Moyo.** 2006. *Isichazamazwi SesoMculo*. Gweru: Mambo Press.
- Sibula, P.** 2007. *Isigama Somthetho / Law Terminology / Regsterminologie*. Stellenbosch: Stellenbosch University Language Centre.
- Simango, S.R.** 2009. Weaning Africa from Europe: Toward a Mother-tongue Education Policy in Southern Africa. Brock-Utne, B. and I. Skattum (Eds.). 2009: 201-212. *Languages and Education in Africa: A Comparative and Transdisciplinary Analysis*. Oxford: Symposium Books.
- Taljard, E. and D.J. Prinsloo.** 2019. African Language Dictionaries for Children — A Neglected Genre. *Lexikos* 29: 199-223.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2005. The Pedagogical Dimension of the Well-conceived Specialised Dictionary. *Iberica. Journal of the European Association of Languages for Specific Purposes* 10: 7-21.
- 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.** 2012. Specialised Lexicography: 20 Years in Slow Motion. *Iberica. Journal of the European Association of Languages for Specific Purposes* 24: 117-128.
- Van Niekerk, T. and J. Wolvaardt (Eds.).** 2010. *Oxford South African Concise Dictionary*. Second edition. Cape Town: Oxford University Press Southern Africa.
- Wababa, Z., K. Welman and K. Press (Eds.).** 2010. *Isichazi-magama seziBalo Sezikolo saseCambridge*. Cape Town: Cambridge University Press.
- Young, D. et al.** 2005. *Understanding Concepts in Mathematics and Science — A Multilingual Learning and Teaching Resource Book in English, IsiXhosa, IsiZulu and Afrikaans. Volume 1*. Cape Town: Maskew Miller Longman.
- Young, D. et al.** 2009. *Understanding Concepts in Mathematics and Science — A Multilingual Learning and Teaching Resource Book in English, isiXhosa, isiZulu and Afrikaans. Volume 2*. Cape Town: Maskew Miller Longman.
- Zondi, K.** 2018. *A Glossary of Law Terms: English–isiZulu*. Durban: University of KwaZulu Natal Press.

Die behandeling van die funksie dekodering in verskillende tipes woordeboeke*

Anna Nel Otto, *Departement Tale en Letterkunde, Nelson Mandela Universiteit, Port Elizabeth, Suid-Afrika (anna.otto@mandela.ac.za)*

en

Jadé Blume, *Departement Tale en Letterkunde, Nelson Mandela Unversiteit, Port Elizabeth, Suid-Afrika (jade.blume@mandela.ac.za)*

Opsomming: Woordeboeke word veral geraadpleeg ter wille van die funksie dekodering. Hierdie artikel verskaf 'n sistematiese beskrywing van die invloed wat die funksie dekodering op die woordeboekstrukture en datatipes in verskillende woordeboektipes het. Tydens hierdie bespreking word veral aandag gegee aan strukture wat in sowel gedrukte as aanlyn woordeboeke voorkom. Alhoewel die belangrikste datatipe vir dekodering betekenisverklarings/vertalings in meertalige woordeboeke is, fokus hierdie artikel veral op die rol van datatipes soos uitspraakleiding, kollokasies, etikette, voorbeelde en etimologiese leiding. In gedrukte woordeboeke is daar 'n groot ooreenkoms ten opsigte van raamstruktuur (minstens sentrale woordelys en gebruikersleiding), dataverspreidingsstruktuur en toegangstruktuur, terwyl verskille veral op die vlak van die makrostruktuur (meer of minder lemmas, verskillende ordenings) en mikrostruktuur (aanduidertipes en hoeveelheid data ten opsigte hiervan) voorkom.

Sleutelwoorde: DEKODERING, WOORDEBOEKTIPES, WOORDEBOEKSTUKTUUR, WOORDEBOEFUNKSIE, OMOVATTENDE WOORDEBOEK, STANDAARDWOORDEBOEK, BEPERKTE WOORDEBOEK, AANLEERDERWOORDEBOEK, WOORDEBOEK VIR SPESIALE DOELEINDES, TWEETALIGE WOORDEBOEK

Abstract: The Treatment of the Function Decoding in Different Types of Dictionaries. Dictionaries are especially consulted for the function of decoding. This article provides a systematic description of the influence that this function has on the dictionary structures and data types in different types of dictionaries. In this discussion attention is paid to structures which appear in both printed and online dictionaries. Although the most important data type for decoding is meaning explanations/translation equivalents in multilingual dictionaries, this article focuses especially on the role of data types such as pronunciation guidance, collocations, labels, illustrations and etymological data. In printed dictionaries there is a resemblance in terms of frame structures (at least a lemma list and usage guidance), data distribution structure and access structure, while differences are more on the level of the macrostructure (quantity of lemmata and different ordering) and microstructure (indicator types and quantity of data).

* Hierdie artikel is 'n verwerkte weergawe van 'n ongepubliseerde M.A.-verhandeling, Blume 2017.

Keywords: DECODING, DICTIONARY TYPES, DICTIONARY STRUCTURE, DICTIONARY FUNCTION, COMPREHENSIVE DICTIONARY, STANDARD DICTIONARY, RESTRICTED DICTIONARY, LEARNER'S DICTIONARY, DICTIONARY FOR SPECIAL PURPOSES, BILINGUAL DICTIONARY

1. Inleiding

Een van die belangrikste funksies in woordeboeke is **dekodering**. Die woord **dekodering** moet hier verstaan word as die verkryging van begrip wanneer 'n gebruiker byvoorbeeld 'n onbekende woord naslaan en dit probeer verstaan (Svensén 2009: 12). Dit word in die eerste instansie vergestalt in definisies/betekenisverklarings in woordeboekkonteks. Morris (1985: 6) verwoord dit indirek soos volg: "The soul of a dictionary, however, is the quality of its definitions."

Die doel van hierdie artikel is om 'n sistematiese beskrywing te verskaf van die invloed wat die funksie dekodering op die woordeboekstrukture en datatipes in verskillende woordeboektipos het. Aangesien hierdie doel verband hou met die daarstel van beter woordeboeke, val dit binne die leksikografie as dissipline. Binne die leksikografie is daar vir 'n lang tydperk gefokus op tradisie en woordeboekinhoud. 'n Tipiese tweetalige woordeboek het byvoorbeeld hoofsaaklik gefokus op die verskaffing van vertaalekwivalente. Mettertyd het leksikograwe meer en meer bewus geword van die feit dat teikengebruikers woordeboeke vir verskillende funksies en in verskillende situasies gebruik. Sommige gebruikers stel slegs in betekenisverklarings (vir dekodering) belang, terwyl ander graag woorde korrek binne konteks wil gebruik (vir produksie). Die rasionaal vir hierdie artikel is om die leemte te vul ten opsigte van 'n sistematiese beskrywing van die invloed wat die funksie dekodering op die woordeboekstrukture en datatipes in verskillende woordeboeke het. Terwyl die belangrikste datatipe vir dekodering betekenisverklarings/vertaalekwivalente in meertalige woordeboeke is en reeds redelik uitvoerig in die literatuur bespreek is, fokus hierdie artikel veral op die rol van datatipes soos uitspraakleiding, kollokasies, etikette, voorbeelde, illustrasies en etimologiese leiding.

Dekodering is maar een van verskeie funksies wat in woordeboeke voorkom. Binne die metaleksikografie is die verskillende funksies reeds geïdentifiseer en die aard daarvan aangetoon, veral deur aanhangers van die funksieteorie (vergelyk onder andere Bergenholtz en Tarp 2003), maar 'n sistematiese beskrywing van die verhouding tussen 'n bepaalde funksie, woordeboekstrukture en datatipes in verskillende woordeboektipos is eers in 2017 voltooi in die vorm van 'n verhandeling, wat fokus op die funksies dekodering, produksie en vertaling (Blume 2017).

In die leksikografie word dit wyd aanvaar dat woordeboeke gespesifiseerde teikengroepe se behoeftes en naslaanvaardighede in ag moet neem. Dit het implikasies vir die funksie(s) wat die woordeboek moet hê en die daarmee gepaardgaande strukture, databehandeling en woordeboektype. In hierdie ver-

band wys Gouws (2018: 180) op die voorkeur van Generasie Z (persone wat na 2000 gebore is) om al hulle data van een bron te verkry, veral die internet. Aanpassings by die digitale omgewing is dus belangrik vir metaleksikograwe (Gouws 2018: 180). Individue binne ander generasies het nie noodwendig dieselfde behoeftes as Generasie Z nie, gevolglik word daar saamgestem met die siening van Gouws (2018: 181) dat aandag gegee moet word aan strukture wat in sowel gedrukte as aanlyn woordeboeke voorkom, maar bepaalde aanpassings in aanlyn woordeboeke vereis, en ook aan die strukture wat slegs in een van die twee tipes voorkom. Dziemianko (2017) wys ook in hierdie verband op die feit dat daar nog nie statisties beduidende bewyse is dat die vorm (gedruk teenoor elektronies) 'n deurslaggewende rol speel by dekodering en produksie nie, maar dat veral vir aanleerders die elektroniese formaat beter blyk te wees vir kitshulp om taalprobleme op te los en vir retensie (Dziemianko 2017: 349).

Die werkswyse is om aan die hand van 'n literatuurstudie en illustrasievoorbeelde te peil hoe die strukture en data-inhoude van woordeboektipes soos omvattende woordeboeke, standaardwoordeboeke, tweetalige woordeboeke, beperkte woordeboeke en woordeboeke vir spesiale doeleindes verpak behoort te wees ten einde suksesvolle dekodering in die hand te werk.

2. Omvattende woordeboeke

Omvattende woordeboeke se primêre funksie is om die taal so volledig moontlik te dokumenteer. Naas dekodering is die kennisgerigte funksie ook ter sake in omvattende woordeboeke. Die kennisgerigte funksie is relevant met betrekking tot die raamstruktuur in gedrukte woordeboeke, terwyl daar veral in 'n aanlyn woordeboek na eksterne bronne, byvoorbeeld die internet, verwys kan word, en in hierdie verband is die datatrekkingstruktuur van belang, waarvolgens gebruikers vanaf enige soeksonde data aan die internet kan onttrek (Gouws 2018: 192).

Wat data-inhoud betref, dui Gouws (2017: 442) aan dat daar geen beperking is op die hoeveelheid lemmas by 'n omvattende woordeboek nie, en Tarp (2008) beklemtoon die relevansie van ongereelde fleksievorme, idioome en spreekwoorde as lemmata wat verklaar behoort te word.

2.1 Die raamstruktuur

Die sentrale woordelys is die belangrikste, verpligtende element in die gedrukte woordeboek (Gouws 2017: 447). Dit moet alfabeties georden word en sluit in enige data wat relevant vir die funksie dekodering is. Die voorteks behoort ook 'n gebruikersleiding te bevat (Gouws 2017: 447).

Die gebruikersleiding moet duidelik aandui hoe sekere datatipes opgespoor kan word. In hierdie verband dui Svensén (2009: 382) die volgende vereistes aan: Daar moet 'n volledige verduideliking van die makro-, mikro- en mediostruktuur van die woordeboek wees, sowel as hoe die verskillende data-

tipes aangebied word. Die manier hoe struktuurmerkers werk, moet aangedui word, en 'n lys redaksionele afkortings met hulle volvorme is nodig. In elektroniese woordeboeke is artikels noodsaaklik en daarnaas ook gebruikersleiding deur middel van skermkote ten einde data optimaal te benut. Dit kan aangevul word deur 'n rubriek *gereelde vrae* of 'n *begeleide toer*, wat die gebruikers deur middel van 'n voorbeeldsoektog touwys maak. Laastens is daar ook die moontlikheid om tydens probleemsituasies 'n leermeester op interaktiewe wyse te raadpleeg (Kemmer 2010: 6).

Enige bykomende data wat belangrik is vir die funksie dekodering, soos ensiklopediese en/of kulturele data, kan verder in die agterteks in gedrukte woordeboeke verskyn, indien dit nie reeds in die artikels opgeneem is nie. Dit is wel waarskynlik dat data soos geografiese name en mitologiese figure in die sentrale teks betrek word omdat omvattende woordeboeke gewoonlik 'n meerdelige produk is, waar een of meer letters by elke deel betrek word en hierdie dele op verskillende tye verskyn (Gouws 2017: 442). In elektroniese woordeboeke kan hierdie data in tekste rondom die sentrale teks verskyn, aangesien dit nie lineêr georden is nie.

2.2 Die dataverspreidingstruktuur

Die teikengebruiker moet die nodige inligting vir dekodering maklik kan onttrek, hoofsaaklik uit die woordeboek se sentrale teks. Relevante bykomende data kan verder in die voor- en/of agterteks verskyn. In die WAT (deel XII) word byvoorbeeld die volgende aspekte in die voorteks hanteer: inleiding, gebruiksleiding en redaksionele afkortings. Die gebruiksleidingafdeling dui vir die teikengebruikers aan watter tipes data in die WAT verskyn en waar in die woordeboekartikels die data opgeneem word. Hierdie toeligting help die gebruiker om lemmas en sublemmas vinniger op te spoor en beter te begryp. Sinonieme en kruisverwysings kan byvoorbeeld begrip bevestig indien die betekenisverklaring nie volkome duidelik is nie. In die toeligting word onder andere **rostrum** as sinoniem van **podium** aangedui. Verder word daar aangedui dat kruisverwysings deur "Vgl." ingelei word. Onder **pikbroek** staan byvoorbeeld: Vgl. MATROOS, SEEROB. Die gebruiker kan ook kennis neem van die feit dat vorme wat soms met die lemma verwar word, deur "onderskei van" aangedui word en dat "OPM." vir verdere verduidelikings gebruik word.

2.3 Die makrostruktuur

Die opname van ensiklopediese en kulturele data in 'n omvattende woordeboek sal die gebruiker help om 'n betrokke woord in die korrekte konteks te verstaan. Klosa (2016: 523) noem dat kulturele data in die betekenisverklarings, voorbeelde en kollokasies opgeneem kan word. Die aanbevelings vir die opname en bewerkings van lemmas in 'n omvattende woordeboek is soos volg, ondersteun deur data uit Svensén (2009: 368-378), Gouws (2017: 442-444), Béjoint (2016: 18),

Klosa (2016: 524-530) en Kemmer (2010: 8):

- Soveel lemmas as moontlik moet opgeneem word.
- Die lemmakeuse moet uit die standaardvariëteit geneem word, asook uit ander variëteite. Etikette moet gebruik word waar daar 'n afwyking van die standaardtaal is.
- In gedrukte woordeboeke is die ordening gewoonlik alfabeties.
- In elektroniese woordeboeke kan naas die alfabetiese lemmalyste, wat voorafgaande en daaropvolgende lemmas kan aandui, ook die volgende verskaf word: óf (soos in tesourusse) parallelle lemmalyste wat volgens woordvelde georganiseer is, óf 'n lemmalys wat volgens frekwensie gebaseer is (Kemmer 2010: 8).

2.4 Die mikrostruktuur

Die artikels, met hulle geordende inskrywings, vorm die mikrostruktuur (Gouws 2017: 452). In ware aanlyn woordeboeke, wat nie net 'n duplisering is van die data in 'n gedrukte woordeboek nie, is daar nie 'n soortgelyke ordening as in 'n gedrukte woordeboek nie.

Die invloed wat die funksie dekodering op die mikrostruktuur in gedrukte omvattende woordeboeke het, word vervolgens bespreek.

2.4.1 Aanduiders en struktuurmerkers

Aanduiders (data) en struktuurmerkers is albei ter wille van dekodering nuttig in alle tipes gedrukte woordeboeke. Aanduiders sluit in onder andere woordsoorte, meervoudsvorme en verkleiningsvorme (Gouws 2003: 35 en Gouws 2017: 452). Tarp (2008: 77) noem dat onder andere woordsoorte, woordkombinasies en fleksievorme datatipes is wat vir dekodering opgeneem behoort te word.

Struktuurmerkers help die gebruiker om so vinnig en so maklik moontlik by 'n spesifieke item uit te kom (Gouws 2017: 453). Tipografiese struktuurmerkers sluit in leksikale items wat in vetdruk, kursief, ensovoorts aangedui is, terwyl nietipografiese struktuurmerkers gewoonlik simbole is, byvoorbeeld 'n klein blokkie of omgekeerde driehoek, wat die gebruiker help om vinnig by 'n item uit te kom. Ronde hakies word byvoorbeeld onder andere as nietipologiese struktuurmerker in die WAT XII gebruik om betekenis te presiseer, soos byvoorbeeld in die volgende geval: **parlement... b** spesifieke samestelling van 'n parlement (PARLEMENT 1 a i).

2.4.2 Uitspraakleiding

In gevalle waar 'n betrokke woord verskillende uitsprake het, behoort die gebruiker uitspraakleiding te kry om tussen die twee betekenis te kan onderskei, soos in die WAT X by **oorweeg**.

2.4.3 Kollokasies

Kollokasies is woorde wat dikwels saam gebruik word en nie 'n vrye verbinding of vaste uitdrukking is nie (Gouws 1989: 97; Bentivogli en Pianta 2002: 786). Volgens Tarp (2008: 254) moet kollokasies nie in 'n woordeboek vir die funksie dekodering opgeneem word nie omdat die aparte dele van die kollokasie vir teksbegrip opgesoek kan word. Otto (2013: 189) en Howarth (1998: 101) dui wel die belang van kollokasies vir onder andere moedertaalsprekers aan, aangesien die praktyk bewys lewer dat moedertaalsprekers probleme ondervind met die korrekte begrip en gebruik van kollokasies. Die gebrekkige begrip van 'n bepaalde kollokator kan tot kontaminasievorme lei, soos byvoorbeeld in die geval van **nood oorbrug**, waar die bedoeling **nood verlig** is en dit verwar word met 'n **probleem oorbrug**.

2.4.4 Etiket

Volgens Gouws (1988: 4) verwys die term etiket "na daardie inskrywing(s) in 'n woordeboekartikel wat as merkers optree om die lemma of 'n ander inskrywing in die artikel se beperkings ten opsigte van byvoorbeeld stylaard, gebruiksfêre en geografiese voorkoms aan te dui". In die WAT XI word byvoorbeeld by **opstopper** deur middel van die etiket (*boks*) onder andere onderskei tussen 'n harde vuishou deur enigiemand en 'n vuishou deur 'n bokser as 'n besondere tipe vuishou.

2.4.5 Illustrasies

Ilsou (1987: 193) dui, met betrekking tot gedrukte woordeboeke, aan dat die begrip "illustrasies" as dekkingsterm dien vir: tabelle, diagramme sowel as prente, met ander woorde vir alle vorme van nielineêre aanbieding waarin die vorm van die voorstelling die items of hulle verbande met mekaar voorstel.

Hoewel dit nuttig kan wees vir dekodering, moet illustrasies in die lig van die hoë onkoste daarvan spaarsamig in gedrukte woordeboeke gebruik word. In elektroniese woordeboeke voeg illustrasies waarde toe, veral ten opsigte van dekodering deur middel van teks, prente, film, geluide, animasie en simulasie (Sager 2000: 588).

2.5 Die toegangstruktuur

Die eksterne toegangstruktuur, wat die soekroete is wat die gebruiker volg om by 'n bepaalde datatipe uit te kom, sal die omslag/naam van die woordeboek, die gebruikersleiding en loopkoppe (in gedrukte woordeboeke) wees.

Bothma (2011: 82-83) dui aan dat die tipiese soekroete in 'n e-woordeboek presies gee waarvoor die gebruiker vra en dat 'n gevorderde soekroete ook

moontlik is. Skakels kan gebruikers help om vinnig tussen kruisverwysings rond te beweeg om data vir dekodering op te spoor, soos in die volgende afdeling aangetoon sal word. Die bruikbaarheidsbenadering is hier van belang (vergelyk Du Plessis 2017: 2).

2.6 Die mediostruktuur

Die mediostruktuur verwys die gebruiker na verwante data deur die gebruiker tussen verskillende dele van 'n woordeboek te neem (Svensén 2009: 388). Volgens Gouws (2018: 186) kan die gebruiker ook na verwante data in 'n ander woordeboek in 'n woordeboekportaal of eksterne kruisverwysingsadresse verwys word. Dit kan op enige plek in die woordeboekartikel voorkom (Bogaards 2013: 25).

Die mediostruktuur het verskillende funksies, waaronder veral ruimtebesparing in gedrukte woordeboeke, maar dit kan terselfdertyd semantiese verbande aantoon. In die WAT XII word byvoorbeeld soms kruisverwysings na sinonieme aangebied. By **padren** word na **padwedren** verwys en die definisie verskyn onder **padwedren**. 'n Wisselvormverwysing het die aanloop "Sien". By **paarsgewyse**, byvoorbeeld, word die kruisverwysing aangedui as: "Sien PAARSGEWYS..." Die definisie verskyn by **paarsgewys**. Daar kan ook kruisverwysings na formele antonieme en verwante of maklik verwarbare leksikale items wees. Antonieme word in die WAT XII deur "teenoor" aangedui, terwyl vorme wat verwar kan word, ingelei word deur "onderskei van". Die antonimiese verwysing by **palingenese 3**, byvoorbeeld, word aangedui as "onderskei van **anateksis**". Dit is veral laasgenoemde vorme wat ter wille van dekodering duidelik van mekaar onderskei moet word.

In e-woordeboeke word die mediostruktuur vergestalt in skakels, wat die gebruiker help om die datatipe waarna verwys word, maklik en vinnig op te spoor (Oppentocht en Schutz 2003: 218). Ten einde dit te kan doen, moet die woordeboekgebruikers weet op watter vlak hulle hulle bevind indien die betrokke omvattende woordeboek ingebed is in 'n woordeboekportaal. Daar moet dus 'n soekenjin wees wat sentraal op die tuisblad georiënteer is en wat aandui hoe daar op elke woordeboekvlak gekom kan word (Klosa, Lemnitzer en Neumann 2008:10).

3. Standaardwoordeboeke

Die meeste gedrukte standaardwoordeboeke het meer as een funksie omdat hulle dikwels help met dekodering en produksie. Die *Verklarende Afrikaanse Woordeboek* (VAW) is primêr op spellingleiding en begrip gerig, maar verskaf wel soms frases, hoewel nie naastenby soveel, en veral volsinne, soos die *Handwoordeboek van die Afrikaanse Taal* (HAT6) nie. Terwyl op die omslag van die HAT6 aangedui word dat daar meer as 80 000 voorbeelde van tipiese taalgebruik is, word in die voorwoord van die VAW gemeld dat die VAW veral op

die trefwoord en sy gepaardgaande grammatikale inligting fokus en dat daar 8000 frases is.

Aangesien gedrukte standaardwoordeboeke duur is, sal gebruikers moontlik soveel as moontlik relevante inligting vir dekodering en produksie daarin wil hê. In e-woordeboeke is dit nie 'n probleem nie omdat slegs die dekoderingsoptie of die produksie-optie (naas ander potensiële opsies) gekies kan word. Die lemmas wat in 'n standaardwoordeboek opgeneem word, kom hoofsaaklik uit die standaardvariëteit van die betrokke taal (Gouws 2017: 444).

3.1 Die raamstruktuur

'n Gedrukte standaardwoordeboek moet sowel 'n sentrale woordelys as minstens 'n voorteks met gebruikersleiding bevat en 'n titelblad en inhoudsopgawe word ook aanbeveel (Ptaszynski 2009: 249). Ptaszynski (2009: 249) argumenteer dat buitetekste (insluitend die titelblad en inhoudsopgawe) funksiebepaald is. Ptaszynski (2009: 249) meld verder dat die funksie van 'n betrokke woordeboek die struktuur en inhoud van hierdie datatipes bepaal, maar nie die opname daarvan nie omdat gesê kan word dat enige woordeboek hierdie buitetekste kan bevat, ongeag die funksie van die woordeboek.

Die inhoudsopgawe kan deur sowel die funksie as die woordeboek-medium beïnvloed word (Ptaszynski 2009: 249-250). 'n Inhoudsopgawe in 'n gedrukte woordeboek kan met 'n keuselys in 'n e-woordeboek vervang word. Fuertes-Olivera en Niño-Amo (2011: 172) noem dat 'n keuselys wat maklik is om te gebruik, noodsaaklik is omdat die gebruikers data vinnig moet kan opspoor. Met hierdie keuselys beweeg gebruikers tussen datatipes rond (Bothma 2011: 81, 83 en Fuertes-Olivera 2013: 331).

3.2 Die dataverspreidingstruktuur

Hierdie struktuur stem gewoonlik ooreen met dié van omvattende woordeboeke.

3.3 Die makrostruktuur

Die kriteria vir die opname van lemmas in standaardwoordeboeke is deur, onder andere, Swanepoel (2003: 56), Svensén (2009: 368-378) en Bergenholtz en Tarp (2003: 71-72) aangetoon. Die lemmakeuse moet hoofsaaklik uit die standaardvariëteit van die taal gemaak word (weens ruimtekewessies) tensy dit van belang is vir dekodering (De Caluwé en Van Santen 2003: 72; Gouws 2017: 444). Die gebruiker moet al die data vir dekodering maklik kan opspoor (Bogaards 2003: 29), wat beteken dat die leksikograaf hoofsaaklik 'n streng alfabetiese makrostruktuur met groepering (sublemmas) sal moet gebruik (Svensén 2009: 370-371). Bergenholtz en Gouws (2013: 72) verduidelik dat samestellings veral

belangrik vir dekodering in 'n standaardwoordeboek is. Indien die gebruiker die werklike betekenis van 'n samestelling wil begryp en slegs die aparte stamme word as lemmas opgeneem, kan die gebruiker ten opsigte van ondeursigtige samestellings 'n begrip-probleem ervaar. Daar moet dus gebruik gemaak word van verklaarde samestellings as hooflemmas, bv. *donskoring*, *skrikkeljaar*, *skutsluis* en *vruggebruik*.

Sublemmas kan slegs deur ander lemmas bereik word, en indien 'n woordeboek 'n kronkelordering toon as gevolg van die horisontale opname van sublemmas, vorm dit 'n artikelkluster (Gouws 2017: 450). Verder kan nislemmas effektief in standaardwoordeboeke vir begrip gebruik word deur die morfologies verwante woorde op te neem (Gouws 2017: 450-451). 'n Voorbeeld hiervan in die HAT6 is **inplons**, gevolg deur **inplooi** op die volgende reël en binne **inplooi** se artikel kom horisontaal die afleiding **inplooiing** voor, gevolg deur **inplooiwerk** op die volgende reël as nuwe lemma.

Neslemmas kom voor waar daar 'n afwyking in die alfabetiese ordeningsstruktuur is. Gouws (2017: 450) dui aan dat neslemmas steeds alfabeties in die kluster gehandhaaf kan word, maar dat die daaropvolgende vertikale lemma dan dikwels nie alfabeties georden is nie (eerstevlakse nesting), of dat die alfabetiese ordening in die kluster self kan ontbreek (tweedevlakse nesting). 'n Voorbeeld van eerstevlakse nesting in die HAT6 is die volgende: **redeloos**, met **redenaar** op die volgende reël en horisontaal die samestellings **redenaars: ~gawe**, **~kuns**, **~talent**, met **redenaarskompetisie** niealfabeties op die volgende reël. Hierteenoor kom tweedevlakse nesting in die HAT6 voor by **hoender**, en wel soos volg: **hoender** op die eerste reël, gevolg deur 'n hele aantal samestellings op horisontale vlak beginnende met **~hoenderafval** en alfabeties tot by **~hoendersop**, gevolg deur die afleiding met verskillende klempatrone, naamlik **hoenderagtig**, **hoenderagtig** en **hoenderbors** op die volgende reël as volgende lemma. Die alfabetiese ordening word tussen **hoendersop** en **hoenderagtig** (in die kluster) onderbreek.

3.4 Die mikrostruktuur

3.4.1 Aanduiders en struktuurmerkers

Vir dekodering is veral semantiese leiding deur middel van betekenisverklarings, woordsoortaanduiding en soms uitspraakleiding, kollokasies en etikette nodig.

Die onderstaande uittreksel vanuit die *VivA*-webwerf gee die elektroniese resultaat uit die HAT6 vir die lemma **kerk**, en dien ter illustrasie van nietipografiese struktuurmerkers in die elektroniese weergawe van die HAT6:

Uitdrukings

Is jy in die kerk gebore?

gevra aan iemand wat nie 'n deur agter hom/haar toemaak nie.

daar is geen kerk met iem. te hou nie

iemand is onbeheerbaar.

die koeël is deur die kerk > koeël

iem. se kerk is uit

dis klaarpraat met iemand.

■ ww. [het gekerk] [veral voltooide deelwoord]

in die kerk getroud raak: ◇ *Vier saamblypaartjies van die plaas Hooggelegen is nou gering, gekerk en behoorlik getroud!* (Rene Erasmus).

◇ *(Maar wat praat ek nou, sy van was nie haar van nie, hulle is nie gekerk nie* (Elsa Joubert).



Die blokkie word voor woordsoorte geplaas en die ◇ voor sitate.

Figuur 1: Die nietipografiese struktuurmerkers vanuit *VivA*

3.4.2 Uitspraakleiding

Daar is gevalle waar uitspraakleiding wel relevant vir dekodering is. Vergelyk byvoorbeeld die betekenisverskil by *oorgiet* in die VAW waar *oorgiet* met die klem op oor- beteken "uit een voorwerp in 'n ander gooi" of "weer giet", teenoor *oorgiet* met die klem op -giet wat "bedek met iets wat uitgegiet word, sprei" beteken.

3.4.3 Kollokasies

L'homme en Bertrand (2000: 498) argumenteer dat kollokasies onvoorspelbaar is, wat beteken dat hulle, vir dekodering, in 'n woordeboek, selfs in 'n standaardwoordeboek, opgeneem moet word. Moedertaalsprekers kan ook probleme met kollokasies hê (Braasch en Olsen 2000: 475) en in die HAT6, byvoorbeeld, word kollokasies in 'n artikel duidelik met die simbool | gemerk (sien uittreksel hieronder):

dubbele standaard model waarvolgens (...) enkele standaard standaard met alleen (...)
dubbele standaard standaard waarby sowel (...)

Figuur 2: Die aanduiding van die kollokasies met *standaard*

3.4.4 Etiket

Etiket moet in standaardwoordeboeke vir begrip betrek word, veral by polisemiese items. In die HAT help etikette byvoorbeeld om die verskillende betekenisonderskeidings van **papegaai** van mekaar te onderskei:

pa.pe.gaai.bek *s.nw.* [-ke] **1** bek van 'n papegaai. **2** (*veearts.*) afwyking by skape, waarby die bokaak oor die onderkaak uitsteek. **3** (*tuinb.*) groot, sterk soort boomskeer... **4** moersleutel vir pype.

3.5 Die toegangstruktuur

Die toegangstruktuur van standaardwoordeboeke stem ooreen met dié van omvattende woordeboeke. Vergelyk verder Gouws (2017: 449) en Svensén (2009: 441).

4. Beperkte woordeboeke

Gouws (2017: 441) sê dat 'n beperkte woordeboek 'n beperktheid het ten opsigte van lemmakeuses wat opgeneem word. 'n Idioom- of sinoniemwoordeboek bevat byvoorbeeld slegs 'n gedeelte van die betrokke taal se leksikon.

4.1 Die raamstruktuur

Die raamstruktuur van 'n beperkte woordeboek hang af van die datatipes wat daarin opgeneem word, want die datatipes en hulle bewerking kan, soos Swanepoel (2003: 58) en Gouws (2017: 442) aantoon, ook beperk wees.

4.2 Die dataverspreidingstruktuur

Die dataverspreidingstruktuur van 'n beperkte woordeboek is soortgelyk aan dié van 'n woordeboek vir spesiale doeleindes (voortaan WSD) omdat dit ook beperkte datatipes kan bevat.

4.3 Die makrostruktuur

Die lemmakeuse sal slegs uit 'n deel van die taal se leksikon bestaan (Swanepoel 2003: 58). In die geval van die gedrukte *Idiomewoordeboek* (De Villiers en Gouws 1988) word 'n beperkte hoeveelheid lemmas opgeneem. In die elektroniese Afrikaanse idiomewoordeboek kan 'n baie groter getal idiome opgeneem word weens die feit dat 'n elektroniese databasis nie ruimteprobleme sal hê nie. Die gebruikers kan self kies watter lemma en datatipes hulle wil opsoek volgens die funksies wat hulle benodig (Ball 2016: 225).

4.4 Die mikrostruktuur

In die geval van *Idiomewoordeboek* word naas die lemma en trefbegrip wat die idioom in 'n algemene betekenisveld plaas, slegs die idiome, hulle betekenis en Engelse vertalings verskaf, soos byvoorbeeld in die geval van *wans*:

wans DADELIK Uit wans uit: Uit die staanspoor. <i>Then and there. On the spot.</i>
--

Figuur 3: Die mikrostruktuur van *Idiomewoordeboek*, geïllustreer deur middel van die lemma *wans*

4.5 Die toegangstruktuur

Die toegangstruktuur vir 'n beperkte woordeboek vir dekodering is soortgelyk aan dié vir 'n WSD vir dekodering.

5. Aanleerderwoordeboeke

Aanleerderwoordeboeke kan vir dekodering, produksie en vertaling gebruik word. Die funksie dekodering is dus slegs een van die funksies wat 'n invloed op die struktuur van aanleerderwoordeboeke het.

Tarp (2008: 137) noem twaalf kriteria wat gebruik kan word om die eien-skappe van 'n spesifieke tipe aanleerder te probeer identifiseer. Volgens hom is onder meer die aanleerder se algemene kulturele kennis asook hoeveel die aanleerder reeds weet van die betrokke kultuur van die taal wat hy/sy wil aanleer, van belang:

- i. Hoe wyd is die aanleerder se algemene kulturele kennis?
- ii. Hoe wyd is die aanleerder se kennis van die kultuur in die vreemde taal?

Moon (2016: 138) en Sánchez (2010: 111) argumenteer dat aanleerders sekere kulturele data benodig om byvoorbeeld die konnotasies van 'n woord korrek te verstaan.

5.1 Die raamstruktuur

As 'n draer van tekssoorte (Gouws 2017: 457) moet aanleerderwoordeboeke meer as slegs die sentrale teks bevat. Die buitetekste is veral belangrik vir die funksie dekodering, aangesien dit addisionele data wat van 'n algemene of kulturele aard is, kan bevat.

Heuberger (2016: 30) dui aan dat die lemmas en hul betekenisonderskeidings op 'n manier aangebied moet word wat sal verseker dat die aanleerder dit maklik kan verstaan en terselfdertyd hul woordeskat verbreed.

Die voortekste (titelblad, inhoudsopgawe, gebruikersleiding (wat die gebruiksliding insluit)), middeltekste en agtertekste vorm alles deel van die buitetekste.

Die gebruikersleiding van 'n aanleerderwoordeboek moet so saamgestel word dat dit relevante data vir die teikengebruiker bevat. Volgens Otto (1989: 406) omsluit die noodsaaklike data wat die gebruikersleiding van 'n eentalige aanleerderwoordeboek behoort te bevat, die volgende: die doel van die woordeboek, hoe dit saamgestel is, die omvang daarvan, die datatipes wat voorsien word, data oor die klanksisteem van die taal, 'n minigrammatika, hoe semantiese en gebruiksliding aangebied word, spellingleiding, lyste wat onreëlmatige vorme bevat, afkortings en SI-eenhede en hulle veelvoude.

5.2 Die dataverspreidingstruktuur

Die dataverspreidingstruktuur is soortgelyk aan dié van die omvattende en standaardwoordeboek, hoewel middel- en/of agtertekste meer benut word (Vergelyk Atkins en Rundell 2008: 177).

5.3 Die makrostruktuur

Die spesifieke makrostruktuurelemente wat opgeneem word, behoort op die teikengebruiker se behoeftes gebaseer te wees. Gouws (2017: 452) sê dat die woordeboeksoort en die naslaan- en woordeboekgebruiksvaardighede van die teikengebruikers 'n bepalende rol speel in die vaststelling van die tipe makrostruktuur. Die lemmas moet onder andere ten minste uit die gebruiklikste woordeskat van die betrokke taal geneem word, maar enige ander lemmas wat ook opgeneem word, behoort nie afgeskeep te word wat die bewerking daarvan betref nie (Rundell 2015: 304). Hoe meer leksikale items opgeneem kan word, hoe meer sal daar aan die behoeftes van die aanleerders voldoen word (Bogaards 1996: 281).

Die *Macmillan English Dictionary for Advanced Learners* (MEDAL) bevat nie slegs gebruiklike woorde wat in rooi aangedui is in die sentrale teks nie, maar verskaf ook ander lemmas buite die kernwoordeskat van die betrokke taal wat die gebruiker dalk nie verstaan nie. Heuberger (2016: 30) sê dat hierdie bykomende woorde veroorsaak dat die gebruikers hulle beter verstaan, wat terselfdertyd ook die gebruiker se woordeskat sal verbreed.

Die polifunksionaliteit van 'n aanleerderwoordeboek is belangrik ten opsigte van die ordeningstruktuur en moet daar, soos Chi (2013: 168) argumenteer, besluit word of 'n streng alfabetiese makrostruktuur met of sonder groepering gebruik gaan word.

5.4 Die mikrostruktuur

5.4.1 Aanduiders en struktuurmerkers

Tarp (2008: 77, 149) sê dat gebruikers lemmas, ortografie (variasie en foute), ongereelde fleksievorme, woordsoorte en pragmatiese en kulturele beperkings vir begrip benodig. Die volgende is uittreksels uit die artikel van *hoog* in *Basiswoordeboek van Afrikaans*:

hoog (hoë, hoër, die hoogste) *adj.* (...) Teenoor: **laag** (...)

Mense en diere is lank en nie hoog nie. Hoe laag iets ook al is wanneer dit gemeet word, sê 'n mens byvoorbeeld *Dit is 10cm hoog*, en nie *Dit is 10cm laag* nie.

Benewens betekenisverklarings en uitdrukkings, wat nie in hierdie uittreksel aangedui word nie, word ook die volgende data verskaf: die lemma, verboë vorm van die adjektief, trappe van vergelyking, woordsoortaanduiding, antoniem en 'n gebruikswaarskuwing in 'n teksblokkie.

Figuur 4: Aanduiders in *Basiswoordeboek van Afrikaans*

5.4.2 Uitspraakleiding

Uitspraakleiding is soms noodsaaklik vir begrip. In *Basiswoordeboek van Afrikaans*, byvoorbeeld, word die klem- en betekenisverskil by *volmaak* (vul) en *volmaak* (perfek) aangedui.

5.4.3 Etimologiese leiding

Daar kan ook etimologiese leiding in 'n aanleerderwoordeboek vir dekodering wees, wat aandui aan watter taal/tale 'n betrokke woord ontleen is (Otto 1989: 411) en hoekom 'n bepaalde woord uiteindelik 'n bepaalde betekenis gekry het. In hierdie verband verwys Ison (1983: 77) na die verduideliking van die oorsprong van **aboveboard** in die Webster's New (8th) Collegiate Dictionary: [Fr. the difficulty of cheating at cards when the hands are above the table], waar die laaste deel uitgebrei kan word om te lees [... the table (the 'board')].

5.4.4 Kollokasies

Kollokasies is wel soms relevant vir begrip, veral waar dit 'n idiomatiese of halfidiomatiese betekenis het, soos by **put up**, wat byvoorbeeld kan kollokeer met "a (great) fight, (little) resistance, a (spirited) defense" (vergelyk MEDAL: 1150). Kollokasies kan verder gebruik word om verskillende betekenisonderskeidings van óf basisse óf kollokatore van mekaar te onderskei, soos by byvoorbeeld **raad vra**, **ouers vra**, **aandag vra** (Otto 1989: 244). Die basis is hier **vra** en die betekenis verander by die kollokatore vanaf "versoek" na "ouers se goedkeuring vir 'n huwelik versoek" en uiteindelik "verg".

Kollokasies kan op meer as een plek verskyn. Otto (1989: 299) argumenteer dat in die geval van 'n bekende kollokasie, soos 'n **lied sing**, dit nie nodig sal wees om die kollokasie ook onder **sing** op te neem nie, terwyl 'n onbekende kollokasie, soos 'n **boom ontwortel**, onder **boom** en **ontwortel** opgeneem behoort te word. Vergelyk ook Dziemianko (2017: 341) vir probleme met opspoorbaarheid.

5.4.5 Etikette

Etikette moet in 'n aanleerderwoordeboek ingesluit word ter wille van dekode-ring ten einde betekenisonderskeidings van mekaar te onderskei, byvoorbeeld waar 'n woord sowel 'n algemene as vaktaalbetekenis het. In MEDAL word byvoorbeeld by die eerste lemma van **mimic** drie betekenisonderskeidings aan-gedui, waarvan die eerste twee algemene betekenisonderskeidings is, terwyl die derde betekenisonderskeiding geëtiketteer word as *science*.

5.4.6 Illustrasies

Die belangrike rol wat illustrasies in aanleerderwoordeboeke kan speel, word deur Otto (2003: 293-304) bespreek en daar word aangedui dat illustrasies en betekenisverklarings, waar nodig, saam gebruik moet word om te verseker dat belangrike taal- en kulturele aspekte op die duidelikste moontlike manier vir aanleerders verduidelik word.

5.5 Die toegangstruktuur

Die toegangstruktuur stem grootliks ooreen met dié van standaardwoorde-boeke.

6. Tweetalige woordeboeke

Tweetalige woordeboeke het gewoonlik meer as een funksie in gedrukte vorm omdat dit die gebruiker met begrip, produksie en vertaling kan help, en is dus meestal nie uitsluitlik 'n woordeboek vir dekodering nie. Volgens Fontenelle (2016: 45) help tweetalige woordeboeke gebruikers met sowel begrip van die moedertaal as die tweede taal.

6.1 Die raamstruktuur

Tweetalige woordeboeke vir dekodering sal 'n soortgelyke raamstruktuur as eentaliges toon.

6.2 Die dataverspreidingstruktuur

'n Tweetalige woordeboek vir dekodering se dataverspreidingstruktuur sal op so 'n manier saamgestel moet word dat die teikengebruiker die nodige inligting vir dekodering maklik kan vind, bv. afkortings en volvorme in 'n middelteks.

6.3 Die makrostruktuur

Gouws (2017: 446-447) stel as vereiste vir tweetalige woordeboeke dat die lem-makeuse vertaalekwivalente in die brontaal effektief moet verwerk en opneem, en dat die lemmas uit die standaardvariëteit van die brontaal geneem moet word. Verder moet hierdie lemmas maklik deur die gebruikers gekry en verstaan kan word, soos Kwary (2014: 67) aandui. Tarp (2008: 182) verduidelik dat 'n lem-makeuse wat op frekwensie gebaseer is, nie voldoende vir dekodering sal wees nie omdat gebruikers 'n breër woordeskat nodig sal hê.

6.4 Die mikrostruktuur

6.4.1 Aanduiders en struktuurmerkers

Tweetalige woordeboeke, juis omdat dit vertaalekwivalente aandui, benodig items wat daardie ekwivalente aandui, soos in die volgende voorbeeld uit *Longman-HAT English–Afrikaans / Afrikaans–Engels Basic Dictionary / Basiswoordeboek* (LHEABDBW) waar *kan* as die vertaalekwivalent van *might* aangegee word:

might modal verb

[USED TO SAY THAT SOMETHING IS POSSIBLE][MAG DALK] ► **kan** *Don't eat those berries — they might be poisonous. Moenie daardie bessies eet nie — hulle kan giftig wees. | You might have been killed! Jy kon dood gewees het!*

Figuur 5: Uittreksel wat die aanduiders en struktuurmerkers van die LHEABDBW illustreer

Aanduiders kan onder andere uitspraakleiding, grammatiese leiding, morfologie, vertaalekwivalente, etimologiese data of voorbeeldmateriaal insluit. In die voorafgaande uittreksel, byvoorbeeld, kan woordsoortaanduiding, betekenisvertaling, ekwivalente en voorbeeldsinne gesien word, wat relevant vir begrip is.

6.4.2 Ekwivalensie

Ekwivalensie is, volgens Adamska-Salaciak (2013: 222), die absolute verpligte element in enige tweetalige woordeboeke omdat dit lemmas in die brontaal voorsien van vertalings in die teikentaal. Adamska-Salaciak (2013: 227) sê verder dat ekwivalensie veral belangrik vir dekodering is.

Honselaar (2003: 324) noem dat tweetalige woordeboeke wat passief is (dit wil sê met die fokus op die funksie dekodering), die gebruikers juis toelaat om die teks te verstaan deur een of twee vertaalekwivalente op te neem en dan slegs die minimum bykomende data voorsien.

6.4.3 Kollokasies

Die volgende uittreksel uit die LHEABDBW insake *hand* dui aan hoe kollokasies, hier *by hand*, *met die hand* en *hulp verleen*, hanteer kan word:

hand noun (**hands**)

1 [PART OF YOUR BODY AT THE END OF YOUR ARM][DEEL AAN DIE PUNT VAN JOU ARM] ► **hand**

2 **by hand** [USING YOUR HANDS] [DEUR JOU HANDE TE GEBRUIK] ► **met die hand**
This shirt was made by hand. Hierdie hemp is met die hand gemaak.

3 **give someone a hand** [HELP SOMEONE] [HULP VERLEEN] ► **iemand help**

Figuur 6: Uittreksel van die kollokasies uit LHEABDBW

6.4.4 Uitspraakleiding

Gevalle waar uitspraakleiding wel tydens dekodering van belang is, is wanneer woorde dieselfde gespel word, maar verskillend uitgespreek word wat die klem betref, en hierdie klemverskil kan lei tot betekenisverskil, soos aangedui in *Pharos Afrikaans–Engels / English–Afrikaans Woordeboek / Dictionary* (PAEW) by *oordryf* wat onderskeidelik as *drift across/blow over* en as *exaggerate* (...) vertaal word.

6.4.5 Etimologiese leiding

Etimologiese leiding kan in 'n tweetalige woordeboek vir begrip opgeneem word, aangesien dit die gebruiker se kennis van 'n betrokke woord kan verbeter. In hierdie verband dui Ilson (1983) en Pierson (1989) die waarde van die studie van etimologiese data in die taalklaskamer aan. Leerders sal natuurlik die gevorderde vaardigheid om etimologiese data in woordeboeke te interpreteer, moet aanleer. Meer spesifiek dui Pierson (1989: 57) op grond van sy persoonlike ondervinding as 'n ESL (English Second Language)-onderwyser in Hong Kong aan hoe die studie van etimologie vir intermediaêre/gevorderde aanleerders benut kan word. Leerders wat weet dat *cranial* van die Griekse woord *kranion* met die betekenis "kopbeen" kom, kan makliker die betekenis *craniate*, *craniology*, *craniometry*, *craniotome*, ensovoorts binne konteks aflei.

6.4.6 Etiket

Etiket tree volgens Burkhanov (2003: 105) op as 'n aanduiding van die beperkings op die gebruik van leksikale items wat tyd, plek, situasies en kommunikatiewe omstandighede betref. In PAEW dui die etiket (*arch.*) voor **espousal** by die lemma **huwelik** aan dat die gebruiker nie hierdie vertaalekwivalent in 'n moderne konteks moet gebruik nie.

Atkins en Rundell (2008: 232) noem dat die opname van etikette ingewikkelder in 'n tweetalige woordeboek is as in 'n eentalige woordeboek omdat die ekwivalente in T1 en T2 gemerk moet word.

6.4.7 Voorbeeldmateriaal

Vrbinc en Vrbinc (2016: 298) dui aan dat leksikograwe verskillende faktore in ag neem wanneer besluite oor die opname van voorbeeldmateriaal gemaak word. Die getal voorbeelde wat die leksikograaf benut om die gebruik van die leksikale item of 'n bepaalde betekenisonderskeiding van 'n leksikale item te illustreer, kan nie vooraf bepaal word nie, maar sekere riglyne is tog ter sake:

- die semantiese en grammatiese kompleksiteit van die betrokke leksikale item — hoe kompleks, hoe meer voorbeelde (Toope 1996: 167)
- die mate waarin twee kulture van mekaar verskil (Jacobsen et al. 1991: 2788)
- die funksie van die woordeboek — enkodering vereis meer voorbeelde as dekodering (Kromann et al. 1991: 2772).

Vrbinc en Vrbinc argumenteer dat voorbeelde in tweetalige woordeboeke eintlik albei funksies vervul, aangesien daar gewoonlik nie vier stelle tweetalige woordeboeke is om voorsiening te maak vir twee stelle moedertaalsprekers nie (Jacobsen et al. 1991: 2786). In terme van woordeboeke vir dekodering behoort 'n tweetalige woordeboek voorbeeldmateriaal te bevat wat die gebruikers help om 'n vreemde taal te begryp of 'n teks van 'n vreemde taal na die moedertaal te vertaal.

6.5 Die toegangstruktuur

Die eksterne toegangstruktuur se soekroete in 'n tweetalige woordeboek vir dekodering is soortgelyk aan dié van eentalige woordeboeke.

Fuertes-Olivera en Niño-Amo (2011: 172-173) verduidelik dat een van die belangrikste elemente in 'n elektroniese woordeboek 'n webtuiste met 'n eenvoudige adres is, en dat die soekfunksie maklik gebruik moet kan word.

7. Woordeboeke vir spesiale doeleindes (WSD)

WSD's kan een of meer funksies hê, afhangende van die teikengebruiker en die vakgebied waarop die betrokke woordeboek gemik is.

7.1 Die raamstruktuur

Die sentrale woordelys moet 'n effektiewe ordeningstruktuur hê, hetsy alfabe-

ties (met of sonder groepering) of sistematies (Svensén 2009: 21). 'n Alfabetiese ordeningstruktuur kom gewoonlik meer dikwels voor, maar Svensén noem dat 'n sistematiese ordeningstruktuur soms beter vir 'n WSD kan wees omdat die lemmas volgens 'n tema/vakgebied gelys word. Bowker (2003: 158) dui aan dat indien 'n WSD van 'n sistematiese ordeningstruktuur gebruik maak, wat die leksikale items volgens vakgebied en/of die verhoudings wat hulle met mekaar toon lys, kan die items wel alfabeties in die indeks gelys word. Hierdie indeks is dan deel van die buitetekste en kom gewoonlik in die agterteks voor, soos in die onderstaande uittreksel uit *A Thesaurus of Medical Words and Phrases* uitgebeeld:

```
w

WAHOO. Euonymus.
WAIST. Cingulum.
WAKEFULNESS. See also Sleep.

Synonyms : Insomnia, Pervigilium, Vigilia, Egrogorsis, Ahypnia, Agrypnia.
agent to produce w. Agrypnotic ; Egertic.
intense w. Egorsis ; Ahypnosis.

form of coma with wakefulness and delirium. Coma vigil; Agrypnocoma.
WAKING. See Wakefulness.
```

Figuur 7: Die indeks in die agterteks van *A Thesaurus of Medical Words and Phrases*

Dit is dus eintlik die teikengebruikers en gebruiksituasies wat die ordeningstruktuur van die betrokke WSD bepaal.

7.2 Die dataverspreidingstruktuur

WSD's is beperk ten opsigte van hulle datatipes, gevolglik moet die leksikograaf data wat vir dekodering relevant is, in die WSD se sentrale teks plaas, terwyl enige bykomende data in buitetekste kan verskyn.

7.3 Die makrostruktuur

WSD's kan alfabeties of sistematies georganiseer word. Die kriteria vir die opname en bewerking van lemmas in 'n WSD vir begrip is soos volg, volgens Bowker (2003: 157-158) en Svensén (2009: 377-378): die lemmas moet slegs opgeneem word as dit relevant vir die vakgebied en teksbegrip is, en die gebruiker moet die data wat vir dekodering opgeneem word maklik kan op-

spoor, wat beteken dat die leksikograaf 'n alfabetiese indeks moet verskaf as die WSD nie alfabeties georden is nie.

7.4 Die mikrostruktuur

Bowker (2003: 159) verduidelik dat hoe meer gespesialiseerd 'n WSD is, hoe minder sal die datatipes wees (soos in die geval van gevorderde gebruikers wat reeds met die betrokke vakgebied en woordeboek tipe vertrou is). Betekenisverklaring is wel noodsaaklik.

In hierdie verband behoort betekenisverklaring eiesoortig in verskillende woordeboek tipes verklaar te word. Die woord *haarskeerder* kan byvoorbeeld deur verskillende gebruikers teëgekomp word. Indien 'n aanleerder hierdie woord in 'n aanleerderwoordeboek sou naslaan en die betekenisverklaring "groot, harige spinnekop" sou kry, sal dit verstaanbaar wees, maar tegnies verkeerd, aangesien dit 'n unieke insek is en nie 'n spinnekop nie. 'n Moedertaalspreker wat die woord in 'n teks teëkom en ook uit die konteks kan raai dat dit 'n tipe insek is, maar wil weet watter soort insek dit is, kan eweneens mislei word deur die eerste kort betekenisverklaring in byvoorbeeld VAW: 1 Groot, harige en vinnige soort spinnekop, jagspinnekop (*Solifuga*). 'n Leek wat besig is om 'n boek te lees en net vinnig die betekenis in 'n WSD wil verifieer, sal waarskynlik nie die moeite doen om *Solifuga* in 'n ander bron te gaan naslaan nie. Buitendien is dit nie gebruikersvriendelik om nog 'n soektog te moet uitvoer as jy net vinnig 'n betekenis wil verifieer nie. In laasgenoemde geval is slegs 'n kort, korrekte verklaring nodig wat aandui dat dit 'n unieke insek is wat lyk soos 'n kruising tussen 'n harige spinnekop en 'n skerpioen as gevolg van groot knypers. Vir 'n semi-deskundige wat 'n teks lees en die woord in 'n WSD naslaan om die betekenis binne die spesifieke konteks te verstaan, is 'n uitvoeriger verklaring nodig, aangesien hierdie persoon oor gespesialiseerde kennis beskik en in fyner besonderhede sal belangstel. Hierdie tipe gebruiker sal waarskynlik wil weet tot watter biologiese familie die insek behoort, of dit giftig is of nie, ensovoorts. Die wetenskaplike naam en ander besonderhede is dus belangrik, byvoorbeeld dat dit niegiftig, baie vinnig is, nie teen glase op kan beweeg nie en baie seer kan byt as dit bedreig word. Vir die ekspert wat 'n WSD raadpleeg, sal feitlik net 'n etiket en die wetenskaplike naam voldoende wees. Vir die leek en semi-deskundige sal 'n etiket in 'n algemene WSD nodig wees om dié term van die algemene woord *haarskeerder* te onderskei.

7.4.1 Aanduiders en struktuurmerkers

Die funksie dekodering het 'n invloed op beide die aanduiders en struktuurmerkers in WSD's. WSD's bevat beperkte datatipes, daarom sal struktuurmerkers die gebruiker help om vinnig by data vir teksbegrip uit te kom.

7.4.2 Uitspraakleiding

Uitspraakleiding is 'n belangrike element van 'n mikrostruktuur, maar omdat WSD's beperkte data bevat, sal hierdie datatipe moontlik nie in alle WSD's voorsien word nie. In die geval van die *Farlex Partner Medical Dictionary for Health Consumers*, word daar wel uitspraakleiding verskaf, soos in die volgende geval:

The screenshot shows the Farlex Partner Medical Dictionary for Health Consumers website. At the top, there is a search bar with the word 'tibia' entered. Below the search bar, there are several navigation options: 'Word / Article', 'Starts with', 'Ends with', and 'Text'. Below these options, there is a horizontal menu with various categories: 'Dictionary', 'Thesaurus', 'Medical Dictionary', 'Legal Dictionary', 'Financial Dictionary', 'Acronyms', 'Idioms', and 'Encyclopedia'. Below the menu, there is a banner for 'ARC View: AspenTech Aims to Optimize Asset Performance in Industrial Process Plants'. Below the banner, there is a section for 'tibia' with a pronunciation guide and a definition. The definition states: 'the inner and larger of the two bones of the lower leg; it articulates with the FEMUR and head of the FIBULA above and with the TALUS below. See appendix 3-3 and see color plates.'

Figuur 8: Die uitspraakgids in *Farlex Partner Medical Dictionary for Health Consumers*

7.4.3 Etiket

Indien die WSD meer as een vakgebied behandel en dus nie besonder gespesialiseer is nie, is die opname van etikette uiters belangrik (Bowker 2003: 159).

7.4.4 Die toegangstruktuur

Die eksterne toegangstruktuur se soekroete vir dekodering in 'n WSD sal die omslag, die gebruikersgids en loopkoppe insluit, en intern sal 'n indeks nuttig wees.

8. Elektroniese woordeboeke

Dieselfde aanduidertipes (en meer) kan in elektroniese woordeboeke se dekoderingsoptie voorkom, maar die strukture verskil. Die toegangstruktuur is

anders weens 'n veelvlakkige artikelstruktuur en gewysigde adresseringstruktuur (Gouws 2018: 182). Die hooftegangstruktuur val ook nie noodwendig saam met die makrostruktuur nie (Gouws 2018:182). Daar is nie 'n raamstruktuur nie, wel 'n uitgebreide dataverspreidingstruktuur, en selfs 'n woordeboekportaalstruktuur en datatrekkingstruktuur (Gouws 2018: 192) is moontlik.

9. Samevatting

Die funksie dekodering het 'n invloed op die woordeboekstrukture en data-inhoud van woordeboeke. Die strukture in gedrukte woordeboeke verskil van dié in elektroniese woordeboeke, soos in die voorafgaande paragraaf aangedui. In ware elektroniese woordeboeke kan woordeboekgebruikers op die dekodeeringsopsie klik, terwyl die gebruiker van 'n gedrukte woordeboek naas die betekenisverklaring ook deur middel van etikette, woordsoortaanduidings, kollokasies en soms uitspraakleiding, etimologiese leiding en voorbeeldmateriaal tot beter begrip kom. In hierdie artikel is veral op gedrukte woordeboeke gefokus. Verdere navorsing behoort op die kriteria vir dekodering in elektroniese woordeboeke te fokus. Die sistematiese beskrywing van die funksies produksie en vertaling word ook in die vooruitsig gestel.

Bronnelys

A. Elektroniese woordeboeke

Farlex Partner Medical Dictionary for Health Consumers 2007: <https://medical-dictionary.thefreedictionary.com/tibia> [2 April 2019].

VivA. Virtuele Instituut vir Afrikaans: www.viva-afrikaans.org.

B. Gedrukte woordeboeke

Barton, W.M. 1903. *A Thesaurus of Medical Words and Phrases*. Philadelphia/New York/Londen: W.B. Saunders & Company.

Botha, W.F. (Hoofred.). 2005. *Woordeboek van die Afrikaanse Taal, Deel XII*. Stellenbosch: Buro van die WAT.

De Villiers, M. en R.H. Gouws. 1988. *Idiomewoordeboek*. Kaapstad: Nasou.

Du Plessis, M. 2005. *Pharos Afrikaans–Engels / English–Afrikaans Woordeboek / Dictionary*. Kaapstad: Pharos.

Gouws, R., I. Feinauer en F. Poneelis. 1994. *Basiswoordeboek van Afrikaans*. Pretoria: Van Schaik.

Labuschagne, F.J. en L.C. Eksteen. 2010. *Verklarende Afrikaanse Woordeboek*. Negende, hersiene uitgawe. Kaapstad: Pharos.

Luther, J. (Red.). 2011. *Longman-HAT English–Afrikaans / Afrikaans–Engels Basic Dictionary / Basiswoordeboek*. Pietermaritzburg: Pearson.

- Luther, J., F. Pfeiffer en R.H. Gouws (Reds.).** 2015⁶. *HAT Verklarende Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson.
- Morris, W.** 1985. *The American Heritage Dictionary*. Boston: Houghton Mifflin.
- Rundell, M. en G. Fox (Hoofreds.).** 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.
- Van Schalkwyk, D.J. (Hoofred.).** 1996. *Woordeboek van die Afrikaanse Taal, Deel X*. Stellenbosch: Buro van die WAT.
- Van Schalkwyk, D.J. (Hoofred.).** 1996. *Woordeboek van die Afrikaanse Taal, Deel XI*. Stellenbosch: Buro van die WAT.

C. Ander bronne

- Adamska-Sałaciak, A.** 2013. Issues in Compiling Bilingual Dictionaries. Jackson, H. (Red.). 2013: 213-231.
- Atkins, B.T.S. en M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Ball, L.** 2016. *An Evaluative Study to Determine to What Extent Technology Can be Used in e-Dictionaries to Provide Relevant Information on Demand*. Ongepubliseerde M.A.-verhandeling. Pretoria: Universiteit Pretoria.
- Béjoint, H.** 2016. Dictionaries for General Users: History and Development; Current Issues. Durkin, P. (Red.). 2016: 7-24.
- Bentivogli, L. en E. Pianta.** 2002. Detecting Hidden Multiwords in Bilingual Dictionaries. Braasch, A. en C. Povlsen (Reds.). 2002: 785-793.
- Bergenholtz, H. en R.H. Gouws.** 2013. The Presentation of Word Formation in General Monolingual Dictionaries. *Lexikos* 23: 59-76.
- Bergenholtz, H. en S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Blume, J.** 2017. 'n Ondersoek na die kriteria vir die realisering van die woordeboekfunksies begrip, produksie en vertaling. Ongepubliseerde M.A.-verhandeling. Port Elizabeth: Nelson Mandela Universiteit.
- Bogaards, P.** 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.
- Bogaards, P.** 2003. Uses and Users of Dictionaries. Van Sterkenburg, P. (Red.). 2003: 26-33.
- Bogaards, P.** 2013. A History of Research in Lexicography. Jackson, H. (Red.). 2013: 19-32.
- Botha, W. (Red.).** 2003. 'n Man wat beur: Huldigingsbundel vir Dirk van Schalkwyk. Stellenbosch: Buro van die WAT.
- Botha, W., P. Mavoungou en D. Nkomo (Reds.).** 2013. *Festschrift Rufus H. Gouws*. Stellenbosch: SUN PReSS.
- Bothma, T.J.D.** 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, P.A. en H. Bergenholtz (Reds.). 2011: 71-102.
- Bowker, L.** 2003. Specialized Lexicography and Specialized Dictionaries. Van Sterkenburg, P. (Red.). 2003: 154-164.
- Braasch, A. en S. Olsen.** 2000. Formalised Representation of Collocations in a Danish Computational Lexicon. Heid, U. et al. 2000: 475-488.
- Braasch, A. en C. Povlsen (Reds.).** 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002*. Copenhagen: Center for Sprogteknologi.

- Brinker, K. et al. (Reds.).** 2000. *Text und Gesprächslinguistik*. HSK 16(1). Berlyn/New York: De Gruyter.
- Burkhanov, I.** 2003. Pragmatic Specifications: Usage Indications, Labels, Examples; Dictionaries of Style, Dictionaries of Collocations. Van Sterkenburg, P. (Red.). 2003: 102-113.
- Carstens, W.A.M. en N. Bosman (Reds.).** 2017. *Kontemporêre Afrikaanse Taalkunde*. Tweede uitgawe. Pretoria: Van Schaik.
- Chi, A.** 2013. Researching Pedagogical Lexicography. Jackson, J. (Red.). 2013: 165-187.
- Cowie, A. (Red.).** 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985*. Tübingen: Max Niemeyer.
- Cowie, A. (Red.).** 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.
- De Caluwé, J. en A. van Santen.** 2003. Phonological, Morphological and Syntactic Specifications in Monolingual Dictionaries. Van Sterkenburg, P. (Red.). 2003: 71-82.
- Du Plessis, A.H.** 2017. Die rol van die bruikbaarheidsbenadering binne die e-leksikografie. *Literator* 38(2). <https://doi.org/10.4102/lit.v38i2.1340>.
- Durkin, P. (Red.).** 2016. *The Oxford Handbook of Lexicography*. New York: Oxford University Press.
- Dziemianko, A.** 2017. Dictionary Form in Decoding, Encoding and Retention: Further Insights. *ReCALL* 29(3): 335-356.
- Fontenelle, T.** 2016. Bilingual Dictionaries: History and Development; Current Issues. Durkin, P. (Red.). 2016: 44-61.
- Fuertes-Olivera, P.A. (Red.).** 2010. *Specialised Dictionaries for Learners*. Berlyn/New York: De Gruyter.
- Fuertes-Olivera, P.A.** 2013. E-lexicography: The Continuing Challenge of Applying New Technology to Dictionary-Making. Jackson, H. (Red.). 2013: 323-340.
- Fuertes-Olivera, P.A. en H. Bergenholtz (Reds.).** 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Londen/New York: Continuum.
- Fuertes-Olivera, P.A. en M. Niño-Amo.** 2011. Internet Dictionaries for Communicative and Cognitive Functions: *El Diccionario Inglés-Español de Contabilidad*. Fuertes-Olivera, P.A. en H. Bergenholtz (Reds.). 2011: 168-186.
- Gouws, R.H.** 1988. Die gebruik van etikette as leksikografiese hulpmiddel. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Geleentheidsuitgawe 6, April 1988: 1-56.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 2003. Types of Articles, their Structure and Different Types of Lemmata. Van Sterkenburg, P. (Red.). 2003: 34-43.
- Gouws, R.H.** 2017. Leksikografie. Carstens, W.A.M. en N. Bosman (Reds.). 2017: 429-461.
- Gouws, R.H.** 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195. <http://Lexikos.journals.ac.za>; <https://doi.org/10.5788/28-1-1461>.
- Hausmann, F.J. et al. (Reds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexicographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*. Berlyn: De Gruyter.
- Heid, U. et al. (Reds.).** 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Heuberger, R.** 2016. Learners' Dictionaries: History and Development; Current Issues. Durkin, P. (Red.). 2016: 25-43.
- Honselaar, W.** 2003. Examples of Design and Production Criteria for Bilingual Dictionaries. Van Sterkenburg, P. (Red.). 2003: 323-332.

- Howarth, P.** 1998. The Phraseology of Learners' Academic Writing. Cowie, A.P. (Red.). 1998: 161-188.
- Ilsion, R.** 1983. Etymological Information: Can it Help our Students? *ELT Journal* 37(1): 76-82.
- Ilsion, R.** 1987. Illustrations in Dictionaries. Cowie, A. (Red.). 1987: 193-211.
- Jackson, H. (Red.).** 2013. *The Bloomsbury Companion to Lexicography*. Londen/Nieu-Delhi/New York/Sydney: Bloomsbury Academic.
- Jacobsen, J.R., J. Manley en V.H. Pedersen.** 1991. Examples in the Bilingual Dictionary. Hausmann, F.J. et al. (Reds.). 1991: 2782-2789.
- Kemmer, K.** 2010. Onlinewörterbücher in der Wörterbuchkritik: Ein Evaluationsraster mit 39 Beurteilungskriterien. *Online publizierte Arbeiten zur Linguistik (OPAL) 2*. Mannheim: Institut für Deutsche Sprache.
- Klosa, A. (Red.).** 2008. Lexikographische Portale im Internet. *Online publizierte Arbeiten zur Linguistik (OPAL) 1*. Mannheim: Institut für Deutsche Sprache.
- Klosa, A.** 2016. Illustrations in Dictionaries; Encyclopaedic and Cultural Information in Dictionaries. Durkin, P. (Red.). 2016: 515-531.
- Klosa, A., L. Lemnitzer en G. Neumann.** 2008. Wörterbuchportale — Fragen der Benutzerführung. Klosa, A. (Red.). 2008: 5-35.
- Kromann, H.-P., T. Riiber en P. Rosbach.** 1991. Grammatical Constructions in the Bilingual Dictionary. Hausmann, F.J. et al. (Reds.). 1991: 2770-2775.
- Kwary, D.A.** 2014. The Evolution of Dictionaries of Economics: From a Glossary to a Lexicographical Information System. *Hermes: Journal of Language and Communication in Business* 27(52): 59-73.
- L'homme, M.-C. en C. Bertrand.** 2000. Specialized Lexical Combinations: Should they Be Described as Collocations or in Terms of Selectional Restrictions? Heid, U. et al. 2000: 497-506.
- Moon, R.** 2016. Explaining Meaning in Learners' Dictionaries. Durkin, P. (Red.). 2016: 123-143.
- Oppentocht, L. en R. Schutz.** 2003. Developments in Electronic Dictionary Design. Van Sterkenburg, P. (Red.). 2003: 215-227.
- Otto, A.N.** 1989. *Kriteria vir 'n Afrikaanse aanleerderwoordeboek*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.
- Otto, A.N.** 2003. Woordeboekillustrasies, met spesiale verwysing na illustrasies in aanleerderwoordeboeke. Botha, W. (Red.). 2003: 293-304.
- Otto, A.N.** 2013. 'n Kritiese waardering van die hantering van kollokasies en idioome in die *HAT* (2005). Botha, W. et al. (Reds.). 2013: 188-195.
- Pierson, Herbert D.** 1989. Using Etymology in the Classroom. *ELT Journal* 43(1): 57-63.
- Ptaszynski, M.O.** 2009. Function-related Secondary User Needs and Secondary Data? A Critical Examination of Some Central Concepts in the Modern Theory of Lexicographical Functions. *Lexikos* 19: 242-261.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Sager, S.F.** 2000. Hypertext und Hypermedia. Brinker, K. et al. (Reds.). 2000: 587-603.
- Sánchez, A.** 2010. The Treatment of Cultural and/or Encyclopaedic Items in Specialised Dictionaries for Learners. Fuertes-Olivera, P.A. (Red.). 2010: 107-129.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Swanepoel, P.** 2003. Dictionary Typologies: A Pragmatic Approach. Van Sterkenburg, P. (Red.). 2003: 44-70.

- 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.
- Toope, M.** 1996. *Examples in the Bilingual Dictionary*. Ongepubliseerde M.A.-verhandeling. Ottawa: Universiteit van Ottawa. [Aanlyn]. Beskikbaar by: <http://www.ruor.uottawa.ca/bitstream/10393/9942/1/mm15769.PDF>.
- Van Sterkenburg, P. (Red.).** 2003. *A Practical Guide to Lexicography. Volume 6*. Amsterdam/Philadelphia: John Benjamins.
- Vrbinc, A. en M. Vrbinc.** 2016. Illustrative Examples in a Bilingual Decoding Dictionary: An (Un)necessary Component? *Lexikos* 26: 296-310.

Exploring the Properties of English Lexical Affixes by Exploiting the Resources of English General-Purpose Dictionaries

Tvrtko Prčić, *Department of English, Faculty of Philosophy, University of Novi Sad, Serbia (tvrtko.prcic@gmail.com)*

Abstract: This paper proposes a new model for exploring the properties of English lexical affixes, based on exploiting the resources of English general-purpose dictionaries. Developed primarily for EFL university students and motivated by the highly inconsistent treatment of affixes in those dictionaries, this model builds around a heuristic self-study method and its accompanying bare-bones inferential dictionary (BBID), purposefully designed, produced and distributed as a guide to students' discoveries. The model has been devised so as to direct students away from affix entries in general-purpose dictionaries towards word entries which contain specific affixes and to lead students to discover the properties of target affixes by analysing corresponding source words provided in BBID. The exposition is divided into four parts, as follows: Section 1 brings some introductory, scene-setting remarks; in Section 2, essential and relational affixal properties are presented; in Section 3, the five major aspects of the new model are explained and exemplified: pragmatic reasons, main objectives, underlying principles, instructional material and real-life functioning; finally, Section 4 offers a summary and a critical assessment of the model and its BBID, together with a glimpse into their future.

Keywords: LEXICAL AFFIXES, ESSENTIAL PROPERTIES, RELATIONAL PROPERTIES, GENERAL-PURPOSE DICTIONARIES, BARE-BONES INFERENTIAL DICTIONARY, HEURISTIC, SELF-STUDY, ENGLISH

Opsomming: Die verkenning van die eienskappe van Engelse leksikale affikse deur die benutting van die hulpbronne in Engelse woordeboeke vir algemene doeleindes. In hierdie artikel word 'n nuwe model, gegrond op die ontginning van Engelse woordeboeke vir algemene doeleindes as hulpbronne, vir die verkenning van die eienskappe van Engelse leksikale affikse voorgestel. Hoofsaaklik ontwikkel vir EVT-universiteitstudente en gemotiveer deur die hoogs inkonsekwente hantering van affikse in hierdie woordeboeke, word die model saamgestel rondom 'n heuristiese selfstudiemetode en die gepaardgaande basiese deduktiewe woordeboek (Engels BBID), wat doelmatig ontwerp, geproduseer en versprei word as 'n gids vir studenteverkenning. Die model is ontwerp om studente weg te stuur van affiksinskrywings in woordeboeke vir algemene doeleindes na woordinskrywings wat spesifieke affikse bevat en om studente te lei na die ontdekking van die eienskappe van doelaffikse deur ooreenstemmende bronwoorde wat in die BBID verskaf word, te analiseer. Die uiteensetting word verdeel in vier dele: Afdeling 1 bevat inleidende, beskrywende opmerkings; in afdeling 2 word noodsaaklike en

relasionele eienskappe van affikse weergegee; in afdeling 3 word die vyf hoofaspekte van die nuwe model uiteengesit en toegelig: pragmatiese motiewe, hoofdoelwitte, onderliggende beginsels, onderrigmateriaal en werklike funksionering; laastens bied afdeling 4 'n opsomming en 'n kritiese beoordeling van die model en sy BBID aan, met terselfdertyd 'n kykie na die toekoms.

Sleutelwoorde: LEKSIKALE AFFIKSE, NOODSAAKLIKE EIENSKAPPE, RELASIONELE EIENSKAPPE, WOORDEBOEKE VIR ALGEMENE DOELEINDES, BASIESE DEDUKTIEWE WOORDEBOEK, HEURISTIES, SELFSTUDIE, ENGELS

1. Opening remarks

In this paper, a solution to the problem of acquiring the properties of English lexical affixes, especially by EFL university students, will be put forward, with the intention of ensuring that all students possess roughly the same working knowledge sufficient for interpreting and understanding both established and new prefixations and suffixations, as they crop up in real-world situations. It will soon be shown that English general-purpose dictionaries, learner's and native-speaker ones alike, for the most part, fail to provide a satisfactory coverage of affixal properties, because their provisions are often inconsistent and/or incomplete, sometimes inaccurate or even lacking altogether. As a result, resources of this kind cannot be recommended for educational use. What is needed in the circumstances is the formulation and adoption of a coherent strategy for a theoretically and methodologically well-founded and well-balanced lexicographic treatment of the properties of lexical affixes, to be implemented in future monolingual general-purpose dictionaries and, better still, in specialized dictionaries of lexical affixes, compiled from scratch.

The said strategy would be based on two key principles: firstly, from a theoretical point of view, the strategy would consider the system of affixes as a network whose elements, i.e. prefixes and suffixes, have their own unique identities and are, concurrently, interconnected with other elements by their comparable forms and/or functions and/or contents (details to be spelled out below); and secondly, from a methodological point of view, the strategy would establish a general — ideally a standardized or at least a widely accepted — lexicographic procedure for representing this unique identity-cum-interconnectedness dichotomy in an integral, effective and user-friendly manner (details yet to be set out). If achieving a common and strong enough consensus among practical and theoretical lexicographers and word-formationists, this innovative approach, or its modified or amended version, could, in the short run, directly contribute to a more consistent, complete and accurate treatment of affixes in general-purpose dictionaries. In the long run, the approach could lead to the compilation of new specialized dictionaries of lexical affixes, or, preferably, a modular electronic dictionary, offering coverage of variable, user-selectable, depth and breadth, to cater for various target groups of users, but paying particular attention to addressing the communicative and educational

needs of advanced learners world-wide — specifically, EFL university students.

With these preconditions in mind, the solution, in the form of a new model, about to be worked out here, is meant to bridge the divide between the present and the future state of affairs, and that is why it should be viewed as an interim solution — until a lasting and optimal solution to the problem is agreed on.

The rest of the exposition will be organized into three parts, thus: English lexical affixes and their properties will be defined and illustrated in Section 2; in Section 3, focusing on its five central aspects, a detailed account of the new model of acquiring affixal properties will be given; and in Section 4, a summary with an appraisal of the achievements of this model, alongside an anticipation of possible future dictionaries of English lexical affixes will round off the discussion.

2. Properties of English lexical affixes: theoretical considerations

By 'English lexical affixes' in this paper are meant derivational prefixes and suffixes in present-day English, synchronically separable sequences of phonemes and fully-fledged morphemes, in which a particular phonological form is systematically associated with at least one particular content and one particular function, irrespective of the affix's origin (cf. Prčić 1999, 2005, 2008, 2019; for alternative views, see Bauer 1983; Bauer and Huddleston 2002; Bauer et al. 2013; Dixon 2014; Lieber and Štekauer 2014; Miller 2014; Štekauer and Lieber 2005); for example, *re-* is a prefix in *replay*, but not in *repeat*, and *-ish* is a suffix in *Finnish*, but not in *finish*, because the latter members of both pairs of these examples are synchronically monomorphemic, unanalysable, words. This triadic conception of the morpheme and, by extension, of the affix is in keeping with the sign-oriented approach to word formation, developed by Marchand (1969), and, ultimately, with the principles of Saussurean structuralism (De Saussure 1916).

Before turning to identifying, describing and exemplifying the properties of English lexical affixes, two vitally important things need to be pointed out (cf. Prčić 1999, 2019). Firstly, affixes have a contributory role in word formation and their properties should be construed mostly in terms of what they contribute graphologically, phonologically, morphosyntactically, semantically and stylistically to properties of the base, in deriving (new) morphologically complex words — either prefixations or suffixations; for example, the actual contribution of the prefix *re-* to the base *play* in the prefixation *replay* or of the suffix *-ish* to the base *Finn* in the suffixation *Finnish*.

Secondly, affixal properties constitute a systematic and predictable contribution in form, content and function to any eligible base, with the same set of modifications involved in all words derived from a certain class of bases. That is why contributions of affixes are considered word-independent — as long as those words are compositional formations, i.e. morphologically and semantically analysable, and therefore predictable in form and content. For example,

the contribution of the prefix *re-*, in the sense 'again', to bases in words like *replay*, *readjust*, *reappear*, *reelect*, *reread*, *revisit* is always the same. However, in formations going through, or having gone through, the process of idiomatization, or lexicalization, gradual changes in form and/or content may in time take place, or have already taken place, within words as wholes. Such modifications, affecting original properties of affixes and bases, are non-systematic and unpredictable, and are considered word-dependent, because they have developed uniquely within particular words. For example, in the semantically idiomatized word *rewrite*, which means 'to write again in a better way', the prefix *re-* still means 'again', but the word itself conveys more meaning than is available on the surface, since it carries the added component 'in a better way'. Due to the working of idiomatization, the morphological and semantic analysability of words thus altered progressively decreases, their compositionality blurs and their predictability lessens (for extensive accounts of idiomatization, see Bauer 1983; Bauer and Huddleston 2002; Hohenhaus 2005; Lieber and Štekauer 2014; Lipka 2002; Prčić 2001, 2016; Quirk et al. 1985). In consequence, any attempt at compiling an inventory of properties of English lexical affixes should concentrate on their contributory role and their systematic contribution to the base, and leave the effects of idiomatization to the treatment of particular idiomatizations.

Affixal properties — or, distinctive properties, in full — define an affix's identity, its place in the lexicon and its relationship with other comparable affixes in that lexicon. In order to be able to put together a comprehensive picture of affixal and interaffixal properties, a two-tier descriptive model composed of two sets will be employed (cf. Prčić 2019). The first consists of essential properties that capture individual form- and content-based behavioural traits of affixes at all levels of linguistic analysis. The second, supplementary, set consists of relational properties, that capture paradigmatic and syntagmatic relations established, or establishable, between individual affixes. Both sets of properties will now be elaborated, with definitions and examples drawing on those in the original proposition (Prčić 2019).

2.1 Essential properties of English lexical affixes

Relevant to all affixes, i.e. prefixes and suffixes, and reflecting affixal individuality, essential properties comprise the identifying characteristics of an affix in isolation and its contribution to a base or a class of bases. According to the descriptive model adopted here, there are six essential properties, all governed by various levels of linguistic analysis:

(1) **Spelling.** In the domain of graphology, this property concerns the realization of an affix in writing. Graphologically, the affix contributes to bases the letter sequence that represents it in isolation, like *-ship*, and within words, like *friendship*. Additionally, where applicable, there are **Spelling Niceties**, referring

to, firstly, variants of the letter sequence, as in *-ize* vs *-ise*; secondly, orthographic changes to the base, including dropping of a silent final *-e* before a suffix beginning with a vowel, as in *drive[e]* + *-er* > *driver*, or vacillation between dropping and non-dropping of a silent final *-e* before a suffix beginning with a vowel, as in *love[e]* + *-able* > *lovable* vs *love* + *-able* > *loveable*; and thirdly, indication of prefix/base hyphenation to distinguish an unanalysable homograph from an analysable one, like *recover* vs *re-cover*, or to signify separate articulations of two identical adjacent vowel letters, like *co-occur* or *re-elect*.

(2) **Pronunciation.** In the domain of phonology, this property concerns the realization of an affix in speech, provided in standard IPA script. Phonologically, the affix contributes to bases the sound sequence that represents it in isolation, like /ʃɪp/, and within words, like /'frendʃɪp/. Additionally, relevant to suffixes, there is **Stress Placement**, referring to the influence of the affix on the stress pattern of the word (cf. Jones 2006; Wells 2008), which may either remain unchanged, with stress-neutral suffixes, as in *-ship*: /⁰frend > 'frendʃɪp/, or it may shift, with stress-imposing suffixes, as in *-ese*: /dʒə'pæn > ,dʒæpə'ni:z/, when the primary stress falls on the suffix itself, or *-ic*: /'ætəm > ə'tɒmɪk/, when it is placed on the penult, or *-ity*: /kri'eɪtɪv > ,kri:ɪr'tɪvɪtɪ/, when it is on the antepenult.

(3) **Attachability.** In the domains of morphosyntax and semantics, this property concerns the ability of an affix to attach to a base, or a class of bases, of a certain part of speech and to derive words of a certain part of speech (cf. Bauer 1983; Bauer and Huddleston 2002; Quirk et al. 1985). Morphosyntactically, the affix contributes to bases the part of speech of the derived word, which may be either the same as that of the base, with class-maintaining affixes, like *dis-* + *connect_v* > *disconnect_v*, or different from that of the base, with class-changing affixes, like *develop_v* + *-ment* > *development_n*. Additionally, where applicable, there is **Subcategorization**, referring to a finer specification of the base, mostly semantically and/or pragmatically determined, like transitivity of verbal bases to which the suffix *-able*, in the sense 'that can be', typically attaches, as in *wash_{v/tr}* + *-able*.

(4) **Descriptive Meaning.** In the domain of semantics, this property concerns the basic, denotative, or systemic, meaning of an affix (cf. Cruse 2010; Lipka 2002; Prčić 2016). Semantically, the affix contributes to bases a set of diagnostic features, spelled out within a brief definition, well-balanced between generality and specificity, like the prefix *re-*, meaning 'again, repetition', as in *reread*, or the suffix *-ish*, meaning 'around, approximately', as in *sixish*.

(5) **Associative Meaning.** In the domain of stylistics, this property concerns the additional, connotative, meaning of an affix (cf. Cruse 2010; Leech 1981; Prčić 2016). Stylistically, the affix contributes to bases, where applicable, restrictions on the word's range of application, indicated most notably by four labels: firstly, dialect, like the suffix *-ize*, in *modernize*, which is world-wide English, vs

the suffix *-ise*, in *modernise*, which is mainly British English; secondly, attitude, like the suffix *-eer*, in *profiteer*, which is disapproving; thirdly, formality, like the suffix *-er*, in *adviser*, which is neutral, vs the suffix *-or*, in *advisor*, which is rather formal; and fourthly, register, like the suffix *-ide*, in *chloride*, which belongs to chemistry.

(6) **Productivity.** In the domains of morphosyntax and semantics, this property concerns the ability of an affix to be synchronically used in the derivation of new words (cf. Bauer 1983, 2001; Bauer et al. 2013; Kastovsky 1986; Lieber and Štekauer 2014). What the affix contributes to bases are the effects of its relative position on a three-zone scale of productivity, spanning from high to restricted to low, with affixes in the high zone having greater likelihood of being chosen in the creation of new words, like the suffix *-er*, deriving agent and/or instrument nouns from dynamic verbs, when it is of high productivity, as in *play_v* + *-er* > *player_n*, or the suffix *-th*, deriving abstract nouns from adjectives, when it is of low productivity, as in *warm_{adj}* + *-th* > *warmth_n*.

2.2 Relational properties of English lexical affixes

Relevant to some affixes, i.e. prefixes and suffixes, only when certain content- and/or form-based conditions are met, and reflecting affixal companionability, relational properties comprise the interacting characteristics of an affix with another or others. According to the descriptive model adopted here, there are four relational properties, two governed by paradigmatics and two by syntagmatics.

Paradigmatic relations between affixes are perceived as interaffixal sense relations, established between individual meanings of affixes, which trigger two relational properties:

(7) **Interaffixal Synonymy.** In the domains of semantics and pragmatics, this property pertains to the fact that at least two affixes share the same descriptive meaning and morphosyntactic function (cf. Cruse 2010; Lipka 2002; Murphy 2003; Prčić 2016), and, as a result, are liable to derive synonymous words when attached to the same base, with all affixes being in the state of rivalry, or competition (cf. Bauer et al. 2013; Kastovsky 1982, 1986). Relationship of this kind can develop in three varieties:

- between prefixes, like *un-*, in *undress*, *de-*, in *destabilize*, and *dis-*, in *disconnect*, all in the sense 'to reverse', or, with the same base, *un-*, *im-* (*in-*), *non-*, *a-*, all in the sense 'not (concerned with)', in *unmoral*, *immoral*, *non-moral*, *amoral*,
- between suffixes, like *-ness*, in *happiness*, *-ity*, in *creativity*, and *-th*, in *width*, all in the sense 'state of being', or, with the same base, *-er*, *-or*, *-ant*, *-ee*, all in the sense 'agent', in *cohabiter*, *cohabitor*, *cohabitant*, *cohabitee*,

- between prefixes and suffixes, like the prefix *en-*, in *enlarge*, and the suffixes *-ize*, in *normalize*, *-ify*, in *purify*, and *-en*, in *widen*, all in the sense 'to make'.

(8) **Interaffixal Antonymy.** In the domains of semantics and pragmatics, this property pertains to the fact that two affixes have opposite descriptive meanings but the same morphosyntactic function (cf. Cruse 2010; Lipka 2002; Murphy 2003; Prčić 2016), and, as a result, are liable to derive antonymous words when attached to the same base. Relationship of this kind can develop in two varieties:

- between prefixes, like *over-* and *under-*, in *overcook* and *undercook*, in the sense 'more than necessary' and 'less than necessary', respectively,
- between suffixes, like *-er* and *-ee*, in *interviewer* and *interviewee*, in the sense 'agent' and 'affected', respectively.

On the other hand, syntagmatic relations between affixes are perceived as interaffixal preferences, which trigger two relational properties:

(9) **Cooccurring.** In the domains of morphosyntax and semantics, this property pertains to the fact that two affixes cooccur in words, because one affix tends to attract another in derived words (cf. Bauer et al. 2013). Relationship of this kind can develop in three varieties:

- between suffixes, like the suffix *-ation*, deriving action nouns, typically gets attracted by the suffixes *-ize*, in *industrialization*, *-ify*, in *purification*, and *-ate*, in *hyphenation* (in the latter case, *-ation* overlaps with *-ate*), with some suffixes attracting at least two synonymous suffixes, hence open to rivalry, like the nominal suffixes *-ness* and *-ity*, of varying degrees of frequency, both attracted by some adjectival bases in *-al*, as in *grammaticalness* and *grammaticality*,
- between prefixes, like the prefix *pro-*, deriving human nouns, in the sense 'deputy', typically gets attracted by the prefix *vice-*, in *pro-vice-chancellor*,
- between suffixes and prefixes, like the prefix *un-*, deriving negative adjectives, typically gets attracted by the suffix *-able*, in *uneatable*, whereas the prefix *in-*, in the same function, is typically attracted by the suffix *-ible*, in *inaccessible*, with some suffixes attracting at least two synonymous prefixes, hence also open to rivalry, like the negative adjectival prefixes *un-* and *a-*, of varying degrees of frequency, both attracted by some adjectival bases in *-al*, as in *untypical* and *atypical*.

The relationship of cooccurring is recursive in nature, especially with certain sets of cooccurrent suffixes, among which regular intersuffixal chain attractions develop, as in *developmentally*, with three suffixes (*-ment*, *-al*, *-ly*), and *organizationally*, with four suffixes (*-ize*, *-ation*, *-al*, *-ly*).

(10) **Correlationing.** In the domains of morphosyntax and semantics, this property pertains to the fact that two affixes correlate in words, because an affix tends to replace another affix in derived words. Relationship of this kind can develop in two varieties:

- between suffixes, like the suffix *-ism*, in the sense 'a system of theory, methodology or practice', deriving human nouns, getting replaced with the suffix *-ist*, in the sense 'an adherent of a system of theory, methodology or practice', as in *structural[ism] > structuralist*; in some instances, the suffix *-ism*, or *-ist*, is replaced with the suffix *-ize*, in the sense 'to follow a system of theory, methodology or practice', deriving action verbs, as in *terror[ism] or terror[ist] > terrorize*,
- between suffixes, like the adjectival suffixes *-able* and *-ible*, getting replaced with their bound allomorphs, *-abil-* and *-ibil-*, before a derivation of state nouns, with the nominal suffix *-ity* attaching, can take place, as in *desirable: desirabil- + -ity > desirability* and *flexible: flexibil- + -ity > flexibility*.

To conclude this overview, a reminder that in handling polyfunctional affixes, in which several distinct, related or unrelated, meanings and distinct functions coincide in a single form, properties (3)-(6) and, where applicable, (7)-(10) need to be specified for each function separately; for example, for the prefix *un-* in words like *unhappy*, *unrest*, *untie*, *unsaddle*, where it means 'not', 'lack of', 'reversal' and 'removal', respectively, and for the suffix *-ish* in words like *childish*, *yellowish*, *fiftyish*, *Finnish*, where it means 'typical of', 'somewhat', 'about', 'coming from', respectively.

3. Description of the new model

By the new model is meant the unity of two interlinked methods of acquiring the properties of English lexical affixes: the first is in the domain of language teaching and it adapts the **heuristic** approach to the acquisition of those properties; and the second method is in the domain of lexicography and it adopts an **inferential** approach to the presentation of affixal properties. The interlinkage of the two methods consists in the fact that acquisition and presentation have been conceived and realized so as to go hand in hand and to rely strongly on each other.

In what follows, the new model will be introduced in a detailed description, which will throw light on its five major aspects: pragmatic reasons (3.1), main objectives (3.2), underlying principles (3.3), instructional material (3.4) and real-life functioning (3.5).

3.1 Pragmatic reasons for the model

Working knowledge of affixal properties, as one type of lexical knowledge, is

of great potential utility for EFL students, including, naturally, EFL university students, in two significant respects: firstly, from a practical (communicative) viewpoint, essential affixal properties help students to understand compositional prefixations and suffixations, established and new alike, by decoding forms and interpreting their meanings (cf. Lehrer 2003; Metcalf 2002; Prčić 2001); later on, at a more advanced level of proficiency, to venture into creating their own new words by encoding meanings into forms. Secondly, from a theoretical ((meta)linguistic) viewpoint, essential and relational affixal properties can help students to understand better the networked organization and operation of (part of) the language system, its lexicon and, within the lexicon, its affixal word-formative module (cf. Bauer 1983; Dixon 2014; Marchand 1969; Miller 2014).

Information about affixal properties is normally included in general-purpose dictionaries, alongside information about words, and in specialized dictionaries of affixes. It is a well-known and easily verifiable fact that today's general-purpose dictionaries of English, both learner's and native-speaker ones, amid fierce competition between publishers and brands, excel at their treatment of words, using for their accurate portrayal ever-improving, corpus-driven and computer-assisted methods of electronic lexicography (cf. Fuertes-Olivera and Bergenholtz 2013; Gouws et al. 2013; Granger and Paquot 2012). In contrast, the treatment of affixes does not follow this, or any other identifiable methodology and, as a result, the picture of affixal properties in many general-purpose dictionaries is far from acceptable.

A study of the lexicographic treatment of two prefixes and two suffixes in four learner's dictionaries (Prčić 1999), known to the trade as the 'big four', has shown that there is much room for improvement in this area. Using the then current, 1995, print editions of *OALD*, *LDOCE*, *COBUILD* and *CIDE*, a comparative analysis of the treatment of four exemplary affixes was carried out — the prefixes *dis-* and *pro-*, and the suffixes *-ish* and *-or*, all selected because of their rich and diversified sets of properties, which pose a descriptive and, above all, lexicographic, challenge. The chief aim of the analysis was to try to identify the system and the method applied for the treatment of these affixes and, by extension, of affixes generally. The results obtained were quite eye-opening, even though not surprising, as it was ascertained that no firm system and no firm method was observed to be at work in any of the 'big four' dictionaries. Consequently, the information provided on affixal properties seems, by and large, to be fairly arbitrary, both in quantity and quality: as stated in the summing-up of the results (Prčić 1999: 274), the expected information may be lacking altogether, it may be given inconsistently and/or incompletely, it may be either under-refined or over-refined, or it may even be incorrect.

Twenty years on, it would be rewarding to conduct a new, two-level analysis of the treatment of affixes, this time using the current online editions of the 'fabulous five' (cf. Prčić 2004), *OALD*, *LDOCE*, *COBUILD*, *CALD* and *MEDAL*, in order, firstly, to find out their provision for the four affixes under scrutiny, and, secondly, to compare and contrast the 'diachronic' provisions in

the now and the then editions — all this with the aim of detecting superficial and substantial signs of improvement, if any.

Meanwhile, until the new analysis is performed and its results ready, here are the findings of a different, recently completed, comparative analysis, involving as many as fifteen general-purpose dictionaries, as few as two lexical affixes and as few as four affixal properties. The properties — Stress Placement (taken from Pronunciation), Base Part of Speech and Word Part of Speech (both taken from Attachability), and Descriptive Meanings, have been selected because they are the most challenging for students to acquire. The affixes — the prefix *un-* and the suffix *-ee*, have been selected because they perfectly exhibit all distinct facets of the four properties. The dictionaries — the online editions of *OALD*, *LDOCE*, *COBUILD*, *CALD*, *MEDAL* (British, learners'), *MWLD*, *RHLD* (American, learners'), *ODE*, *CED*, *C21CD* (British, native speakers'), *RHUD*, *AHD*, *MWCD*, *WNWCD* (American, native speakers') and *Wiktionary* (global), have been selected because they are universally and deservedly considered to be among the most respected, consulted and trusted dictionaries of English.

The coverage of the four properties has been checked in these fifteen dictionaries with the following expectations:

- Stress Placement — for *un-*, indication of the syllable in the word bearing primary stress (the prefix is stress-neutral); for *-ee*, indication of the last syllable in the word, bearing primary stress under the influence of *-ee* (the suffix is stress-imposing).
- Base Part of Speech — for *un-*, indication of adjective, adverb, noun or verb, as appropriate; for *-ee*, indication of verb, adjective or noun, as appropriate (both affixes are class-changing in some functions and class-maintaining in others).
- Word Part of Speech — for *un-*, indication of adjective, adverb, noun or verb, as appropriate; for *-ee*, indication of noun.
- Descriptive Meanings — for *un-*, indication of four meanings (not, lack of, reversal, removal); for *-ee*, indication of three meanings (affected, agent, one involved in).

The findings are summarized in Table 1, where the letter 'Y' stands for 'yes, provision is given', the letter 'N' for 'no, provision is not given' and the letter 'P' for 'partial provision is given', all implying variably fulfilled expectations; for the meanings, if provision is below the total expected, a figure appears to denote the number of meanings covered; the best two overall provisions are highlighted.

		<i>OALD</i>	<i>LDOCE</i>	<i>COBUILD</i>	<i>CALD</i>	<i>MEDAL</i>	<i>MWLD</i>	<i>RHLD</i>	<i>ODE</i>	<i>CED</i>	<i>C2/CD</i>	<i>RHUD</i>	<i>AHD</i>	<i>MWCD</i>	<i>WNWCD</i>	<i>Wiktionary</i>
<i>un-</i>	Stress Placement	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	Base Part of Speech	N	N	Y	Y	Y	N	P	Y	P	Y	P	N	Y	P	P
	Word Part of Speech	Y	Y	P	N	N	N	P	N	P	P	P	N	Y	N	P
	Descriptive Meanings	3	3	2	2	2	3	Y	Y	Y	3	Y	Y	Y	Y	Y
<i>-ee</i>	Stress Placement	N	N	/	N	N	N	N	N	N	N	N	N	N	N	N
	Base Part of Speech	N	N	/	Y	Y	N	Y	P	N	P	P	N	N	N	P
	Word Part of Speech	Y	Y	/	P	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N
	Descriptive Meanings	2	Y	/	2	2	2	Y	2	2	2	Y	Y	Y	Y	Y

Table 1: Coverage of selected properties of selected affixes in selected dictionaries

The tabulated data clearly shows that the word best capturing the lexicographic treatment of affixes and their properties, both in learner's and native-speaker's dictionaries, both British and American, is — **inconsistency**. As can be noticed, one dictionary leaves out the suffix entirely; not a single dictionary regards stress placement as important enough for inclusion; contrariwise, around a half have quite adequate provision for the meanings (8 prefixal vs 7 suffixal), although not all are neatly presented and defined; however, the most inconsistency mars surveys of base and word parts of speech. Needless to say, an in-depth examination of the information retrieved from the fifteen dictionaries would require a full-length paper, to demonstrate empirically the quantity and the quality of the coverage of affixal properties. What can already be asserted with confidence is that inconsistency seems generally to be a strong and prevailing trend — and for no obvious reason.

The conclusion, reached two decades ago, that affixes are, on the whole, treated "as nothing more than poor relations of words and phrases" (Prčić 1999: 274), sadly, remains valid today and, to all intents and purposes, is likely to persist. In consequence, it would be both pointless and useless to recommend to students, especially EFL university students, general-purpose dictionaries, containing bafflingly erratic sketches of affixes, as sources of information about the properties of English lexical affixes.

In order to lessen the effects of this state of affairs the simplest solution is a repurposing of dictionaries. More specifically, considering the fact that general-

purpose dictionaries are highly reliable in their treatment of words, the self-same dictionaries should *somehow* be made usable for extracting information about affixes from the information they provide about words. That *somehow*, the driving force behind the new model of exploring affixal properties, will be elucidated in the following sections.

3.2 Main objectives of the model

The new model, intended especially for EFL university students, has been continuously tried and tested for over a decade with third-year undergraduates attending a course in Lexical Morphology at the Department of English, Faculty of Philosophy, University of Novi Sad. The objectives of this model are threefold:

- to provide students with an efficient, effective and user-friendly way of familiarizing affixal properties;
- to provide students with an intriguing and stimulating teaching aid, that will guide them through the intricacies of English lexical affixes and thus allow them a fuller and more consistent insight into affixal properties than is presently available;
- to devise and compile a novel **bare-bones inferential dictionary** of English affixes (henceforward, **BBID**), a hybrid between a teaching aid and a dictionary, or, more precisely, a teaching aid in the shape of a dictionary, in which two systems for gaining information are deployed: supplying users with information in the dictionary itself and directing users to look for information elsewhere, in English learner's and/or native-speaker general-purpose dictionaries.

3.3 Underlying principles behind the model

From a theoretical point of view, underpinning this model are the definition of the affix, the determination of the nature of its properties and their classification into essential and relational ones, as outlined in Section 2.

From a methodological point of view, underpinning this model is the heuristic approach to teaching and learning (cf. Kumaravadivelu 2003; Sale 2015; Seliger 1975; Takimoto 2008), whereby students are guided step by step, by appropriately given prompts in BBID, to actively discover, or infer, information and/or gain knowledge on their own, rather than being simply their passive recipients. In this particular case of self-study, the heuristic builds around two key lexical objects — target affix and source word.

Occupying central position in the model is the **target affix**, a prefix or a suffix, whose properties are there to be discovered, like the prefix *un-* and the suffix *-ee*. Closely connected with the target affix is the **source word**, one or more of them, depending on the number of meanings and functions involved,

like 1 *un.happy*, *un.English*, *un.beaten*, *un.caring*, 2 *un.rest*, 3 *un.tie*, 4 *un.saddle*_v, for the prefix, and 1 *employ.ee*_N, *pay.ee*_N, 2 *escap[e].ee*_N, 3 *absent.ee*_N, for the suffix. Source words are representative of the entire set of affixal properties and they play a dual role: firstly, they provide a natural surroundings for the target affix and contextualize its available meaning(s) and function(s); and, secondly, and much more importantly, source words contain target-affix properties to be discovered and exhibit them in a clear and prototypical form, as systematic and predictable contributions in compositional formations.

The required properties can be inferred in two complementary ways, interconnected with the provision of two complementary types of information that students are supposed to discover, collate and organize into a coherent whole: the first is by reading immediate information off the source words themselves, mostly relating to form and partly function, thereby inferring **obvious information**, explicitly given and just waiting to be recognized and registered; and the second way is by exploiting information, offered in English learner's and/or native-speaker general-purpose dictionaries, *about* the source words and, by implication, the target affixes, mostly relating to content and partly function, thereby inferring **hidden information**, implicitly signalled and prompted to be sought and found elsewhere.

The affixal properties deemed necessary and sufficient for EFL undergraduates address their passive knowledge of the processes of English derivation and their concrete results, and therefore equip students to understand, interpret or decode derived words — established and new alike. Out of the two sets of essential and relational properties discussed in Section 2, at the end of their work based on the model being outlined, what is exactly expected from students is familiarity with essential affixal properties, except for Subcategorization and Productivity, which, alongside relational properties, belong rather to the active knowledge of postgraduates. According to this model, the following seven properties are to be mastered by students and, for this reason, are regarded as core affixal properties: (1) Spelling, the only property presumed to be known by students and therefore dispensable from this checklist, (2) Spelling Niceties, (3) Pronunciation, (4) Stress Placement, (5) Parts of Speech, (6) Descriptive Meaning, and (7) Associative Meaning.

3.4 Instructional material for the model

This model is meant to be realized with the support of uncommonly modest instructional material, in the form of a single handout, taking up a mere two A4 pages. This information-packed handout — containing, in fact, the full BBID — offers all that is to be known and done about becoming familiar with the core properties of English lexical affixes. BBID adopts two earlier created lists of prefixes and suffixes (cf. Prčić 2005, 2008), which constitute English synchronic prefixes and suffixes, frequent and necessary enough to warrant inclusion into an EFL university students' teaching aid of this kind. The dictionary's structure

will now be explained and exemplified (for theoretical, methodological and practical aspects of structuring dictionaries, see Atkins and Rundell 2008; Fontenelle 2008; Hartmann 2001; Hartmann and James 1998; Landau 2001; Svensén 2009); for greater ease of reference, in Appendix 1 and 2, typographically reflowed contents of the original handout are reproduced in their entirety.

As it can be seen in the reproductions, the macrostructure of this dictionary is organized around two numbered alphabetical lists of affixes, populated with altogether 100 prefixes and 113 suffixes. After brief initial notes on the top, the lexicographic treatment of these affixes is divided between four structural features, like this:

- acting as headwords, placed at the beginning of their lines and printed in bold are target prefixes and suffixes, followed or preceded, respectively, by a hyphen, as in (elements under discussion are emphasized by shading):

prefix:

6. **be-**: 1 be.friend_v, be.little_v | 2 be.spectacled

suffix:

68. **-ish**: 1 child.ish_{ADJ} | 2 yellow.ish_{ADJ} | 3 fifty.ish_{NUM} | 4 Finn.ish_{ADJ/N}

- acting as form- and content-related comment, continuing on the same line and printed in ordinary type are numbered source words, one or more of them, mirroring the meanings and functions of an affix, with a dot inside showing the boundary between a base and an affix, as in:

prefix:

6. **be-**: 1 be.friend_v, be.little_v | 2 be.spectacled

suffix:

68. **-ish**: 1 child.ish_{ADJ} | 2 yellow.ish_{ADJ} | 3 fifty.ish_{NUM} | 4 Finn.ish_{ADJ/N}

- acting as between-entry one-way cross-references, placed at the beginning of their lines, printed in italics and directed to by the prompt 'see', are allomorphs of affixes cross-referred to their canonical forms, where they receive full treatment, as in:

prefixes:

24. *em-* see *en-*

25. **en-**: 1 en.large_v | 2 en.throne_v; **em-**: 1 em.power_v | 2 em.body_v

suffixes:

44. *-et* see *-ette*

45. **-ette**: 1 disk.ette_N | 2 leather.ette_N | 3 usher.ette_N, Paul.ette_N;

-et: 1 baron.et_N / baron.et_N, falcon.et_N;

- only in the suffix segment, acting as within-entry two-way cross-references, placed at the very end of the same lines, printed in italics and

directed to by the prompt 'see also', are suffixes cross-referred to formally and/or semantically related suffixes, and vice versa, as in:

1. **-able**: 1 wash.able_{ADJ}, pay.able_{ADJ} | 2 change.able_{ADJ} | 3 fashion.able_{ADJ};
see also *-ible*
- ...
54. **-ible**: 1 digest.ible_{ADJ}; see also *-able*.

Provision of concrete core properties in BBID is based on its notation, specially devised to be simple, self-explanatory and easy-to-remember. The role of the notation is twofold: to convey obvious, readily recognizable, information about the properties of affixes, as given in the dictionary, and to convey hidden information, indirectly prompting users to look for it in another dictionary. In the following paragraphs, explanations will be given of the types of notation used to introduce specific pieces of obvious, hidden and mixed information about the seven core affixal properties:

(1) Spelling:

- (obvious information) provided for all target affixes is the written form of an affix in isolation, printed in bold, with prefixes followed by a hyphen and suffixes preceded by a hyphen, to show their bound nature: e.g. **equi-**, **-ness**,
- (obvious information) provided for all source words containing target affixes is the written form of an affix within the source word, with the dot to be ignored here: e.g. **equi**.distant, kind.**ness**_N,
- (obvious information) if two forms of a target affix or a source word are separated by a slash, it indicates written and/or spoken variants: e.g. **deca-/deka-**, lion.**ess**_N / lion.**ess**_N,
- (obvious information) if a letter or a sequence of letters, with or without morphemic status, in a source word is enclosed within square brackets, it indicates omission of that letter or sequence: e.g. happ[y]i.**ness**_N, particip[ate].**ant**_N,
- (obvious information) if an additional consonant letter, preceded by a plus sign, is inserted into the source word, it indicates doubling of that letter: e.g. hat+t.**er**_N.

(2) Spelling Niceties:

- (obvious information) if a target prefix is followed by two hyphens, it indicates that the prefix and its base may occur hyphenated in words: e.g. **micro--**.

(3) Pronunciation:

- (hidden information) to be discovered by checking the pronunciation of a target affix within the source word: e.g. **quasi--** quasi.scientific, **-ish** child.**ish**_{ADJ}.

- (4) Stress Placement:
- (mixed information) provided for all source words containing stress-imposing target suffixes is an underline at the syllable bearing primary stress, to be exactly discovered by checking the pronunciation of the source word: e.g. contempt.**uous**_{ADJ}, atom.**ic**_{ADJ}, employ.**ee**_N,
 - (mixed information) if no underline is provided, it indicates that primary stress of the source word agrees with that of its base, to which a stress-neutral target affix has attached, to be exactly discovered by checking the pronunciation of the source word: e.g. **mis**.spell, wash.**able**_{ADJ}.
- (5) Parts of Speech:
- (obvious information) provided for all source words containing class-changing target prefixes and all target suffixes is a subscript after a source word showing its part of speech: e.g. **be**.friend_V, gold.**en**_{ADJ},
 - (obvious information) if no subscript is provided after a source word, it indicates that its part of speech agrees with that of the base, to which a class-maintaining target prefix has attached: e.g. **co**.author,
 - (hidden information) to be discovered for all target affixes is the part of speech of the base by checking the source word while ignoring the affix and foregrounding the base: e.g. **ante**.date, **de**.ice_V, usher.**ette**_N.
- (6) Descriptive Meaning:
- (hidden information) to be discovered for all target affixes by checking the descriptive meaning of the source word: e.g. **un**.saddle_V, reader.**ship**_N.
- (7) Associative Meaning:
- (hidden information) to be discovered for all target affixes by checking the associative meaning of the source word: e.g. profit.**eer**_N.

As it can be gathered, the central structural feature of BBID, which is responsible for conveying some of the obvious information and all of the hidden information, is its triple cross-referential system. It is meant to work on two levels — internally, when handling obvious information, provided in this dictionary, and externally, when handling hidden information, provided in other dictionaries. Internal cross-references can be twofold: the within-entry type operates bidirectionally and is announced by the prompt 'see also', while the between-entry type operates unidirectionally and is announced by the prompt 'see'. In contrast, external cross-references always operate unidirectionally, have no announcing prompts, but their specific notation, directing users to sources beyond BBID.

3.5 Real-life functioning of the model

It will shortly become apparent that this unusual, if not peculiar, self-study

model is proving both usable and useful, bringing tangible benefits for students, since the slightly roundabout route to exploring the properties of English lexical affixes by exploiting the resources of English general-purpose dictionaries is in several respects superior to the seemingly quick direct route of checking affixal entries in dictionaries only to be faced with inconsistent and incomplete information for which hardly any use can be found.

With the objectives formulated and the accompanying instructional material prepared, the new model is put into actual practice. This process comprises four phases:

- Setting the task in class by the teacher, which involves explaining to the students, with the aid of an instruction sheet and concrete examples, firstly, the affixal properties to be found out; secondly, the organization and use of BBID; thirdly, the method of finding out affixal properties step by step; fourthly, a list of English learner's and native-speaker dictionaries to be consulted (comprising the fifteen titles mentioned in Section 3.1); and fifthly, the reasons for taking this circuitous, heavily cross-referential, route over the well-trodden path to looking up affixes alone.
- Doing the task by the students at home, which involves, firstly, checking source words and discovering the properties of target affixes; and secondly, putting together the information thus obtained into a profile of each affix — this phase is expected to be completed within three weeks.
- Discussing the students' findings in class, which involves, firstly, analysing the results achieved by reviewing their affix profiles; secondly, resolving points of difficulty and/or uncertainty, preferably student-to-student or, as a last resort, teacher-to-students; and thirdly, eliciting student feedback on their conducting research in this particular way.
- Testing the students' knowledge, which involves, firstly, tackling different example words either in isolation or in texts, in follow-up exercises, mock exams and the end-of-course exam; secondly, evaluating the students' level of affixal proficiency; and thirdly, reassessing the merits and demerits of this model, with a view to constantly improving its efficiency — theoretically, methodologically and practically.

4. Summing up, appraising and looking ahead

In this paper, a viable interim solution to acquiring the properties of English lexical affixes by EFL university students has been presented. The solution involves a new model of heuristic self-study of affixal properties, which builds around a novel BBID, intended to guide students on their discoveries. Without doubt the most important advantage of this model is that the results obtained by students are, on aggregate, the same or very similar, regardless of the dictionary, or dictionaries, consulted. In this way, all students have come to share

roughly the same working knowledge of affixes and their properties, which has resulted in the majority being able to apply with aplomb their newly-acquired knowledge to other derived words, established and new, arising in concrete situations. This objective, and the prime motive for initiating this project, has apparently been accomplished, alongside the other objective, that an appropriate teaching aid in the shape of a dictionary be conceived and produced.

Judging by the students' spontaneous reactions, they see the relative merits of this model in the following: they can work undistracted, at their own pace; they can gain knowledge through their own individual research; they can freely exploit the dictionaries they already have or browse on the web, with no extra expenses incurred; they can recognize affixal properties in different words of the same types; they can understand and interpret established and new words containing prefixes and/or suffixes. As relative demerits of this model the students have singled out the following: time-consuming and never-ending task; dauntingly large number of affixes and their properties to be dealt with; problems with first discovering and then identifying and formulating definitions of the meaning(s) of some affixes. An additional cautionary note from the teacher may be of interest: over the years, it has become known that a few students are put off by BBID and choose to quietly perform their research in the conventional manner, looking up affixes proper and inevitably achieving limited success; still more frustratingly, very few students are put off by self-study and choose to abandon the research completely.

The students' opinions on BBID have always started from an initial sense of dismay and disappointment with the novel dictionary's austere appearance and lack of information found in typical, normal, dictionaries. Then, after explanations which helped them to grasp the system behind BBID and its expected use, and further clarifications of the unique role of the source words and the hidden, cross-referential information, there ensued a gradual swing of opinion towards liking the dictionary and growing to appreciate the wealth of information it offers explicitly and implicitly. However, one student remarked wryly that she had enjoyed discovering about affixal properties very much but she would have preferred to have a print or online dictionary with all the information they had had to dig up for themselves. Not surprisingly, this remark met with overwhelming approval of her fellow students.

Echoing the title of the renowned *Longman Language Activator* (Summers 2004), the purposefully designed innovative BBID, by its function perhaps best designated *English Affix Activator*, offers all that is necessary for students' heuristic self-study process — explicit information and implicit instructions for discovering information about the properties of English lexical affixes in general-purpose dictionaries. Although it has proved to be an efficient teaching aid in its current form, BBID could be expanded in scope, by the inclusion of indications for the properties of Subcategorization and Productivity, from the set of essential properties, and internal two-way cross-references for Interaffixal Synonymy and Interaffixal Antonymy, from the set of relational properties. These

additions would turn BBID, a (highly) minimalist dictionary, into an enlarged and more informative reference work, an (almost) optimal dictionary, with an educational role.

Lastly, in order to make this optimal dictionary into a full-fledged, comprehensive dictionary of affixal properties, all properties, both essential and relational, discussed above should be included in full, with an effective presentation and fine exemplification, where silent source words are promoted into telling examples. In this way, dispensing with implicitly given information, which prompts users to some activity, and concentrating solely on explicit information, a genuinely useful piece of lexicography, preferably in electronic form, entitled provisionally *An Advanced Learner's Dictionary of English Lexical Affixes* (or, *ALDELA*, for short), could be produced, drawing on the strong (and not so strong) theoretical and methodological points underpinning *Collins COBUILD English Guides 2: Word Formation* (Sinclair 1991), *Word Parts Dictionary: Standard and Reverse Listings of Prefixes, Suffixes, and Combining Forms* (Sheehan 2000), *Ologies and Isms: A Dictionary of Word Beginnings and Endings* (Quinion 2002) and *Affixes: The Building Blocks of English*, its online variety — and, especially, *A Dictionary of English Affixes: Their Function and Meaning* (Stein 2007).

Such a dictionary would be a significant step in the right direction and would greatly reduce the need for acquiring the properties of English lexical affixes by relying on BBID and prompting users to turn to general-purpose dictionaries for a substantial and valuable assistance. However, discussion of this dictionary of the future and its kindred predecessors lies outside the scope of the present paper.

Acknowledgements

This is an expanded version of the paper presented at the 8th International Conference *Languages and Cultures in Time and Space*, Novi Sad, November 2018. The paper is part of the research on Project No. 178002, entitled *Languages and Cultures in Time and Space*, which is financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

For constructive comments and suggestions the author wishes to thank two anonymous reviewers.

References

A. Dictionaries

Printed

Jones, D. 2006. *Cambridge English Pronouncing Dictionary*. Seventeenth edition. Edited by Roach, P., Hartman, J. and J. Setter. Cambridge: Cambridge University Press.

- Quinion, M.** 2002. *Ologies and Isms: A Dictionary of Word Beginnings and Endings*. Oxford: Oxford University Press.
- Sheehan, M.J.** 2000. *Word Parts Dictionary: Standard and Reverse Listings of Prefixes, Suffixes, Roots and Combining Forms*. Jefferson/London: McFarland.
- Sinclair, J. (Ed.)**. 1991. *Collins COBUILD English Guides 2: Word Formation*. London: HarperCollins.
- Stein, G.** 2007. *A Dictionary of English Affixes: Their Function and Meaning*. Munich: Lincom Europa.
- Summers, D. (Ed.)**. 2004. *Longman Language Activator*. Second edition. Harlow: Pearson Education.
- Wells, J.C.** 2008. *Longman Pronunciation Dictionary*. Third edition. Harlow: Pearson Education.

Online

- Affixes: The Building Blocks of English*: <http://www.affixes.org/>.
- The American Heritage Dictionary of the English Language, the (AHD)*: <https://ahdictionary.com/>.
- Cambridge Advanced Learner's Dictionary and Thesaurus (CALD)*: <https://dictionary.cambridge.org/>.
- Chambers 21st Century Dictionary (C21CD)*: <https://chambers.co.uk/>.
- COBUILD Advanced English Dictionary (COBUILD)*: <https://www.collinsdictionary.com/>.
- Collins English Dictionary (CED)*: <https://www.collinsdictionary.com/>.
- Longman Dictionary of Contemporary English (LDOCE)*: <https://www.ldoceonline.com/>.
- Macmillan English Dictionary for Advanced Learners (MEDAL)*: <http://www.macmillandictionaries.com/>.
- Merriam Webster's Collegiate Dictionary (MWCD)*: <https://www.merriam-webster.com/>.
- Merriam Webster's Learner's Dictionary (MWLD)*: <http://learnersdictionary.com/>.
- Oxford Advanced Learner's Dictionary (OALD)*: <https://www.oxfordlearnersdictionaries.com/>.
- Oxford Dictionary of English (ODE)*: <https://en.oxforddictionaries.com/>.
- Random House Learner's Dictionary of American English (RHLD)*: <http://www.wordreference.com/definition/>.
- Random House Unabridged Dictionary (RHUD)*: <https://www.dictionary.com/>.
- Webster's New World College Dictionary (WNWCD)*: <https://www.yourdictionary.com/>.
- Wiktionary, the Free Dictionary*: <https://en.wiktionary.org/>.

B. Other literature

- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bauer, L.** 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, L.** 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
- Bauer, L. and R. Huddleston.** 2002. *Lexical Word Formation*. Huddleston, R. and G.K. Pullum (Eds.). 2002. *The Cambridge Grammar of the English Language: 1621-1721*. Cambridge: Cambridge University Press.
- Bauer, L., R. Lieber and I. Plag (Eds.)**. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Cruse, A.** 2010. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Third edition. Oxford: Oxford University Press.

- Dixon, R.M.W. 2014. *Making New Words: Morphological Derivation in English*. Oxford: Oxford University Press.
- Fontenelle, T. (Ed.). 2008. *Practical Lexicography: A Reader*. Oxford: Oxford University Press.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2013. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*. London: Bloomsbury Academic.
- Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries: An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin/New York: De Gruyter Mouton.
- Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. Harlow: Longman / Essex: Pearson Education.
- Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hohenhaus, P. 2005. Lexicalisation and Institutionalisation. Štekauer, P. and R. Lieber (Eds.). 2005. *Handbook of English Word-Formation*: 353-373. Dordrecht: Springer.
- Kastovsky, D. 1982. Word-Formation: A Functional View. *Folia Linguistica* 16: 181-198.
- Kastovsky, D. 1986. The Problem of Productivity in Word Formation. *Linguistics* 24: 585-600.
- Kumaravadivelu, B. 2003. *Beyond Methods: Macrostrategies for Language Teaching*. New Haven: Yale University Press.
- Landau, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. New York/Cambridge: Cambridge University Press.
- Leech, G. 1981. *Semantics: The Study of Meaning*. Second edition. Harmondsworth: Penguin Books.
- Lehrer, A. 2003. Understanding Trendy Neologisms. *Italian Journal of Linguistics* 15(2): 369-382.
- Lieber, R. and P. Štekauer (Eds.). 2014. *The Oxford Handbook of Derivational Morphology*. Oxford: Oxford University Press.
- Lipka, L. 2002. *English Lexicology: Lexical Structure, Word Semantics, and Word-Formation*. Third edition. Tübingen: Gunter Narr.
- Marchand, H. 1969. *The Categories and Types of Present-Day English Word-Formation*. Second, completely revised and enlarged edition. München: C. H. Beck'sche Verlagsbuchhandlung.
- Metcalf, A. 2002. *Predicting New Words: The Secrets of their Success*. Boston/New York: Houghton Mifflin.
- Miller, D.G. 2014. *English Lexicogenesis*. Oxford: Oxford University Press.
- Murphy, M.L. 2003. *Semantic Relations and the Lexicon: Antonymy, Synonymy, and Other Paradigms*. Cambridge: Cambridge University Press.
- Prčić, T. 1999. The Treatment of Affixes in the 'Big Four' EFL Dictionaries. *International Journal of Lexicography* 12(4): 263-279.
- Prčić, T. 2001. Interpreting Morphologically Complex Lexemes Revisited. Németh, T.E. and K. Bibok (Eds.). *Pragmatics and the Flexibility of Word Meaning*: 225-244. Oxford: Elsevier.
- Prčić, T. 2004. Enter the 'Big Fifth' EFL Dictionary: *Macmillan English Dictionary for Advanced Learners*. *Lexicographica* 20: 303-322.
- Prčić, T. 2005. Prefixes vs Initial Combining Forms in English: A Lexicographic Perspective. *International Journal of Lexicography* 18(3): 313-334.
- Prčić, T. 2008. Suffixes vs Final Combining Forms in English: A Lexicographic Perspective. *International Journal of Lexicography* 21(1): 1-22.

- Prčić, T.** 2016. *Semantika i pragmatika reči*. 3., elektronsko, izdanje [*Semantics and Pragmatics of the Word*. Third, digital, edition]. Novi Sad: Filozofski fakultet. Accessed at: <http://digitalna.ff.uns.ac.rs/sadrzaj/2016/978-86-6065-356-9> (30 March 2019).
- Prčić, T.** 2019, in press. Distinctive Properties of English Lexical Affixes: An Overall View. *Belgrade English Language and Literature Studies (Belgrade BELLS), Special Issue in Honour of Boris Hlebec*.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik.** 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.
- Sale, D.** 2015. *Creative Teaching: An Evidence-Based Approach*. Singapore: Springer.
- Saussure, F. de.** 1916. *Cours de linguistique générale*. Paris: Payot.
- Seliger, H.W.** 1975. Inductive Method and Deductive Method in Language Teaching: A Re-examination. *International Review of Applied Linguistics in Language Teaching (IRAL)* 13(1): 1-18.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Štekauer, P. and R. Lieber (Eds.).** 2005. *Handbook of Word-Formation*. Dordrecht: Springer.
- Takimoto, M.** 2008. The Effects of Deductive and Inductive Instruction on the Development of Language Learners' Pragmatic Competence. *The Modern Language Journal* 92: 369-386.

Appendix 1: A bare-bones inferential dictionary of English lexical prefixes

NOTE: Two hyphens following a prefix indicate that the prefix often attaches to the base with a hyphen.

1. **a-**: 1 a.typical | 2 a.shore_{ADV} | 3 a.loud_{ADV}
2. **ante-**: 1 ante.date | 2 ante.chamber
3. **anti--**: 1 anti.democratic | 2 anti.bacterial, anti.pyretic | 3 anti.climax, anti.clockwise
4. **arch-**: 1 arch.duke | 2 arch.enemy | 3 arch.conservative
5. **auto--**: 1 auto.focus | 2 auto.suggestion
6. **be-**: 1 be.friend_V, be.little_V | 2 be.spectacled
7. **bi--**: 1 bi.lingual | 2 bi.annual
8. **by--/bye--**: 1 by.product, by.road | 2 by(e).election
9. **centi-**: 1 centi.metre / centi.meter
10. **circum-**: 1 circum.navigate
11. **cis-**: 1 cis.alpine
12. **co--**: 1 co.author, co.produce | 2 co.pilot
13. **contra-**: 1 contra.indication | 2 contra.bassoon
14. **counter--**: 1 counter.attack | 2 counter.clockwise | 3 counter.part
15. **cyber-**: 1 cyber.talk
16. **de--**: 1 de.stabilize | 2 de.ice_V | 3 de.verbal
17. **deca-/deka-**: 1 deca.gram / deka.gram
18. **deci-**: 1 deci.bel
19. **demi-**: 1 demi.god | 2 demi.semiquaver
20. **di-**: 1 di.syllabic
21. **dis-**: 1 dis.connect | 2 dis.similar, dis.agree | 3 dis.bud_V
22. **dys-**: 1 dys.functional
23. **e--**: 1 e.banking
24. *em-* see *en-*
25. **en-**: 1 en.large_V | 2 en.throne_V; **em-**: 1 em.power_V | 2 em.body_V
26. **endo-**: 1 endo.centric
27. **equi-**: 1 equi.distant
28. **ex--**: 1 ex.president
29. **exo-**: 1 exo.centric
30. **extra--**: 1 extra.marital | 2 extra.large
31. **fore-**: 1 fore.see | 2 fore.arm
32. **giga-**: 1 giga.hertz | 2 giga.byte
33. **half--**: 1 half.cooked | 2 half.Irish | 3 half.sister
34. **hecto-**: 1 hecto.litre / hecto.liter
35. **hemi-**: 1 hemi.sphere
36. **hetero-**: 1 hetero.sexual
37. **hexa-**: 1 hexa.decimal
38. **homo-**: 1 homo.sexual
39. **hyper--**: 1 hyper.sensitive | 2 hyper.market
40. **hypo-**: 1 hypo.allergenic

41. *il-* see *in-*
42. **ill--**: 1 ill.chosen, ill.fitting
43. *im-* see *in-*
44. **in-**: 1 in.expensive | 2 in.activity; **il-**: 1 il.legal; **im-**: 1 im.patient; **ir-**: 1 ir.regular
45. **infra-**: 1 infra.red
46. **inter--**: 1 inter.national | 2 inter.relate
47. **intra--**: 1 intra.galactic
48. *ir-* see *in-*
49. **kilo-**: 1 kilo.gram | 2 kilo.byte
50. **macro--**: 1 macro.organism | 2 macro.climate | 3 macro.biotic
51. **mal-**: 1 mal.nutrition, mal.treat
52. **maxi--**: 1 maxi.skirt
53. **Mc-**: 1 Mc.Book
54. **mega--**: 1 mega.store, mega.hit | 2 mega.ohm | 3 mega.byte
55. **meta-**: 1 meta.linguistics
56. **micro--**: 1 micro.organism, micro.chip | 2 micro.electronics | 3 micro.climate | 4 micro.ampere
57. **milli-**: 1 milli.second
58. **mini--**: 1 mini.bus
59. **mis-**: 1 mis.spell, mis.use | 2 mis.trust
60. **mono-**: 1 mono.chromatic
61. **multi--**: 1 multi.racial
62. **nano-**: 1 nano.second
63. **neo--**: 1 neo.classical, neo.Darwinism
64. **non--**: 1 non.alcoholic, non.English, non.aggression | 2 non.smoker, non.fiction | 3 non.skid_{ADJ} | 4 non.entity
65. **omni-**: 1 omni.present
66. **out-**: 1 out.swim
67. **over-**: 1 over.cook, over.ambitious | 2 over.rule
68. **pan--**: 1 pan.African
69. **para-**: 1 para.normal | 2 para.medical | 3 para.gliding
70. **penta-**: 1 penta.syllabic
71. **pico-**: 1 pico.farad
72. **poly-**: 1 poly.syllabic
73. **post--**: 1 post.graduate | 2 post.alveolar
74. **pre--**: 1 pre.marital | 2 pre.shrunk | 3 pre.molar
75. **preter-**: 1 preter.natural
76. **pro--**: 1 pro.European | 2 pro.active | 3 pro.consul
77. **proto-**: 1 proto.language
78. **pseudo--**: 1 pseudo.science | 2 pseudo.intellectual
79. **quadri-**: 1 quadri.lateral
80. **quasi--**: 1 quasi.scientific | 2 quasi.official
81. **re--**: 1 re.write | 2 re.call
82. **retro-**: 1 retro.active
83. **self--**: 1 self.respect, self.taught | 2 self.propelling
84. **semi--**: 1 semi.circle, semi.quaver | 2 semi.automatic, semi.skilled | 3 semi.annual

85. **step-**: 1 step.son
86. **sub--**: 1 sub.standard | 2 sub.section, sub.lease | 3 sub.editor | 4 sub.human | 5 sub.aquatic
87. **super--**: 1 super.intelligent | 2 super.market | 3 super.star, super.power | 4 super.sonic, super.natural | 5 super.impose
88. **supra-**: 1 supra.segmental
89. **sur-**: 1 sur.charge
90. **tele-**: 1 tele.kinesis | 2 tele.conferencing
91. **tera-**: 1 tera.watt | 2 tera.byte
92. **tetra-**: 1 tetra.meter
93. **trans-**: 1 trans.continental | 2 trans.national | 3 trans.alpine | 4 trans.form
94. **tri-**: 1 tri.partite
95. **uber-**: 1 uber.cool
96. **ultra--**: 1 ultra.modern | 2 ultra.violet
97. **un-**: 1 un.happy, un.English, un.beaten, un.caring | 2 un.rest | 3 un.tie | 4 un.saddle
98. **under--**: 1 under.cook | 2 under.secretary
99. **uni-**: 1 uni.directional
100. **vice--**: 1 vice.president

Appendix 2: A bare-bones inferential dictionary of English lexical suffixes

NOTE: Underlined syllables indicate the place of primary stress in words with stress-imposing suffixes.

1. **-able**: 1 wash.able_{ADJ}, pay.able_{ADJ} | 2 change.able_{ADJ} | 3 fashion.able_{ADJ}; see also *-ible*
2. **-aceous**: 1 curv[e].aceous_{ADJ}; see also *-ous*
3. **-ade**: 1 block.ade_N | 2 lemon.ade_N
4. **-age**: 1 cover.age_N | 2 mile.age_N | 3 post.age_N | 4 orphan.age_N | 5 patron.age_N | 6 peer.age_N, sewer.age_N
5. **-aire**: 1 million.aire_N
6. **-al**: 1 emotion.al_{ADJ}, magic.al_{ADJ}, parent.al_{ADJ}, origin.al_{ADJ} | 2 renew.al_N; **-ial**: 1 manager.ial_{ADJ}, professor.ial_{ADJ}; **-ual**: 1 fact.ual_{ADJ}, aspect.ual_{ADJ}; see also *-ical*
7. **-alia**: 1 kitchen.alia_{N(PL)}
8. **-an**: 1 Tibet.an_{ADJ/N}, Americ[a].an_{ADJ/N} | 2 Chomsky.an_{ADJ} | 3 republic.an_{ADJ/N}; **-ean**: 1 Europ[e].ean_{ADJ/N}, Carib+b.ean_{ADJ/N} / Carib+b.ean_{ADJ/N} | 2 Keynes.ean_{ADJ}, Shakespear[e].ean_{ADJ}; **-ian**: 1 Brazil.ian_{ADJ/N}, Austr[ia].ian_{ADJ/N}, Hungar[y].ian_{ADJ/N} | 2 Dickens.ian_{ADJ}, Shakespear[e].ian_{ADJ} | 3 mathematic.ian_N; see also *-ician*
9. **-ana**: 1 Americ[a].ana_N; **-iana**: 1 Mozart.iana_N
10. **-ance**: 1 import[ant].ance_N, brilli[ant].ance_N | 2 assist.ance_N; **-ancy**: 1 pregn[ant].ancy_N, brilli[ant].ancy_N; see also *-cy*; see also *-ence*
11. *-ancy* see *-ance*
12. **-ant**: 1 contest.ant_N, particip[ate].ant_N | 2 expect.ant_{ADJ}; see also *-ent*
13. **-ar**: 1 pol[e].ar_{ADJ} | 2 see *-er*
14. **-ard**: 1 drunk.ard_N
15. **-arian**: 1 parliament.arian_N, humanit[y].arian_{ADJ/N}
16. **-arium**: 1 planet.arium_N
17. **-ary**: 1 moment.ary_{ADJ} | 2 revolution.ary_{ADJ/N} | 3 dispens[e].ary_N
18. **-aster**: 1 critic.aster_N
19. **-ate**: 1 hyphen.ate_V | 2 passion.ate_{ADJ} | 3 doctor.ate_N | 4 elector.ate_N | 5 emir.ate_N
20. **-ation**: 1 expect.ation_N, uni[fy]fic.ation_N, demonstr[ate].ation_N; see also *-ion*
21. **-ative**: 1 talk.ative_{ADJ}, connot[e].ative_{ADJ} / connot[e].ative_{ADJ}; see also *-ive*
22. **-atory**: 1 commend.atory_{ADJ} | 2 observ[e].atory_N; see also *-ory*
23. **-cy**: 1 bankrupt.cy_N, accura[te].cy_N | 2 baronet.cy_N | 3 infan[t].cy_N, pira[te].cy_N; see also *-ancy*; see also *-ency*
24. **-dom**: 1 star.dom_N, free.dom_N | 2 earl.dom_N | 3 king.dom_N | 4 official.dom_N
25. *-ean* see *-an*
26. **-ed**: 1 excit[e].ed_{ADJ} | 2 talent.ed_{ADJ}, kind-heart.ed_{ADJ}
27. **-ee**: 1 employ.ee_N, pay.ee_N | 2 escap[e].ee_N | 3 absent.ee_N
28. **-eer**: 1 profit.eer_N | 2 mountain.eer_N
29. **-eme**: 1 morph.eme_N
30. **-en**: 1 wid[e].en_V | 2 gold.en_{ADJ}
31. **-ence**: 1 differ[ent].ence_N, consist[ent].ence_N | 2 interfer[e].ence_N; **-ency**: 1 effici[ent].ency_N, consist[ent].ency_N; see also *-cy*; see also *-ance*
32. *-ency* see *-ence*

33. **-ent**: 1 resid[e].ent_{N/ADJ} | 2 depend.ent_{ADJ/N}; see also *-ant*
34. *-eous* see *-ous*
35. **-er**: 1 teach.er_N, comput[e].er_N | 2 hat+t.er_N | 3 London.er_N | 4 double-deck.er_N | 5 rug[by]+g.er_N | 6 disclaim.er_N | 7 din[e].er_N; **-ar**: 1 li[e].ar_N; **-ier**: 1 cash.ier_N, cloth.ier_N; **-yer**: 1 law.yer_N; see also *-or*
36. **-erel**: 1 cock.erel_N
37. **-ern**: 1 north.ern_{ADJ}
38. **-eroo**: 1 switch.eroo_N
39. **-ers**: 1 preg[nant]+g.ers_{ADJ}, Rod[ney]+d.ers_N
40. **-ery**: 1 rob+b.ery_N | 2 cook.ery_N | 3 brew.ery_N | 4 slav[e].ery_N | 5 brav[e].ery_N, tomfool.ery_N | 6 machin[e].ery_N; **-ry**: 1 dentist.ry_N | 2 rival.ry_N | 3 savage.ry_N | 4 gadget.ry_N
41. **-ese**: 1 Japan.ese_{ADJ/N} | 2 journal.ese_N
42. **-esque**: 1 Byron.esque_{ADJ}, pictur[e].esque_{ADJ}
43. **-ess**: 1 count.ess_N / count.ess_N, lion.ess_N / lion.ess_N
44. *-et* see *-ette*
45. **-ette**: 1 disk.ette_N | 2 leather.ette_N | 3 usher.ette_N, Paul.ette_N; **-et**: 1 baron.et_N / baron.et_N, falcon.et_N
46. *-ey* see *-y*
47. **-fold**: 1 three.fold_{ADJ/ADV} | 2 three.fold_{ADV}
48. **-ful**: 1 care.ful_{ADJ} | 2 forget.ful_{ADJ} | 3 spoon.ful_N
49. **-hood**: 1 child.hood_N | 2 brother.hood_N | 3 false.hood_N
50. **-i**: 1 Bengal.i_{ADJ/N}
51. *-ial* see *-al*
52. *-ian* see *-an*
53. *-iana* see *-ana*
54. **-ible**: 1 digest.ible_{ADJ}; see also *-able*
55. **-ic**: 1 atom.ic_{ADJ}, optimist.ic_{ADJ}, histor[y].ic_{ADJ} | 2 Milton.ic_{ADJ} | 3 Celt.ic_{ADJ/N}
56. **-ical**: 1 nonsens[e].ical_{ADJ}, histor[y].ical_{ADJ}; see also *-al*
57. **-ician**: 1 diet.ician_N, beaut[y].ician_N; see also *-ian* 3
58. **-ics**: 1 linguist[ic].ics_N | 2 athlet[ic].ics_N
59. *-ie* see *-y*
60. *-ier* see *-er*
61. **-ify**: 1 solid.ify_V, simpl[e].ify_V, beaut[y].ify_V | 2 speech.ify_V
62. **-ile**: 1 protract.ile_{ADJ} | 2 infant.ile_{ADJ}
63. **-ine**: 1 hero.ine_N, Bernard.ine_N
64. **-ing**: 1 charm.ing_{ADJ}, prevail.ing_{ADJ} | 2 read.ing_N, gather.ing_N | 3 bank.ing_N | 4 curtain.ing_N
65. **-ion**: 1 rebel+l.ion_N, connect.ion_N; **-ition**: 1 add.ition_N, compet[e].ition_N; **-sion**: 1 deci[de].sion_N; **-xion**: 1 conne[ct].xion_N; see also *-ation*
66. *-ious* see *-ous*
67. *-ize* see *-ize*
68. **-ish**: 1 child.ish_{ADJ} | 2 yellow.ish_{ADJ} | 3 fifty.ish_{NUM} | 4 Finn.ish_{ADJ/N}
69. **-ism**: 1 cognitiv[e].ism_N, ideal.ism_N, Darwin.ism_N | 2 sex.ism_N | 3 hero.ism_N | 4 critic.ism_N | 5 colloqual.ism_N | 6 alcohol.ism_N

70. **-ist**: 1 *cognitiv[e].ist*_{N/ADJ}, *ideal.ist*_{N/ADJ}, *Darwin.ist*_{N/ADJ} | 2 *sex.ist*_{N/ADJ} | 3 *novel.ist*_N | 4 *semantic.ist*_N, *zoolog[y].ist*_N | 5 *violin.ist*_N | 6 *cycl[e].ist*_N, *plagiar[ize].ist*_N
71. **-ista**: 1 *fashion.ista*_N, *Blair.ista*_N
72. **-ite**: 1 *Thatcher.ite*_{N/ADJ} | 2 *Brooklyn.ite*_N
73. **-ition** see *-ion*
74. **-itis**: 1 *tonsil+1.itis*_N | 2 *television.itis*_N
75. **-ity**: 1 *stupid.ity*_N, *accept[able]abil.ity*_N, *vis[ible]ibil.ity*_N
76. **-ive**: 1 *attract.ive*_{ADJ}, *creat[e].ive*_{ADJ}; see also *-ative*
77. **-ize**: 1 *modern.ize*_V | 2 *critic.ize*_V; **-ise**: 1 *modern.ise*_V | 2 *critic.ise*_V
78. **-kin**: 1 *lamb.kin*_N
79. **-less**: 1 *taste.less*_{ADJ} | 2 *count.less*_{ADJ}
80. **-let**: 1 *book.let*_N, *pig.let*_N | 2 *neck.let*_N
81. **-ling**: 1 *duck.ling*_N | 2 *prince.ling*_N, *weak.ling*_N
82. **-ly**: 1 *slow.ly*_{ADV}, *gent[le].ly*_{ADV} | 2 *mother.ly*_{ADJ} | 3 *hour.ly*_{ADJ/ADV} | 4 *poor.ly*_{ADJ}
83. **-manship**: 1 *brink(s).manship*_N, *games.manship*_N; see also *-ship*
84. **-ment**: 1 *develop.ment*_N
85. **-ness**: 1 *kind.ness*_N, *happ[y].ness*_N
86. **-nik**: 1 *peace.nik*_N
87. **-o**: 1 *comb[ination].o*_N | 2 *cheap.o*_N
88. **-ock**: 1 *hill.ock*_N
89. **-oid**: 1 *human.oid*_{ADJ/N}
90. **-or**: 1 *invent.or*_N, *generat[e].or*_N; see also *-er*
91. **-ory**: 1 *advis[e].ory*_{ADJ}, *migrat[e].ory*_{ADJ} | 2 *deposit.ory*_N; see also *-atory*
92. **-ous**: 1 *poison.ous*_{ADJ}; **-eous**: 1 *gas.eous*_{ADJ}; **-ious**: 1 *uproar.ious*_{ADJ}; **-uous**: 1 *contempt.uous*_{ADJ}; see also *-aceous*
93. **-ry** see *-ery*
94. **-s**: 1 *banana.s*_{ADJ}, *Deb[orah].s*_N
95. **-ship**: 1 *professor.ship*_N | 2 *Lord.ship*_N | 3 *musician.ship*_N, *statesman.ship*_N; see also *-manship* | 4 *friend.ship*_N | 5 *reader.ship*_N | 6 *hard.ship*_N
96. **-sion** see *-ion*
97. **-some**: 1 *quarrel.some*_{ADJ} | 2 *three.some*_N
98. **-ster**: 1 *young.ster*_N | 2 *trick.ster*_N
99. **-sy**: 1 *cute.sy*_{ADJ} | 2 *Pat.sy*_N
100. **-th**: 1 *warm.th*_N | 2 *grow.th*_N
101. **-ton**: 1 *simple.ton*_N
102. **-trix**: 1 *avia[tor].trix*_N
103. **-ty**: 1 *certain.ty*_N
104. **-ual** see *-al*
105. **-uous** see *-ous*
106. **-ure**: 1 *fail.ure*_N | 2 *prefect.ure*_N
107. **-ward**: 1 *east.ward*_{ADV/ADJ}, *wind.ward*_{ADV/ADJ}; **-wards**: 1 *east.wards*_{ADV}
108. **-wards** see *-ward*
109. **-ways** see *-wise*
110. **-wise**: 1 *clock.wise*_{ADV/ADJ} | 2 *length.wise*_{ADV/ADJ} | 3 *weather-wise*_{ADV}; **-ways**: 1 *length.ways*_{ADV/ADJ}
111. **-xion** see *-ion*

112. **-y**: 1 milk.y_{ADJ} | 2 green.y_{ADJ} | 3 stick.y_{ADJ} | 4 dog+g.y_N, night[dress].y_N, Tom+m.y_N |
5 tough.y_N, town.y_N | 6 modest.y_N | 7 expir[e].y_N; **-ey**: 1 clay.ey_{ADJ}, dic[e].ey_{ADJ} |
2 Charl[es].ey_N; **-ie**: 1 dog+g.ie_N, night[dress].ie_N, dear.ie_N, Sus[an].ie_N,
Charl[es].ie_N | 2 tough.ie_N, town.ie_N
113. *-yer* see *-er*

New-line and Run-on Guiding Devices in Print Monolingual Dictionaries for Learners of English

Bartosz Ptasznik, *University of Warmia and Mazury, Olsztyn, Poland*
(bartosz.ptasznik@uwm.edu.pl)

and

Robert Lew, *Faculty of English, Department of Lexicography and Lexicology,
Adam Mickiewicz University, Poznań, Poland* (rlew@amu.edu.pl)

Abstract: The paper focuses on guiding devices in print monolingual dictionaries for learners of English. It aims to find answers to six research questions. The main aim is to investigate how the location of guiding devices within entries — starting from a new line versus run-on — affects consultation time and sense selection accuracy of dictionary users in entry navigation. In addition, the paper looks at the effect that part of speech (noun versus verb entries) has on consultation time and sense selection accuracy; further, the relationship between entry length and consultation time is investigated, as well as that between consultation time and sense selection accuracy.

Keywords: DICTIONARY, DICTIONARY CONSULTATION, DICTIONARY ACCESS, LEXICOGRAPHY, GUIDING DEVICES, SIGNPOSTS, SIGNPOSTING, MONOLINGUAL DICTIONARIES, LEARNERS' DICTIONARIES, ENGLISH, POLISH LEARNERS

Opsomming: Nuwereël- en deurloop-hulpmiddels vir gebruiksheiding in gedrukte eentalige woordeboeke vir Engelse leerdere. In hierdie artikel word gefokus op gebruiksheidingsinstrumente in gedrukte eentalige woordeboeke vir Engelse leerdere. Dit het die beantwoording van ses navorsingsvrae ten doel. Die hoofdoel daarvan is om te ondersoek hoe die plasing van gebruiksheidingsinstrumente in inskrywings — beginnende by 'n nuwe reël versus deurloop — die woordeboekgebruiker se konsultasietyd en akkuraatheid van betekenis-seleksie in inskrywingsnavigasie beïnvloed. Daarbenewens word daar ook in hierdie artikel gekyk na die effek wat woordsoort (selfstandige naamwoord- versus werkwoordinskrywings) op die konsultasietyd en akkuraatheid van betekenis-seleksie het. Vervolgens word die verband tussen die lengte van die inskrywing en die konsultasietyd ondersoek, asook die verband tussen konsultasietyd en akkuraatheid van betekenis-seleksie.

Sleutelwoorde: WOORDEBOEK, WOORDEBOEKKONSULTASIE, WOORDEBOEKTOEGANG, LEKSIKOGRAFIE, GEBRUIKHEIDINGSINSTRUMENTE, AANWYSERS, AANWYSING, EENTALIGE WOORDEBOEKE, AANLEERDERSWOORDEBOEKE, ENGELS, POOLSE LEERDERS

1. Background

1.1 Research on guiding devices in English monolingual learners' dictionaries

Empirical studies that deal with guiding devices in English monolingual learners' dictionaries are becoming increasingly common. One motivation for this is the pursuit of empirically motivated principles to inform the design of dictionary entries so that they satisfy the needs of dictionary users who regularly consult a dictionary, and who in the process of doing so repeatedly have to browse through lengthy entries that take up much of their time. Although the gradual transition to digital dictionaries makes this issue less relevant, in many parts of the world, print dictionaries are still far from disappearing. Thus far, two broad types of guiding devices have been used in English learners' dictionaries: (1) signposts, which briefly define their respective senses, with each signpost being placed next to its sense; and (2) menus, which form a list of sense cues appearing as a block at the top of the entry.

Signpost-related research goes back to Tono's (1992) study and his finding that less proficient English learners benefit from menus in entries to a greater extent than do more proficient learners. In another study (1997), Tono found that LDOCE3 signposts are superior to CIDE guide words, especially with regard to the accuracy of sense selection and time needed for consultation; however, not necessarily so in the case of longer entries. Further, LDOCE3 guiding devices appeared to convey semantically more meaningful information. One year later, Bogaards (1998) concluded that semantically-guided LDOCE3 and CIDE sense-navigation devices generally outperformed grammar-based COBUILD2 access devices, as well as the access devices adopted by OALD5 lexicographers. Interestingly, the subjects who were assisted by OALD5 guiding devices performed the worst, despite having been exposed to the shortest entries.

More recent research includes a series of studies carried out by Lew. Lew and Pajkowska (2007) stressed the need to conduct further research on meaning access devices in learners' dictionaries, and noted in their research that pre-intermediate English learners may benefit from signposts more when entries were of shorter length, whereas intermediate learners may benefit more from signposts in longer entries. Lew and Tokarek (2010) focused their attention on entry menus in web-based bilingual dictionaries and observed that by and large menus with highlighting of target senses were both more effective (accurate) and efficient (faster) than menus which appeared without highlighting. In the same year, Lew (2010) compared the signpost and menu systems (based on OALD7 entries) and found that signposts resulted in superior sense selection and translation accuracy compared with menus. One explanation offered for this apparent advantage of signposts was the location of the information within entries. Signposts appear next to their respective senses, which might make it

easier for dictionary users to consult the signpost along with the related definition and examples that appeared next to it. By contrast, in menu-based systems, dictionary users need to go over the list of sense cues located above the entry, and only then are they able to proceed to the actual visual scan of the senses. Thus, in the case of menus, the guiding information appears out of the immediate context of the respective senses: this could lead to dictionary users not being able to reach the type of information they are searching for in the senses, and might leave users confused between the menu and the actual senses.

Tono (2011) applied the eye-tracking technique to observe patterns of dictionary use and test the effectiveness of the signposting and menu systems. One finding was that, depending on how signposts were phrased, they could either be helpful or misleading to dictionary users. Furthermore, in view of earlier findings (Tono 1992), Tono expected that menus would be of greater benefit to less advanced English students. In the eye-tracking study, proficient students tended to ignore menus, but at the same time consulted signposts willingly. Less proficient students, however, did not use the signposts in entries, possibly because they had no clue as to what the purpose of the signposts was.

A study by Nesi and Tan (2011) confirmed Lew's (2010) finding of the superiority of signposts over menus. It also found that learners were able to select senses more accurately and quickly when the target senses were located at the beginning or towards the end of entries, while the middle of entries was the most problematic. Unexpectedly, sense selection success and consultation times were best for the target senses appearing at the end of entries. Nesi and Tan attributed these results to the so-called bathtub effect, which is a known effect of the early and final parts of words being easier to remember (Aitchison 1997); in the context of entry consultation, this would translate into facilitation of initial and final senses in entries. In terms of the advantage of final senses over initial senses, one explanation could be advanced learners starting their entry search with the final senses, rather than in a top-to-bottom fashion. The authors speculated that such a strategy may have arisen from the discovery through regular dictionary consultation that the senses placed at the beginning of entries tend to be familiar to advanced users, due to the fact that in many dictionaries they tend to be the frequent senses. Nesi and Tan (2011) also found that entry or definition length have no effect on accuracy of sense selection. Likewise, consultation time remained unaffected by entry length, while definition length was positively related to consultation time. An additional general finding was that adjective and verb entries were more problematic than noun entries for dictionary users. To some extent, this finding was confirmed by Ptasznik's research (2015), according to which consultation of verb entries took more time than consultation of noun entries.

Ptasznik and Lew (2014) compared entries equipped with a combination of signposts and menus against entries with signposts only. The study found that adding entry-initial menus to entries with signposts did not improve the

rate of correct sense selection, nor did it reduce the time needed for entry consultation. Hence, Ptasznik and Lew concluded that it is best when entries in paper dictionaries are equipped with signposts only.

The two most recent experimental studies of signposting are Dziemianko (2016) and Dziemianko (2017). The former study undertook an empirical comparison of the alternative methods of presenting signposts featured in three online dictionaries: LDOCE5 (white capitals on a blue background), OALD8 (crimson capitals above a crimson line), and OALD9 (black lower-case letters above a dark orange line). The LDOCE5 signpost highlighting strategy was found to be the most beneficial of all three strategies with regard to consultation time, while the OALD8 strategy was the least helpful. As far as accuracy of sense selection is concerned, all three methods achieved comparable scores, and so none of the signpost highlighting strategies came out as most beneficial. The study also tested meaning retention; for this outcome measure, white capitals on a blue background and lower-case letters above a dark orange line resulted in highest retention rates. This study also touched upon sense positioning in entries, with the finding consonant with Nesi and Tan (2011) that dictionary users consulted entry-final senses the fastest. However, signposts that appeared in crimson capitals above a crimson line and black lower-case letters above a dark orange line did not exhibit a significant effect of sense position on the time of consultation. The findings were not compatible with the interpretation that proficient learners would scan entries from the bottom up, as shorter times for entry-final senses were only recorded for senses equipped with signposts in the form of white capitals on a blue background. In addition, accuracy of sense selection was highest for senses located in the middle of entries, which contradicts Tono's (1992) finding of initial senses being the best in this respect. All three methods of signpost highlighting produced more or less similar sense selection success rates in all types of sense positions, which does not support Nesi and Tan's (2011) claim that initial and final senses lead to most successful results. Dziemianko (2016) concurs with Ptasznik (2015) in warning against simplistically assuming the advantage of homogenous form of signposts throughout the entry and stresses the importance of conducting further research on heterogeneous and homogenous sense-navigation devices, as well as the length of sense indicators.

Dziemianko's most recent study (Dziemianko 2017) did not confirm Nesi and Tan's (2011) finding regarding the bathtub effect. Dziemianko found no evidence of a bathtub effect on either sense selection, reception, or production. Furthermore, selection of senses, as well as decoding and encoding were not affected by the length of entries. On the other hand, the study found a tendency for participants to identify longer target senses with greater success.

1.2 Which dictionaries have new-line and run-on guiding devices?

By and large, guiding devices that start from a new line¹ have dominated in

print monolingual English learners' dictionaries (see **Table 1**). Presumably, the thinking was that starting senses on a new line would make it easier for English learners to distinguish one sense of an entry from another. The only English monolingual learners' dictionary that has departed from this general strategy in some of its editions was the *Cambridge Advanced Learner's Dictionary* [CALD]. In general, CALD3 entries have been equipped with run-on guidewords (guidewords are the specific brand of guiding devices that appear in CALD3), though in the most highly polysemous entries (i.e. those with the greatest number of senses) new-line guidewords are employed, which may suggest that CALD3 lexicographers agreed with the view of the competing dictionaries that adding a line break might be advantageous, but compromised in the interest of saving space. The next edition, however (CALD4), only included run-on guidewords, irrespective of entry length. **Table 1** gives the solutions adopted in the specific dictionary editions with respect to the presentation of senses from a new line or as run-on.

Dictionary edition	Year of publication	Guiding device position
<i>Longman Dictionary of Contemporary English (LDOCE3)</i>	1995	New-line
<i>Longman Dictionary of Contemporary English (LDOCE4)</i>	2003	New-line
<i>Longman Dictionary of Contemporary English (LDOCE5)</i>	2009	New-line
<i>Longman Dictionary of Contemporary English (LDOCE6)</i>	2014	New-line
<i>Cambridge International Dictionary of English (CIDE)</i>	1995	New-line
<i>Cambridge Advanced Learner's Dictionary (CALD1)</i>	2003	New-line
<i>Cambridge Advanced Learner's Dictionary (CALD2)</i>	2005	New-line
<i>Cambridge Advanced Learner's Dictionary (CALD3)</i>	2008	Run-on ²
<i>Cambridge Advanced Learner's Dictionary (CALD4)</i>	2013	Run-on
<i>Oxford Advanced Learner's Dictionary of Current English (OALD4)</i>	1989	New-line
<i>Oxford Advanced Learner's Dictionary of Current English (OALD5)</i>	1995	New-line
<i>Oxford Advanced Learner's Dictionary of Current English (OALD6)</i>	2000	New-line
<i>Oxford Advanced Learner's Dictionary of Current English (OALD7)</i>	2005	New-line
<i>Oxford Advanced Learner's Dictionary of Current English (OALD8)</i>	2010	New-line
<i>Oxford Advanced Learner's Dictionary of Current English (OALD9)</i>	2015	New-line
<i>Macmillan English Dictionary for Advanced Learners (MED1)</i>	2002	New-line
<i>Macmillan English Dictionary for Advanced Learners (MED2)</i>	2007	New-line

Table 1: New-line and run-on guiding devices in specific editions of print monolingual English learners' dictionaries

Although new-line guiding devices seem to be the dominant practice for print monolingual English learners' dictionaries (**Table 1**), we believe it is still impor-

tant to test if it is indeed the better way to structure senses when these start with signposts. Two outcomes are measured: consultation time and sense selection accuracy. The next section outlines the aims and methods, followed by results and discussion.

3. The study

3.1 Aims of the study

The aims of the present study were to investigate:

- (1) how CONSULTATION TIME is affected by GUIDING DEVICE POSITION (new-line guiding devices vs. run-on guiding devices) during entry navigation; we would expect that positioning the guiding devices from a new line should facilitate visual search and thus reduce the time needed to find the relevant sense
- (2) how CONSULTATION TIME is affected by PART OF SPEECH³ (nouns vs. verbs); we would expect verbs to be more challenging than nouns, the latter being *first-order words* (Piotrowski 1989: 102);
- (3) whether GUIDING DEVICE POSITION is a factor determining SENSE SELECTION ACCURACY in entry look-ups; one might expect that the new-line format would potentially allow the user to avoid more errors due to its less cluttered, unambiguous layout;
- (4) whether SENSE SELECTION ACCURACY varies by PART OF SPEECH; again, nouns might be expected to trigger fewer errors, due to their semantic primacy;
- (5) if CONSULTATION TIME depends on ENTRY LENGTH measured in words; we would expect longer entries to require longer consultation, thus a positive correlation between the variables would be expected;
- (6) whether SENSE SELECTION ACCURACY of lexical items is related to CONSULTATION TIME; we would expect a negative correlation, with longer times corresponding to lower accuracy; both being expected consequences of increased item difficulty.

3.2 Method, participants, and procedure

The study was carried out at the Faculty of Humanities at the University of Warmia and Mazury in Olsztyn. 100 students of English who were third and fourth year students participated in the study. Their English language proficiency level was assessed by their academic teachers as B2 to C1 by the Com-

mon European Framework of Reference for Languages standards; Polish was their native language.

Four variables were selected for the study, two predictor variables: GUIDING DEVICE POSITION and PART OF SPEECH; and two outcome variables: CONSULTATION TIME and SENSE SELECTION ACCURACY. GUIDING DEVICE POSITION had two levels: guiding devices that started on a new line (new-line guiding devices) and devices that continued within the entry without a line break (run-on guiding devices). The number of senses was not manipulated in this study, but was controlled and set at the reasonable number of seven senses. All entries selected for the study consisted of at least seven senses; if there were more, some irrelevant senses were removed, so that the final number of senses was always seven. Half of the entries were nouns and half were verbs. All subjects were exposed to the same target items (each task was the same for each subject, items had the same cue sentences, they were of the same part of speech and the lexicographic data were the same in all entries), but specific entries appeared with either new-line or run-on guiding devices. The order of presentation of items was randomized to one of two versions of the test, the versions then being assigned randomly (the effect of version was checked and it was not significant). CONSULTATION TIME and SENSE SELECTION ACCURACY were recorded for each subject and test item.

A pilot test was carried out on 10 subjects to see if the whole procedure worked well. The subjects had enough time for each task during their class hours and by and large the subjects achieved a SENSE SELECTION ACCURACY within the range of about 55–80%. None of the test items exhibited a floor or ceiling effect, and there were no other problems, and so the main study was conducted after the pilot testing phase.

Twenty test items were used in the main study: *bear, introduction, jump, hole, match, mark, slip, power, lift, weight, withdraw, wave, transfer, unit, beat, base, print, grip, cast, root*. The lexicographic data of the entries were taken from LDOCE6, while the cue sentences were taken from four dictionaries: OALD9, the free online version of LDOCE [LDOCEO], and — sporadically — from the *Macmillan English Dictionary Online* [MEDO] and *Cambridge Dictionary Online* [CDO]. Each single test item (from among the twenty test items mentioned above) that was chosen for the study needed to have at least seven senses with guiding devices and had to include at least one less common sense that the subjects would not be likely to be familiar with. These less common senses of words were the target senses in the cue sentences of the study, as the subjects had to be exposed to tasks that would depend on entry consultation for successful completion.

The test was administered in a classroom by the first author, in five groups of between fifteen and twenty students in each session, over the course of one week. Before the actual test, each group received step-by-step instructions on how to proceed. The instruction was delivered verbally from a printed script in Polish, the native language of all participants. The subjects were told

in their native language that they were expected to read the cue sentence. The cue sentences provided the subjects with some context in the target language and an underlined word (the target item), which appeared in a less-known sense (see Appendix A and B). Subsequently, they were instructed to identify the target item and its context, search for the meaning of the underlined target word (the target items in the cue sentences were provided with a dictionary entry underneath the cue sentences: see Appendix A and B), write down on the test sheet the number of the target sense in which the target item is used, record their own time and move on to the next task item. The participants were informed that each test had twenty test items on separate sheets and that there were no time restrictions for the completion of the test. Participants recorded their own time using the stopwatch function of their mobile phones (it would not be possible for the experimenter to record twenty participants at the same time, and the school only allowed testing in groups). The next step was a complete practice run on two items (these were not used in the data analysis), followed by the actual testing.

All participants were exposed to the same test items. New-line and run-on guiding devices were rotated across test versions. In any single test, half of the test items had new-line guiding devices, and the other half had run-on guiding devices. Also, in any single test, half of the test items were noun entries, and the other half were verb entries. Across the 100 participants, the same number of responses were recorded for new-line and run-on guiding devices, as well as noun and verb entries. CONSULTATION TIME and SENSE SELECTION ACCURACY were recorded for each subject and test item (that is 2,000 data points for each variable).

4. Results and discussion

4.1 Consultation time

Consultation time is markedly non-normal and skewed to the right (skew = 1.94, kurtosis = 5.1), which — as is usual for temporal variables such as human reaction time or time-on-task — normalizes well when the time values are expressed as natural log values. Our data is no exception: after logarithmizing the times, the distribution becomes symmetrical and nearly normal (skew = -0.16, kurtosis = 0.0). This is also evident in the normal Quantile–Quantile plots (see **Figure 1**). Therefore, we will be operating in logged values, but for the readers' convenience converting them back to raw time expressed in seconds. Likewise, average times will be calculated as mean logged values, which is equivalent to geometric means on the original scale.

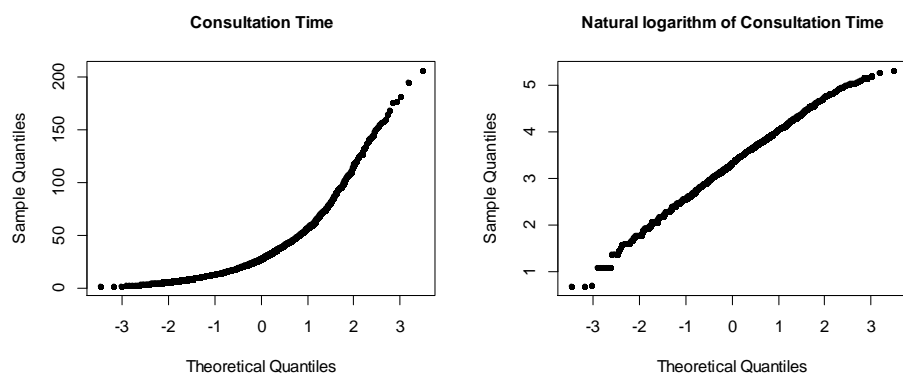


Figure 1: Normal Quantile–Quantile plots for CONSULTATION TIME: raw (panel on left) and logged values (panel on right)

The grand (geometric) mean for CONSULTATION TIME, averaged across all entry look-ups, was 27.7 seconds. When broken down by part of speech and guiding device position, the (geometric) means pattern as in **Table 2**.

	new-line	run-on	any position
Noun	23.1	25.5	24.3
Verb	32.8	30.6	31.7
any POS	27.5	27.9	27.7

Table 2: Average CONSULTATION TIME for combinations of GUIDING DEVICE POSITION and PART OF SPEECH (geometric mean given in seconds)

The marginal values in **Table 2** suggest a difference due to PART OF SPEECH, with verb entries being, on average, some thirty percent slower than noun entries. By contrast, the position of the guiding device does not seem to matter at all, since the CONSULTATION TIMES for the new-line and run-on conditions are virtually identical. This impression is confirmed in the so-called pirate plots (**Figure 2**) generated with the help of the `yarr` package (Phillips 2017). A pirate plot is an advanced plot that visually conveys information on (1) the central tendency (in our case, the geometric mean); (2) the detailed distribution of raw data through a jittered plot of all the data points (here, grey dots); and (3) a probability density estimate (the "beans"). In addition, the narrow boxes around the mean bar represent inference bands, in this case computed as 95% Bayesian Highest Density Intervals. The subjects in the study needed on average 32 seconds for the consultation of verb entries and about eight seconds less for noun entries (24 seconds).

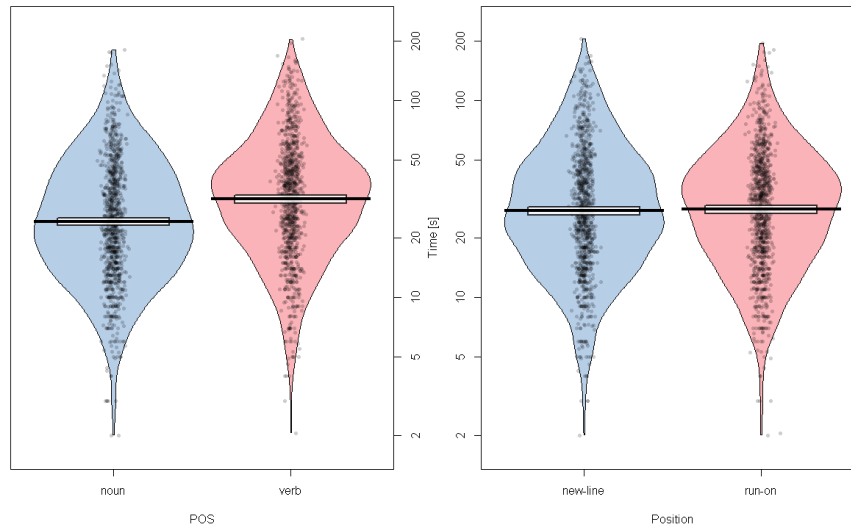


Figure 2: Pirate plots of CONSULTATION TIME by PART OF SPEECH (left) and GUIDING DEVICE POSITION (right). Inference bands around the means are 95% Bayesian Highest Density Intervals.

These results suggest that PART OF SPEECH affects CONSULTATION TIME in entry navigation, with verb entries requiring on average eight seconds more consultation time, which is about a third of the time more compared to noun entries. This may mean that verb entries tend to be more problematic for dictionary users than noun entries.

4.2 Best model for CONSULTATION TIME

A series of linear mixed models were fitted using `lme4::lmer` (Bates et al. 2015) and `afex::mixed` (Singmann et al. 2018) for logarithmized CONSULTATION TIME as the outcome variable, starting with complete models with interactions. By both BIC and AIC criteria, the best model included only PART OF SPEECH as a fixed effect (though not either LENGTH or GUIDING DEVICE POSITION), as well as random intercepts for SUBJECT and ITEM ($\log.\text{Time} \sim \text{POS} + (1|\text{Subject}) + (1|\text{Item})$). In this model, residuals had an approximately normal distribution (see **Figure 3**). Part of speech was marginally significant ($F_{(18,1)} = 4.26, p = 0.05$) using the Kenward-Roger approximation (Judd et al. 2012; the same method was used for computing p-levels in subsequent analyses). This confirms the significance of the effect of PART OF SPEECH on CONSULTATION TIME, with verbs taking on average more time to consult than nouns by about one third.

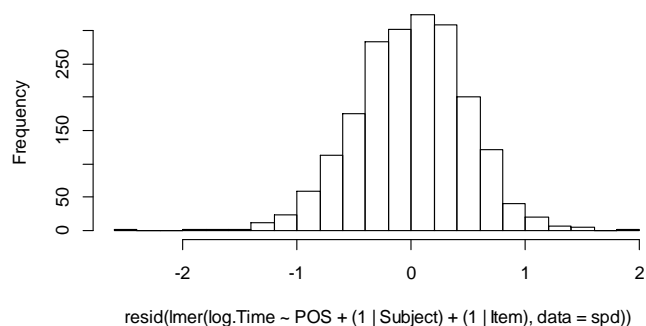


Figure 3: Histogram of residuals for the best model $\log.\text{Time} \sim \text{POS} + (1 | \text{Subject}) + (1 | \text{Item})$. The distribution of residuals is symmetrical and approximately normal.

4.3 Selection accuracy

In terms of SENSE SELECTION ACCURACY, the grand mean was 62 percent, and it was exactly the same in the case of run-on and new-line entries. Nouns exhibited a somewhat higher mean accuracy (66 percent) than verbs (58 percent). Broken down by PART OF SPEECH and GUIDING DEVICE POSITION, the means patterned as in **Table 3**.

	new-line	run-on	any position
Noun	67%	66%	66%
Verb	57%	59%	58%
any POS	62%	62%	62%

Table 3: Mean SENSE SELECTION ACCURACY for combinations of GUIDING DEVICE POSITION and PART OF SPEECH (in percent)

A series of binary logistic models with random factors was fitted with SENSE SELECTION ACCURACY as the outcome variable. The best model was the intercept-only model with random intercepts for SUBJECT and ITEM ($\text{Correct.sense} \sim 1 + (1 | \text{Subject}) + (1 | \text{Item})$). The best model that was not intercept-only was the model that included PART OF SPEECH as a predictor, with nearly the same *Akaike Information Criterion* value ($AIC = 2416.7$) as the intercept-only model ($AIC = 2416.4$). This may be taken to interpret that PART OF SPEECH may play some minor role in determining sense selection accuracy (with nouns yielding

more success on average than verbs). By contrast, GUIDING DEVICE POSITION appears not to matter for the accuracy of sense selection within entries.

4.4 Relationship between entry length and consultation time

In **Figure 4**, we plot the relationship between ENTRY LENGTH and CONSULTATION TIME for the individual test items (for CONSULTATION TIME, these are unlogged fitted values). The slope is not significantly different from zero ($p = 0.15$, n.s.), which suggests CONSULTATION TIME is not related to ENTRY LENGTH; this may be seen as somewhat surprising: a longer entry might be expected to require longer study time due to its sheer length. In addition, a longer entry could also mean that the lexical item itself might be a more challenging one, and that might again be expected to require longer consultation. No such effect, however, is evident in our data, and there is in fact a (non-significant) hint of a reverse trend.

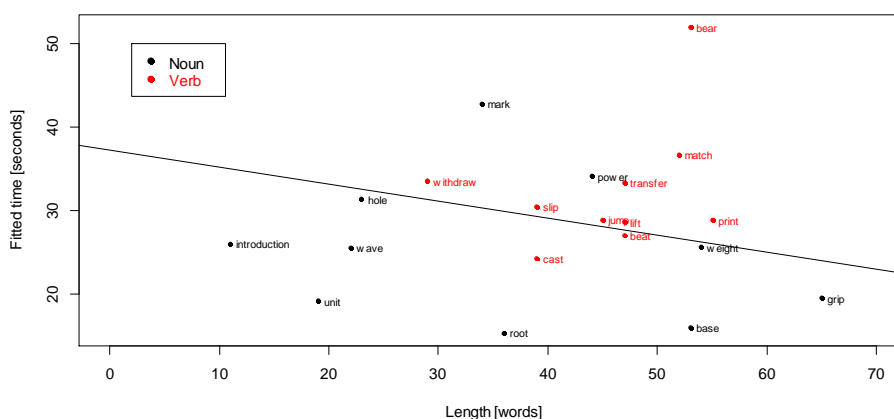


Figure 4: Plot of CONSULTATION TIME (fitted values) and ENTRY LENGTH for the individual items (nouns are shown in black, verbs in red font)

4.5 Relationship between selection accuracy and consultation time

The final question refers to the relationship between SELECTION ACCURACY and CONSULTATION TIME for our test items. We might expect that the relatively easy items would be dealt with relatively more quickly and with better success rates than the more challenging items. Therefore, we would expect a negative relationship between success rates for SELECTION ACCURACY and CONSULTATION TIME. In **Figure 5**, we plot the relationship between these two variables fitted in a linear mixed model with CONSULTATION TIME as outcome, SELECTION ACCURACY

as predictor and random intercepts for subject and item (this turned out to be the best model: POS was only marginally significant and was removed from the model, and the POS by SELECTION ACCURACY interaction was not significant). The slope of the best line of fit included in **Figure 5** is significantly different from zero ($p = 0.005$), and indicates a robust negative linear relationship between CONSULTATION TIME and SELECTION ACCURACY. Roughly speaking, an improvement of 0.1 in SELECTION ACCURACY corresponds to a decrease of 3.2 seconds in CONSULTATION TIME.

Looking at the plot, the noun item *base* has the best SELECTION ACCURACY, and is also the fastest; in contrast, the noun *mark* is the least successful and is the second slowest item, with the verb *bear* being the slowest of all.

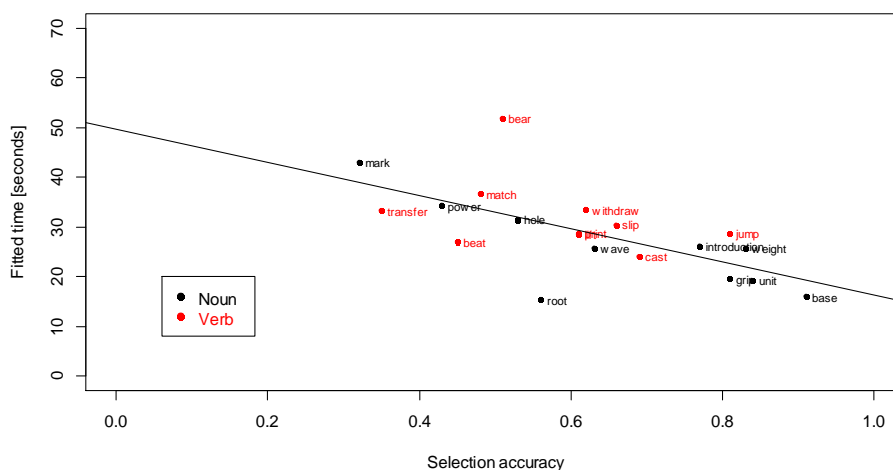


Figure 5: Plot of CONSULTATION TIME (unlogged fitted values) and SENSE SELECTION ACCURACY for the individual items (nouns are shown in black, verbs in red font)

5. Discussion and conclusion

The most important and practical research finding of this study is that no effect of GUIDING DEVICE POSITION on either CONSULTATION TIME or SENSE SELECTION ACCURACY was found. Although this is a null effect, the substantial sample size gives us some confidence that these hypothetical effects, even if present, are small in magnitude. In practical terms, this means that there is no evidence that adding new lines in an attempt to better signal new senses to the user adds further benefit, as the line breaks did not significantly improve either speed or success of consultation (they were actually the same). We might then tenta-

tively speculate that when entry senses are prefixed with signposts rendered in typography similar to that adopted in the present study (i.e. in framed capitals following sense numbers: see Appendix), the typography is salient enough to offer good sense discrimination in a visual search, and thus gets no further help from breaking the line for each sense. Since adding line breaks uses up extra space, then — based on the present findings — we would advise *against* their inclusion in dictionary entries, as long as sufficiently salient signposts are present, until real evidence is presented of any associated benefits. This finding is of direct relevance to lexicographic practice, since the variable manipulated is one that lexicographers can actually control in practical lexicographic work: they can decide to make entries with or without the extra line breaks. Our findings suggest that *without* might be good enough.

The remaining findings of the present study are interesting, though they cannot be directly translated into improved lexicographic practice. In terms of the part of speech of the entry, verb entries needed about thirty percent more time than noun entries, confirming findings from previous research (Nesi and Tan 2011; Ptasznik 2015) that verb entries tend to be more problematic than noun entries and that consulting verb entries takes more time. When it comes to the other outcome measure, accuracy of sense selection, nouns outperformed verbs by about fourteen percent. Although this difference did not turn out to be significant, with the intercept-only model offering the best explanation by the usual trade-off between model accuracy and parsimony, the direction of the difference again contributes to the overall tendency evident from other studies demonstrating verb entries to be more problematic, in general, than noun entries, and this can translate into differences in terms of both speed and success of dictionary consultation. This is no fault of lexicographers and nothing they can fix by improving dictionary entries, as the underlying cause lies in an inherent difference between nouns and verbs. Other things being equal, nouns are more fundamental to human experience (prototypically, they designate objects) than verbs (which prototypically represent actions). This insight was captured by Lyons (1977: 7.4, 8.1) and re-iterated in the lexicographic context by Piotrowski (1989: 102). The inherent relative difficulty of verbs presents a challenge to the definer: definitions of verbs tend to be longer than those of nouns, and, given the present and previous findings, presumably more challenging to process and comprehend. In this context, an interesting extension would be to test adjective entries in a similar manner and, even more interestingly, to see if the problem persists to the same extent if translation equivalents are used rather than same-language definitions.

As stated above, the more challenging lexical items typically require longer entries; however, our study did not find any indication of consultation time increasing with greater entry length. This finding invites further research into the phenomenon of information overload in the context of dictionary entries, as in our study there is no evidence that an entry of some 70 words in length is problematic in this regard. On the other hand, the two main outcome

variables in our study did correlate quite highly: consultation time was inversely related to the accuracy of sense selection, meaning that the longer the consultation, the lower the accuracy: a relationship that is to be expected.

This study has investigated entries in the print format. In digital dictionaries, sense navigation devices may work somewhat differently, and it may be of relevance to re-test some of the present research questions in digital dictionary environments.

Notes

1. New-line and run-on guiding devices are demonstrated in the appendices of the paper.
2. Only the most polysemous CALD3 entries have new-line guidewords.
3. Variables are written in small capital letters throughout this article (PART OF SPEECH, GUIDING DEVICE POSITION, CONSULTATION TIME, SENSE SELECTION ACCURACY).

References

Online dictionaries

- Cambridge Dictionary Online*. (<https://dictionary.cambridge.org/>). [CDO]
Longman Dictionary of Contemporary English Online. (<http://www.ldoceonline.com/>). [LDOCEO]
Macmillan English Dictionary Online. (<http://www.macmillandictionary.com>). [MEDO]

Print dictionaries

- Cowie, A.P. (Ed.)**. 1989. *Oxford Advanced Learner's Dictionary of Current English* (4th edition). Oxford: Oxford University Press. [OALD4]
Crowther, J. (Ed.). 1995. *Oxford Advanced Learner's Dictionary of Current English* (5th edition). Oxford: Oxford University Press. [OALD5]
Delacroix, L. (Ed.). 2014. *Longman Dictionary of Contemporary English* (6th edition). Harlow: Longman. [LDOCE6]
Deuter, M. (Ed.). 2015. *Oxford Advanced Learner's Dictionary of Current English* (9th edition). Oxford: Oxford University Press. [OALD9]
Gillard, P. (Ed.). 2003. *Cambridge Advanced Learner's Dictionary* (1st edition). Cambridge: Cambridge University Press. [CALD1]
Mayor, M. (Ed.). 2009. *Longman Dictionary of Contemporary English* (5th edition). Harlow: Longman. [LDOCE5]
McIntosh, C. (Ed.). 2013. *Cambridge Advanced Learner's Dictionary* (4th edition). Cambridge: Cambridge University Press. [CALD4]
Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press. [CIDE]
Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners* (1st edition). Oxford: Macmillan Education. [MED1]

- Rundell, M. (Ed.)**. 2007. *Macmillan English Dictionary for Advanced Learners* (2nd edition). Oxford: Macmillan Education. [MED2]
- Summers, D. (Ed.)**. 1995. *Longman Dictionary of Contemporary English* (3rd edition). Harlow: Longman. [LDOCE3]
- Summers, D. (Ed.)**. 2003. *Longman Dictionary of Contemporary English* (4th edition). Harlow: Longman. [LDOCE4]
- Turnbull, J. (Ed.)**. 2010. *Oxford Advanced Learner's Dictionary of Current English* (8th edition). Oxford: Oxford University Press. [OALD8]
- Walter, E. (Ed.)**. 2005. *Cambridge Advanced Learner's Dictionary* (2nd edition). Cambridge: Cambridge University Press. [CALD2]
- Walter, E. (Ed.)**. 2008. *Cambridge Advanced Learner's Dictionary* (3rd edition). Cambridge: Cambridge University Press. [CALD3]
- Wehmeier, S. (Ed.)**. 2000. *Oxford Advanced Learner's Dictionary of Current English* (6th edition). Oxford: Oxford University Press. [OALD6]
- Wehmeier, S. (Ed.)**. 2005. *Oxford Advanced Learner's Dictionary of Current English* (7th edition). Oxford: Oxford University Press. [OALD7]

Other references

- Aitchison, Jean**. 1997. *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: Blackwell.
- Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker**. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1): 1-48.
- Bogaards, Paul**. 1998. Scanning Long Entries in Learner's Dictionaries. Fontenelle, Thierry, Philippe Hilgsmann, Archibal Michiels, André Moulin and Siegfried Theissen (Eds.). 1998. *EURALEX '98 Actes/Proceedings*: 555-563. Liège: English and Dutch Departments, University of Liège.
- Dziemianko, Anna**. 2016. An Insight into the Visual Presentation of Signposts in English Learners' Dictionaries Online. *International Journal of Lexicography* 29(4): 490-524.
- Dziemianko, Anna**. 2017. Dictionary Entries and Bathtubs: Does It Make Sense? *International Journal of Lexicography* 30(3): 263-284.
- Judd, Charles M., Jacob Westfall and David A. Kenny**. 2012. Treating Stimuli as a Random Factor in Social Psychology: A New and Comprehensive Solution to a Pervasive but Largely Ignored Problem. *Journal of Personality and Social Psychology* 103(1): 54-69.
- Lew, Robert**. 2010. Users Take Shortcuts: Navigating Dictionary Entries. Dykstra, Anne and Tanneke Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*: 1121-1132. Ljouwert: Afûk.
- Lew, Robert and Julita Pajkowska**. 2007. The Effect of Signposts on Access Speed and Lookup Task Success in Long and Short Entries. *Horizontes de Lingüística Aplicada* 6(2): 235-252.
- Lew, Robert and Patryk Tokarek**. 2010. Entry Menus in Bilingual Electronic Dictionaries. Granger, Sylviane and Magali Paquot (Eds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications*: 193-202. Louvain-la-Neuve: Cahiers du CENTAL.
- Lyons, John**. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Nesi, Hilary and Kim Hua Tan**. 2011. The Effect of Menus and Signposting on the Speed and Accuracy of Sense Selection. *International Journal of Lexicography* 24(1): 79-96.
- Phillips, Nathaniel**. 2017. *Yarrrr!: The Pirate's Guide to R* (R Package Version 0.1.5).

- Piotrowski, Tadeusz.** 1989. Monolingual and Bilingual Dictionaries: Fundamental Differences. Tickoo, Makhan L. (Ed.). 1989. *Learners' Dictionaries: State of the Art*: 72-83. Singapore: SEAMEO Regional Language Centre.
- Ptasznik, Bartosz.** 2015. *Signposts and Menus in Monolingual Dictionaries for Learners of English*. Poznań: Adam Mickiewicz University Press.
- Ptasznik, Bartosz and Robert 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.
- Singmann, Henrik, Ben Bolker, Jake Westfall and Frederik Aust.** 2018. *afex: Analysis of Factorial Experiments* (R Package Version 0.19-1).
- Tono, Yukio.** 1992. The Effect of Menus on EFL Learners' Look-up Processes. *Lexikos* 2: 230-253.
- Tono, Yukio.** 1997. Guide Word or Signpost? An Experimental Study on the Effect of Meaning Access Indexes in EFL Learners' Dictionaries. *English Studies* 28: 55-77.
- Tono, Yukio.** 2011. Application of Eye-Tracking in EFL Learners' Dictionary Look-up Process Research. *International Journal of Lexicography* 24(1): 124-153.

APPENDIX A: New-line guiding devices

Zapisać **NUMER** znaczenia podkreślonego wyrazu, który został użyty w podanym kontekście: _____

Zapisać **CZAS** wymagany do wykonania zadania: _____

Women are advised not to drink more than fourteen units of alcohol per week.

UNIT noun

1 **GROUP** a group of people working together as part of a larger group: *The man is in the hospital's **intensive care unit**.*

2 **MEASURING** an amount of something used as a standard of measurement: [**+of**] *The watt is a **unit** of electrical power.*

3 **PART** a thing, person, or group that is regarded as one single whole part of something larger: *a Russian **army unit** | [**+of**] *The family is the basic social **unit** of modern society.**

4 **PART OF A BOOK** one of the numbered parts into which a **TEXTBOOK** (= a book used in schools) is divided

5 **PRODUCT** *technical* a single complete product made by a company: *The factory's output is now up to 150,000 **units** each month.*

6 **PART OF A MACHINE** a piece of equipment which is part of a larger machine: **control / display / filter etc unit**

7 **SCHOOL/UNIVERSITY** *AmE* an amount of work that a student needs to do in order to complete a particular course

APPENDIX B: Run-on guiding devices

Zapisać **NUMER** znaczenia podkreślonego wyrazu, który został użyty w podanym kontekście: _____

Zapisać **CZAS** wymagany do wykonania zadania: _____

Women are advised not to drink more than fourteen units of alcohol per week.

UNIT noun

1 **GROUP** a group of people working together as part of a larger group: *The man is in the hospital's **intensive care unit**.* **2** **MEASURING** an amount of something used as a standard of measurement: **[+of]** *The watt is a unit of electrical power.* **3** **PART** a thing, person, or group that is regarded as one single whole part of something larger: *a Russian **army unit*** | **[+of]** *The family is the basic social unit of modern society.* **4** **PART OF A BOOK** one of the

numbered parts into which a **TEXTBOOK** (=a book used in schools) is divided **5** **PRODUCT** *technical* a single complete product made by a company: *The factory's output is now up to 150,000 units each month.* **6** **PART OF A MACHINE** a piece of equipment which is part of a larger machine: **control / display / filter etc unit** **7** **SCHOOL/UNIVERSITY** *AmE* an amount of work that a student needs to do in order to complete a particular course

African Language Dictionaries for Children — A Neglected Genre

Elsabé Taljard, *Department of African Languages, University of Pretoria,
Pretoria, South Africa (elsabe.taljard@up.ac.za)*

and

Danie Prinsloo, *Department of African Languages, University of Pretoria,
Pretoria, South Africa (danie.prinsloo@up.ac.za)*

Abstract: Children's dictionaries are instrumental in establishing a dictionary culture and are the gateway to sustained and informed dictionary use. It is therefore surprising that very little attention is paid to these dictionaries in scholarly research. In this article we reflect on the design of two series of dictionaries and one free-standing dictionary, all presumably aimed at first-time dictionary users, specifically looking at how selected design elements are aligned with the lexicographic needs of the target users. We argue that the conceptualization of children's dictionaries for African-language-speaking children should be a bottom-up process, and that an Afrocentric approach, taking the target user's Frame of Reference as the point of departure, is preferable to a Eurocentric approach, which often leads to a mismatch between conceptual relationships and linguistic form and function in African language dictionaries.

Keywords: CHILDREN'S DICTIONARIES, AFRICAN LANGUAGE DICTIONARIES, USER'S PERSPECTIVE, THEORY OF LEXICOGRAPHIC COMMUNICATION, AFROCENTRIC APPROACH TO DICTIONARY COMPILATION, USER'S FRAME OF REFERENCE (FOR)

Opsomming: Afrikataalwoordeboeke vir kinders — 'n miskende genre. Kinderwoordeboeke speel 'n belangrike rol in die vestiging van 'n woordeboekkultuur en gee toegang tot volgehoue en kundige woordeboekgebruik. Dis is daarom vreemd dat daar min aandag gegee word aan hierdie tipe woordeboeke in akademiese navorsing. In hierdie artikel besin ons oor die ontwerp van twee woordeboekreekse en 'n vrystaande woordeboek, wat almal oënskynlik gemik is op beginnergebruikers van woordeboeke. Ons kyk spesifiek na die belyning van die ontwerp-aspekte van die woordeboeke met die gebruikers se leksikografiese behoeftes. Ons dui aan dat die konseptualisering van kinderwoordeboeke vir sprekers van Afrikatale op grondvlak behoort te begin, en dat 'n Afrosentriese benadering verkieslik is bo 'n Eurosentriese benadering. Laasgenoemde benadering lei dikwels tot 'n mispassing tussen konseptuele verhoudings, en linguïstiese vorm en funksie in Afrikataalwoordeboeke.

Sleutelwoorde: KINDERWOORDEBOEKE, AFRIKATAALWOORDEBOEKE, GEBRUIKERS-PERSPEKTIEF, TEORIE VAN LEKSIKOGRAFIESE KOMMUNIKASIE, AFROSENTRIESE BENADERING TOT WOORDEBOEKKOMPILASIE, GEBRUIKER SE VERWYSINGSRAAMWERK

Introduction

Judging by the dearth of research on children's dictionaries, lexicographers do not seem to regard this dictionary genre as worthy of serious academic attention. The situation is even worse for children's dictionaries published in the (South) African languages. With the exception of the publication of Gouws, Prinsloo and Dlali (2014) and an MA dissertation by Putter (1999), children's dictionaries have not been the focus of lexicographic research, attracting little more than a sidelong reference in general discussions on lexicographic theory and practice. Gouws (2013) does however refer to the important role that foundation phase dictionaries, i.e. dictionaries aimed at 7–9 year olds, play in the establishment of a dictionary culture in South Africa. He rightly remarks that "a general theory of lexicography [...] should assist any lexicographer planning any new dictionary, also foundation phase dictionaries". Dictionaries for children are the gateway to the establishment of a dictionary culture and it is therefore surprising that the compilation and evaluation of such dictionaries receive so little attention from researchers in lexicography.

The Oxford First Bilingual Dictionary series, e.g. the *Oxford First Bilingual Dictionary Setswana + English* (2008) reflects a layout by topic, indexes in English and Setswana and has a curriculum focus. Such a thematic ordering as well as the provision of indexes was also followed in a series of Foundation Phase dictionaries published by Maskew Miller Longman in 2010 and the Official Foundation Phase CAPS dictionaries of the South African National Lexicography Units (2018).

In this article we reflect on the design of two series of dictionaries and one free-standing dictionary, all presumably aimed at first-time dictionary users, specifically looking at how selected design elements are aligned with the lexicographic needs of the target users.

Our study takes the user's perspective as its point of departure, but we also refer to Beyer's theory of lexicographical communication, specifically referring to the importance of the user's Frame of Reference (Beyer 2014). The dictionaries on which our investigation will focus are:

- A series of Foundation Phase dictionaries, 2010, published by Maskew Miller Longman (henceforth MML dictionaries);
- The Official Foundation Phase CAPS dictionaries, 2018, published by the South African National Lexicography Units (henceforth NLU dictionaries); and
- The *Ju/'hoan Children's Picture Dictionary*, 2014, published by University of KwaZulu-Natal Press (henceforth JCP dictionary).

The first two titles both constitute a series of bilingual dictionaries, with each dictionary treating English and an African language. Both dictionaries consist of three major sections, i.e. a picture section on specific themes such as 'my body', 'clothes', 'my family', 'our home', 'in the kitchen', etc., followed by sec-

tions where lemmas are presented in alphabetical order in English and the relevant African language (with or without treatment), and page numbers linking the word with its illustration in the thematic section. Figure 1 is an extract from the thematic section of MML.

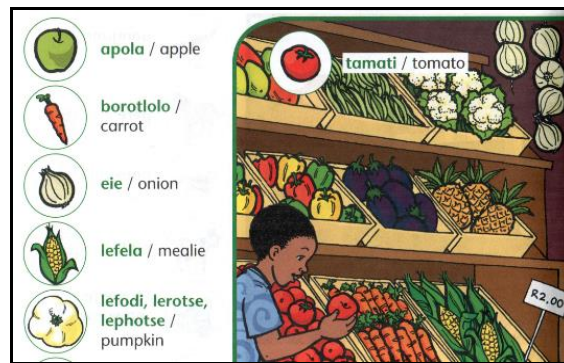


Figure 1: An extract from the thematic section 'Fruit and vegetables' in MML

In the NLU dictionaries a similar thematic pictorial approach is followed, but pictures are mainly presented individually as in figure 2. Indexes in English and the African language of the words and their illustrations are given, with page numbers linking the lemma with its pictorial illustration.



Figure 2: An extract from the thematic section "At school" in NLU dictionary

The third dictionary is free-standing, treating three languages, i.e. Ju|'hoan, Afrikaans and English. Ju|'hoan is a Northern Khoesan language, spoken by San people in Namibia and Botswana. It is an endangered language, spoken by

11 000 speakers. The Ju |'hansi people are to a large extent a marginalized society, trying to survive under difficult socio-economic circumstances. In the JCP the thematic section focuses on animals, birds, insects, hunting, dance, and home and family — themes of importance in the everyday life of Ju |'hoan children, cf. figure 3. At the back of the dictionary a translated word list per theme of the Afrikaans words in Ju |'huansi and English is given with page references to words and their pictorial illustrations in the main section of the dictionary.



Figure 3: An extract from the thematic section animals in JCP

In all of these dictionaries, treatment revolves around pictorial illustrations. Cianciolo (1981: 1) says picture books in general communicate with and appeal to children; they enrich, extend and expand children's background of experiences. Al-Kasimi (1977: 97) however says that the importance of pictorial illustrations is not always taken seriously and that the use of pictorial illustrations is rarely dealt with in the literature on lexicography. Gouws et al. (2014: 39) conclude that "the challenge for the lexicographer is to maintain a sound balance between the selection of the terms, the extent of the treatment, the detail of the distinction and the target user's skills and existing knowledge." Gouws et al. (2014: 25), emphasize that no single dictionary can be everything for everyone. This also holds true for dictionaries directed at specific target users, and each dictionary has to be considered as part of the broader family of dictionaries and not as an isolated product.

In order to contextualize our study, we will first present some terminological clarification as to the use of the terms 'children's dictionaries' versus 'picture dictionaries' since it has transpired that some confusion exists regarding the use of these terms. This will be followed by a classification of the dictionaries according to Tarp's (2011) suggested dictionary typology. Since we approach our discussion from a user's perspective, we then attempt to establish the profile of the intended users and their lexicographic needs, because these aspects are crucial in the planning and conceptualization of a dictionary. It has to be mentioned that whilst monolingual picture dictionaries might be regarded as the norm in the rest of the world, learners in South Africa are expected to use bilingual dictionaries, which immediately place them in relation to the dominant language (English). A detailed discussion of the preference for and/or desirability of monolingual versus bilingual dictionaries, how-

ever, fall outside the scope of this article. We discuss the issue of an Afrocentric versus a Eurocentric approach to compiling dictionaries for African language children. A critical analysis of the selection of visual content viewed against the initial conceptualization is followed by suggested guidelines for illustrations in children's dictionaries. Since paper dictionaries are still the norm for African language speaking children, our discussion is restricted to paper dictionaries.

Children's dictionaries/picture dictionaries: terminological clarification

As Tarp (2011: 229) points out, commercial publishing houses cannot be expected to use scientifically correct terms for their products. Their use of terminology is determined by their target market rather than by a desire to be scientifically correct. In the academic literature, it seems that a terminological distinction is made between picture dictionaries on the one hand, and children's dictionaries on the other. The term 'picture dictionary' is used to refer to a special type of dictionary, especially designed to assist people with speech disabilities or no speech to communicate. These dictionaries therefore do not only have children as their target users and they are often tailor-made to the needs of a specific user and usage situation. Even so, the term 'picture dictionary' is used by publishers to refer to and market dictionaries aimed at children, where pictures or illustrations provide access to the lexicographic information provided in the dictionary. Two of the dictionaries under discussion in this article contain the words 'picture dictionary' in their titles, despite their intended target users not fitting the profile of users of picture dictionaries as described in the academic literature. The term 'children's dictionary' is perhaps a more apt designation for the dictionaries under discussion in this article. All three dictionaries are onomasiological in nature, moving from a concept via a pictorial illustration to a word and its meaning. The onomasiological nature of these specific dictionaries is the essential feature that distinguishes them from other dictionaries, such as what is generally termed school dictionaries. An onomasiological approach provides a natural fit for a thematic — instead of an alphabetical — ordering, since concepts and their accompanying illustrations, related to a specific theme, are grouped together. Access to the information contained in the dictionary is therefore via the illustration, and on finding the relevant picture, the user should be able to connect the concept to its linguistic representation, i.e. the lemma.

Typological contextualization

In terms of Tarp's (2011) typology of pedagogical dictionaries the dictionaries under investigation straddle two typological categories. In terms of the age of the target group, these dictionaries are classified as school dictionaries, and if the type of learning that the dictionaries are supposed to support is taken as a typological criterion, they should be classified as dictionaries for both native language and non-native language learners. The category 'school dictionary'

can, in terms of the South African education system, be further subdivided according to the educational phase of the target users, i.e. Foundation Phase (Grades 1 to 3, although Grade 0 or R is sometimes informally included), Intermediate Phase (Grades 4–7), Senior Phase (Grades 7–9), which all fall within the General Education and Training Band, and the Further Education and Training Band (Grades 10–12). The first two dictionaries that we refer to in our discussion are both aimed at Foundation Phase learners, i.e. learners in the age bracket of roughly 7 to 9 years, or 6 to 9 years if Grade 0/R is included. This is also reflected in the titles of the two dictionaries. These learners are typically pre-literates to early literates. Linking a dictionary to a specific educational phase is a generally recognized marketing strategy amongst (South African) publishers — it is aimed at assisting parents and teachers in the selection of an appropriate dictionary. Furthermore, schools (via the provincial departments of education) are the main potential clients of dictionary publishing houses. The *Ju/'hoan Children's Picture Dictionary*, published for Namibian children, does not refer explicitly to any educational phase of the target users, other than simply indicating that it is a dictionary for children.

Profiling the user

Identifying the target user of the dictionaries under discussion seems a straightforward exercise. We have already referred to the age group at which these dictionaries are aimed. From a pedagogical perspective, the target users of the MML and NLU dictionaries are learners who — according to the official policy of the Department of Basic Education — receive instruction in their home language for the first three years of formal schooling (i.e. from grade 1).¹ However, as South Africa is a multilingual environment, the learning of a second or additional language plays an important role, since, for many learners, the language of instruction switches to English in their fourth year of schooling. It is therefore important that learners be exposed to English from early on. Gouws et al. (2014: 26) however, point out that the need for dictionaries directed at learners' home languages should never be denied. In order to promote multilingualism, it is important that South African learners have a solid grounding in and knowledge of their home language before moving on to the acquisition of a second language, cf. Gouws et al. (op. cit.). The emphasis when planning a dictionary for these target users should therefore be on assisting users with the understanding of basic concepts, appropriate for their age level, in their home language, thus bridging the gap between pre-literacy and literacy. A secondary aim should be the provision of access to English as an additional language. Such an approach would also be in line with the official Language in Education policy, in which the promotion of additive bilingualism is stated as one of the aims of the Ministry of Education's language policy, cf. <https://www.education.gov.za/Portals/0/DoE%20Branches/GET/GET%20Policies/LanguageEducationPolicy1997.pdf?ver=2008-06-28-134223-253>.

The target user of the JCP dictionary has a different profile. The title of the

dictionary refers to children as target users, without specifying their age. In the front matter of the dictionary, it is stated that native-speakers of Ju|'hoansi and learners of Ju|'hoansi as an additional language can use the dictionary. In the case of the MML and NLU dictionaries, the target user is contextualized within the formal South African education system, and the dictionary is presented as a tool that can support learners within a school environment. These aspects are lacking in the case of the JCP dictionary. In fact, it is not clear to what extent the children in the Ju|'hoan community are exposed to formal schooling, or what language that schooling is offered in.

Although the delineation of the intended user of the MML and NLU dictionaries seems clear, one can argue that it represents a rather over-simplified view of the actual grass roots situation in South African schools. Target users of school dictionaries in South Africa are extremely diverse, reflecting the inequalities that continue to plague South African society, with educational inequalities being among the most prominent. Webb (1999: 358) identifies three sociolinguistic categories relevant to the education sector, i.e. larger urban areas, smaller urban areas and deep rural areas. Each of these categories has its own profile with regard to, *inter alia*, the level of exposure to an additional language, the general literacy level and the prominence of African languages as languages of teaching and learning. Although it would be unfair to expect of lexicographers and publishing houses to make provision for the total diversity of the target users, we are of the opinion that this aspect needs to be considered, specifically with reference to series of dictionaries, such as the MML and NLU dictionaries. In each dictionary of the series, a different language pair is treated. Even within the parameters set by the curriculum of the Foundation Phase that forms the basis of these dictionaries, some leeway for differentiation between language pairs should exist and should be explored by lexicographers and publishers. This issue is discussed below and illustrated by means of practical examples.

Conceptualization of the dictionary: top down or bottom up?

It is generally accepted within the lexicographic community that the compilation of any dictionary must be preceded by proper planning, cf. Gouws and Prinsloo (2005: 13-19). During the planning phase, issues such as the organisation plan, the genuine purpose and lexicographic function, and the conceptualization of the dictionary are considered. Dictionary planning is thus a pragmatic and concrete process, with tangible outputs. These processes are however, very seldom documented and any researcher interested in these inner workings of the lexicographic process for any specific dictionary or dictionary project has mostly to rely on information provided in the front matter of the dictionary, in media releases and marketing blurbs. Additionally, this type of information has to be inferred from an analysis of the structure and content of the dictionary. Fortunately for the MML dictionaries, a description of the planning process is provided in Gouws et al. (2014: 26, 27). They indicate that the dictionary-

ies are aimed at native-speakers of the respective languages, but provision is to be made for access to English as an additional language. This means that, for example, the Sepedi–English version of the dictionary series is aimed at native-speakers of Sepedi who are learning English as a first additional language. The dictionaries are furthermore envisaged as assisting users with text reception and text production, as well as serving their cognitive needs. This then constitutes the genuine purpose of the dictionary. In the user's guide of these dictionaries, the information about the genuine purpose of the dictionary is confirmed and also conveyed to the user. Compare the following excerpt from the Sepedi–English dictionary:

This dictionary will help you to:

- speak, read and write in Sepedi as Home Language
- understand and use English as Additional Language
- learn and use Sepedi as Additional Language.

In addition to the explicit formulation of the genuine purpose of these dictionaries, the lexicographers have stated a complementary aim, i.e. the establishment and promotion of a dictionary culture, familiarizing target users with the dictionary as a practical instrument that can play an important role in the process of lifelong learning, cf. Gouws et al. (2014: 27).

As is the case regarding the MML dictionaries, the NLU dictionaries' genuine purpose is that of production, reception and cognition, and they also echo the sentiment of establishing a dictionary culture from an early age. However, the NLU dictionaries are less clear about the language profile of their intended target user, stating in the user's guide to these dictionaries that they were designed to assist learners with learning the relevant African language or English as an additional or second additional language. This seems to imply that either the African language or English could be the home language of the intended target user. However, it is also indicated that the dictionary can be instrumental in the transition from the African language as medium of instruction in the Foundation phase to English as medium of instruction in the Intermediate phase, thus implying that the actual target users are speakers of an African language as home language, who are learning English as their first additional language. This would be in line with the actual situation in South African classrooms, where 81.9% of learners in the Foundation phase have an African language as home language². As dictionaries are compiled for specific users, it is extremely important that lexicographers should be clear about the language profile of the target user, since the language through which the user will access the information in the dictionary will have a direct impact on the structure of the dictionary — it would, *inter alia*, determine the language of lemmatization and also the metalanguage used in the dictionary.

Gouws (2016: 365) points out that the genuine purpose of a dictionary may extend beyond the mere extraction of information from the data that is provided for the lexical items represented by the lemma signs. The information

leaflet that accompanies the JCP dictionary states that the compilation of the dictionary is a "community-driven project that highlights Ju|'hoan culture". From the rudimentary microstructure, the nature of the illustrations and layout of the dictionary, it is clear that extraction of information through a detailed lexicographic treatment is not the main aim of the compilers of this dictionary. The editorial team consisting of academics, project advisors and illustrators also includes members of the Ju|'hoan community, and another stated aim of the team is to create a broader awareness about the Ju|'hoan language and culture, and to prevent this endangered language from disappearing forever.

It is essential that the genuine purpose of the dictionary needs to be realized in the compilation thereof. The extent to which the final lexicographic product is aligned with its genuine purpose will determine its usefulness as a lexicographic utility tool.

Since the MML and NLU dictionaries are school dictionaries, it is to be expected that their content must be aligned and integrated with the curriculum for the relevant educational phase at which these dictionaries are aimed. Both the MML and NLU dictionaries emphasize the fact that the content is structured to align with information provided in the Curriculum Assessment Policy Statement (CAPS) documents. These documents are official documents, from the Department of Basic Education, setting out the policy on curriculum content and assessment for each educational phase. The users' guides to both dictionaries indicate that the dictionaries are theme based, and that the selection of themes was done according to the CAPS for the Foundation Phase.

Basing a dictionary and its contents on a prescribed school curriculum implies that the conceptualization of the dictionary was in essence top-down. The contents of the dictionary, i.e. the themes to be addressed, as well as the lemma list are to a certain extent predetermined. The user's guide to the NLU dictionaries states that the themes include the basic words of objects and actions that learners may encounter in everyday life. This statement raises an issue that is particularly pertinent to children's dictionaries, i.e. the notion that literature aimed at children needs to relate to the child's lived experience and that lexicographers should take cognisance of the user's frame of reference (FoR). The other side of the lexicographic coin is that lexicographers also operate from within their own frame of reference. According to Beyer's (2014) theory of lexicographic communication (TLC), the FoR of the lexicographer is part of the expressive information conveyed in a dictionary. Expressive information refers to information about the lexicographer as sender, one aspect of which is the lexicographer's frame of reference (FoR). He states the following in this regard: "Both the sender's and the receiver's *frame of reference* (=FoR) come into play here, since messages are construed and encoded in utterances against the sender's FoR, and utterances are decoded and messages reconstructed against the receiver's FoR" (Beyer 2014: 47). In the case of children's dictionaries, it is almost inevitable that the (adult) lexicographer's frame of reference and that of the child as target user will not coincide, resulting in an unsuccessful transfer of the lexicographic message to the user, or simply put, an unsuccessful dic-

tionary consultation. Nevertheless, determining the FoR of the target user should form part of the dictionary planning process and should not be left to supposition. Granted, determining the FoR of children in this specific age bracket is no mean task. One possible solution is suggested by Cignoni et al. (1996): "It is important to evidence the significance and determining aid that children can give to the compilation of a dictionary designed for their purposes and needs". They suggest that children as target users of the planned dictionary should not only be seen as users of the final product, but they should also participate in the compilation of the dictionary as a utility tool. Involving the target user in the planning and conceptualization of the dictionary would constitute a bottom-up process, which could contribute to bridging the gap between the perceived FoR of children and their actual FoR. Determining the FoR of South African Foundation Phase learners is complicated by the heterogeneous nature of the target user, as pointed out earlier. In this regard, the planning and conceptualization of the JCP dictionaries represent an innovative and creative approach. The planners of this dictionary had the advantage of a small, homogeneous body of target users, whose FoR is easier to determine. In an information pamphlet accompanying the dictionary, it is stated that members of the community led the way in the selection of themes, lexical entries, design and layout, and that children were actively involved in developing the JCP dictionary.

The themes that have been selected in the JCP use the real life experience(s) of the target user as their point of departure and reflect the lived experiences of the target users. These themes are: animals, birds, insects, reptiles and creepy crawlies, home and the family, hunt, gather and dance. Selecting themes that children as dictionary users can relate to and which fall within the ambit of their FoR can contribute not only to a successful dictionary consultation process, but also to the establishment of a dictionary culture.

A second aspect of the top-down approach is the motivation for the compilation and publishing of the dictionary. For both the MML and the NLU dictionaries, the initial motivation for the dictionaries was to a certain extent externally motivated. Gouws et al. (2014: 24) point out that it is common practice in South Africa for commercial publishers to take responsibility to continue providing dictionaries to satisfy the lexicographic needs of different user groups with different needs and reference skills. With regard to the MML dictionaries, Gouws et al. (op. cit.) state that "[i]n order to meet the lexicographic needs of a new generation of South African learners, the publishing house Maskew Miller Longman (MML) decided to launch a series of dictionaries in different official languages which are aimed at learners in the foundation phase of primary school". Even though externally (and commercially) motivated, such an approach should not be viewed in a negative light. The publishing of dictionaries as a commercial venture actually works to the advantage of the production of good dictionaries. A dictionary that satisfies the needs of the intended user should — at least in theory — add to the commercial value of the dictionary, although we recognize the fact that quality content is not the only factor that determines whether a dictionary will sell or not. Even so, it is in the

interest of publishers to avail themselves of the services of lexicographers, experts in education and language experts in order to ensure the publication of a good dictionary. This is exactly the approach taken by the publishers of the MML dictionaries.

The NLU dictionaries, on the other hand, have been compiled to fulfil a legislative obligation, and are thus also externally motivated. The back cover of each of the dictionaries in the series states that the dictionary is a product of the relevant NLU, who is "constitutionally and legislatively mandated to develop dictionaries and other material, on behalf of the South African government, in order that all Government Departments and Agencies, public and private schools, tertiary institutions and libraries have the resources needed to fulfil their obligations to all our country's indigenous languages". We do recognize that the fulfilment of a legislative obligation does not exclude other motivational factors such as commercial interests or competition from other publishers, but these considerations are not always made public. A strong point in favour of these dictionaries is the fact that they have been compiled by trained lexicographers, who are also first language speakers of the relevant African languages. However, a question that needs to be asked is to what extent the input of the lexicographers was restricted by the publisher.

Although the compilation of the JCP dictionary is also not entirely internally motivated, there is evidence of a much greater involvement and support from the Ju|'hoan community. As already indicated, the selection of the themes to be addressed was done in collaboration with members of the community, but perhaps even more significantly, the illustrations, as well as the sound and video recording which appears on a CD-ROM that accompanies the dictionary, were all done with the direct assistance of community members. This dictionary is probably the most successful in terms of establishing the FoR of the target users, and secondly in terms of the involvement of the speaker community in the initial planning and conceptualization of the dictionary.

Eurocentric or Afrocentric?

In a discussion regarding the Eurocentric versus the Afrocentric nature of African language dictionaries, Prinsloo (2017: 5) sketches a number of scenarios for African language dictionaries. These scenarios range from dictionary compilation by foreigners abroad, which represents a true Eurocentric approach, to dictionary compilation by African language speakers, which represents a true Afrocentric approach. According to Gouws (2007: 315), the linguistic situation in postcolonial Africa has resulted in a "drastic swing from externally motivated to internally motivated dictionaries, resulting in a situation which sees the majority of new lexicographic projects in Africa characterized by an Afro-centred approach that deviates from the Eurocentred approach". This seems to be a rather over-optimistic view of the current state of African language dictionaries. Being internally motivated and compiled by African language speakers does not guarantee an Afrocentric approach to dictionary compilation. Prah (2007: 23)

is rather blunt in his condemnation of the current situation, referring to "cultures and languages of the majorities [which] are suppressed and silenced in favour of a dominant Eurocentric high culture, which everybody is willy nilly obliged by force of circumstance to emulate".

A children's dictionary that is truly Afrocentric is much more than a dictionary characterized by a superficial adaptation of illustrations depicting black children instead of their white counterparts. It is a dictionary that is sensitive to portraying anything European as the default and as an ideal that has to be emulated. In the following example, the concept 'my house', taken from the NLU series is represented by an illustration that is unmistakably a typical European dwelling:



Figure 4: Illustration of the lemma 'house' in NLU dictionary

There is no denying the fact that many African language speaking children do grow up in houses such as these, but the reality in South Africa is that many children grow up in townships, informal settlements and rural areas where dwellings look much different from the one depicted above. A dictionary that is truly Afrocentric should make provision for the fact that a less formal or traditional dwelling can also represent the concept 'my house', cf. the illustration for the concept 'house' in the JCP dictionary:



Figure 5: Illustration of the lemma 'house' in the JCP dictionary

The dictionary should furthermore also be sensitive to what can at best loosely be termed African values. We use the term 'African values' in a very circum-spect manner, since we are quite aware of the dangers of generalization and stereotyping, but again, this relates to depicting and thus defining concepts in terms of the FoR of the target user. Instead of the usual nuclear family consisting of a mother, father and their offspring, the family depicted in the NLU dictionary is the extended family, including the grandparents, which is often typical of African societies. Considering the South African reality, a family is sometimes child-headed, or headed by a grandmother; these are factors that should be considered by lexicographers working within an African environment.

Dictionaries for African children should be much more than dubbed over versions of dictionaries initially conceptualized for English- (or Afrikaans-) speaking target users. African language lexicographers are all too often relegated to the role of translator. They are presented with a lemma list, and a preconceived dictionary layout and dictionary structure as a *fait accompli*. Lexicographers are then tasked with the provision of translation equivalents of the lemmas and also with the translation of, for example, definitions and usage examples, with very little regard to the challenges characteristic of African language lexicography. The remark by Gouws et al. (2014: 26) suggests that this was the case for the MML dictionaries: "MML realised the need for a dictionary directed at the specific needs and usage situation of South African foundation phase learners. This publishing house already had a monolingual English dictionary for first-language foundation phase learners, and a decision was made to embark on a series of foundation phase dictionaries for the other official South African languages, resulting in the current series of dictionaries". It would seem that the existing monolingual English dictionary was used as a blueprint for the bilingual dictionaries in the series. Although a pragmatic and efficient procedure with regard to time and money, such an approach does not always work to the advantage of African language dictionaries, the reason being that the whole conceptualization of the dictionary is based on a non-African language and a non-African epistemology. Such a practice can result in a mismatch between conceptual relationships and (linguistic) form and function. This is especially relevant in children's dictionaries, where the emphasis is on the establishment of conceptual relationships. The conceptual category 'attributes' is a case in point. In English, there is a one-to-one correspondence between the conceptual category 'attributes', the linguistic category 'adjective', and the syntactic structure used to express these concepts. In the African languages, however, there is a one-to-many correspondence between the conceptual category 'attributes' and the linguistic categories that are used as nominal descriptors. Attributes in African languages are expressed by a number of different constructions — some attributes are expressed by means of adjectives (*lebone le lehubedu* 'red light'), some by means of so-called possessive constructions (*ngwana wa boomo* 'headstrong child') or nominal relatives (*ngwana yo boomo* 'headstrong child'), and others by means of verbal relative constructions (*meetse a a elago* 'flowing water'). Furthermore, due to the complex system of concordial

agreement in the African languages, each syntactic structure can diverge into potentially 15 structures, depending on the noun that is being described. Compare the following diagram:

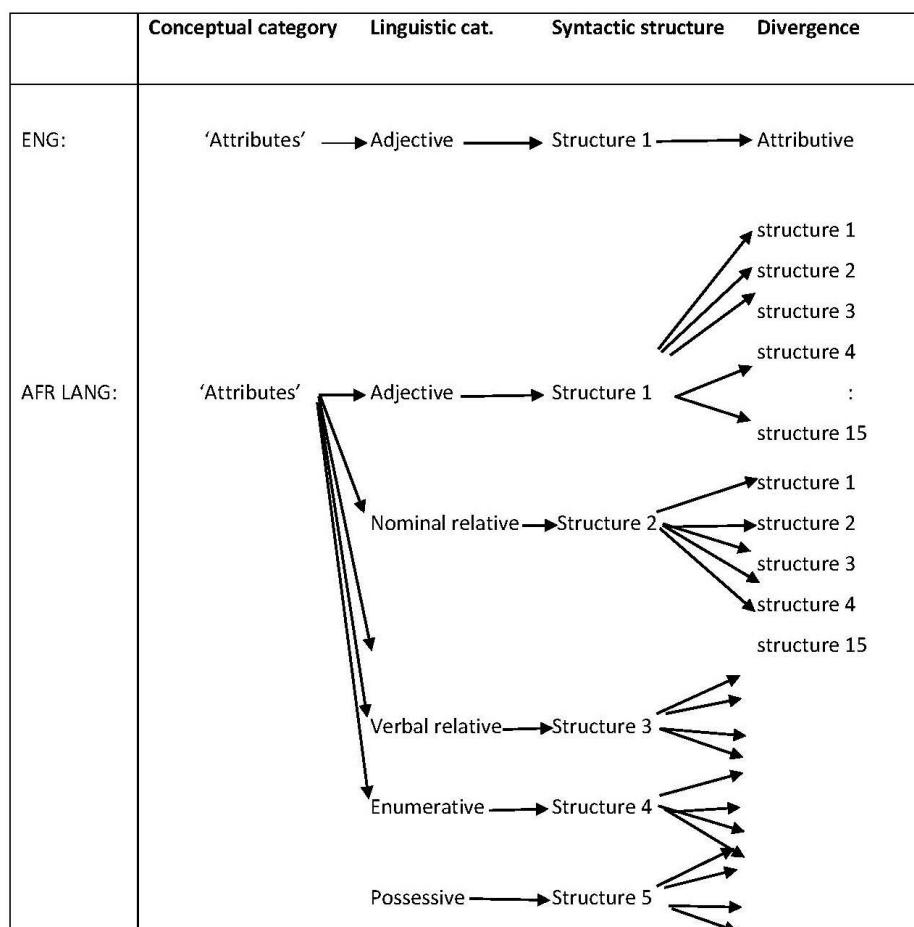


Figure 6: Correspondence between conceptual categories and syntactic structures

Whereas the concept 'red', for example, is expressed by means of the attributive use of the adjective in English, in Sepedi, this concept can be expressed as *wo mohubedu*, *se sehubedu*, *le lehubedu*, *a mahubedu* and *ye khubedu*, depending on the object which is described as being red. In cases such as these, the lexicographer needs to give additional guidance to the African language user, to compensate for the conceptual mismatch. This additional guidance could be by means of providing usage examples.

Matching concepts to pictures, matching pictures to lemmas

With regard to the use of illustrations in dictionaries in general, Gouws (2014: 164, 165) emphasises that illustrations must be functional, i.e. they must assist the user with the retrieval of (semantic) information; if not, they are non-functional, and can have a cosmetic function at best. The most important function of illustrations is therefore supporting verbal descriptions.

The illustrations in children's dictionaries are different to ostensive definitions in so-called illustrated dictionaries because of the onomasiological nature of children's dictionaries. An onomasiological approach implies that the point of departure is the concept, which is represented by the lemma as a linguistic sign and visually illustrated by means of a picture or illustration. The illustration therefore forms an integral part of the treatment of the lemma since it is the only element providing additional semantic information. The underlying assumption in the compilation of children's dictionaries is that the user — being pre-literate or early literate — will move from the illustration to the lemma. According to Gouws et al. (2014: 30), seeing the picture and word together allows users to make the necessary link between the concept as represented by the picture, and the relevant word. The illustration is probably the first point of entry for the user, specifically for pre-literate and early literate users. Only in cases where the user has already attained some level of literacy does it become possible to use the lemma as the guiding element or first point of entry. In such cases, the illustration serves as a support for the definition and/or example material.

In the NLU and JCP dictionaries, verbal descriptions are quite sparse, consisting of a single lemma and its translation equivalent(s) only. In these cases, the illustration is the primary element in providing additional semantic information. Compare the treatment of the lemmas 'blackboard' (*letlapa*) and 'hunter' (*!aqekxao*) in the NLU (Northern Sotho–English) and JCP dictionaries respectively:

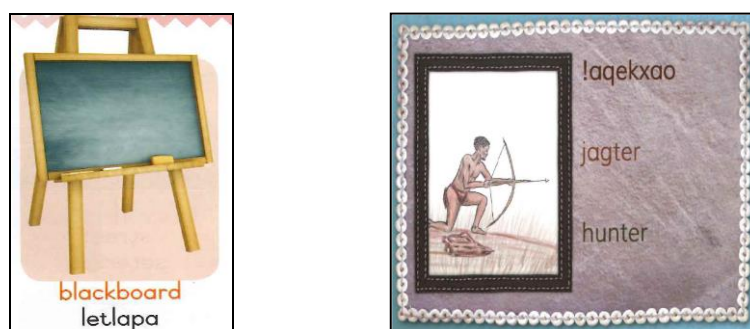


Figure 7: Lemmas 'blackboard' and 'hunter' in NLU and JCP dictionaries respectively

In the MML dictionaries, where definitions and example sentences are given, illustrations are additional support mechanisms and should be aligned with and confirm the information provided in the contents of these data categories. In the alphabetical lemma list of the MML dictionary, definitions are provided in the target language, followed by a usage example and its translation in English. Compare the treatment of the lemma 'snake' (*noga*) in the Northern Sotho-English dictionary:

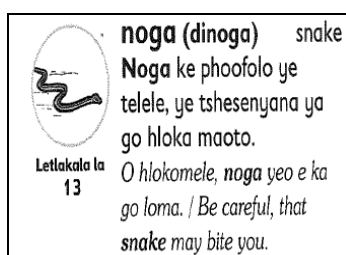


Figure 8: Lemma 'snake' in MML dictionary

In this example, the picture of the snake confirms the information given in the definition: a snake is a long, thin animal and does not have feet.

It is important that the illustration is a clear and unambiguous representation of the relevant concept, i.e. that the link between the concept and the illustration is clear, otherwise the illustration is non-functional. Compare the following examples from the NLU dictionary in which the days of the week are treated:

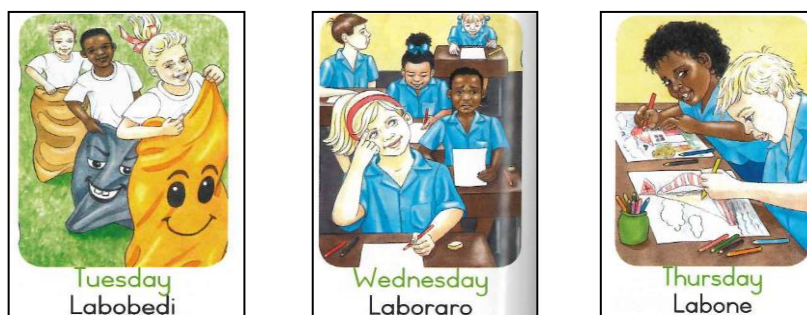


Figure 9: Days of the week in NLU dictionaries

The functionality of these illustrations is questionable. For example, the illustration of the lemma *Labobedi* 'Tuesday' does not provide users with additional semantic information that would enable them to distinguish Tuesday from the

other weekdays. Lexicographers would do well to follow the accepted lexicographic practice of defining the days of the week in terms of days that precede or follow a particular day, and find an innovative way to visually represent this. Having a picture of a calendar and indicating the days in relation to each other would be more functional. The utilization of the same illustrations for different, non-related concepts is a further indication that these particular illustrations may not be functional. The illustration used for Thursday, for example, is also used elsewhere in the same dictionary to illustrate the verb *itlwaetša* 'practice'; the illustration for Wednesday is used to illustrate the verb *šoma* 'work'. In the latter instances, the functionality of the illustrations is also doubtful.

Apart from the overarching function of illustrations of supporting verbal descriptions, a number of functions specifically relevant for children's dictionaries can be identified. In the first instance, illustrations should contribute to the cognitive development of the user, cf. Putter (1999: 92) in this regard. The thematic ordering of lemmas lays the basis for the establishment of semantic networks. Concepts that are semantically related can be presented in the same visual space, and illustrations can be utilized to highlight the distinguishing features of each concept. Differences and similarities between different concepts related to a specific semantic category can be illustrated. Compare the following examples of different animal tracks, categorized under the theme 'hunting' from the JCP dictionary:



Figure 10: Lemmas on animal tracks in JCP dictionary

The illustrations provide the user with enough information to distinguish cognitively between the tracks of a lion, a gemsbok and a giraffe. These lemmas, however, are treated on different pages, even though they belong to the same theme. Providing them on the same page would have strengthened the cognitive links between these concepts and would have made the similarities and differences immediately visually evident to the user.

The thematic pages of the MML dictionaries have the added advantage that related concepts are represented within a context and in the same visual space, thus visually supporting the relatedness of the concepts. Compare in this regard the theme on transport in the Northern Sotho–English dictionary:

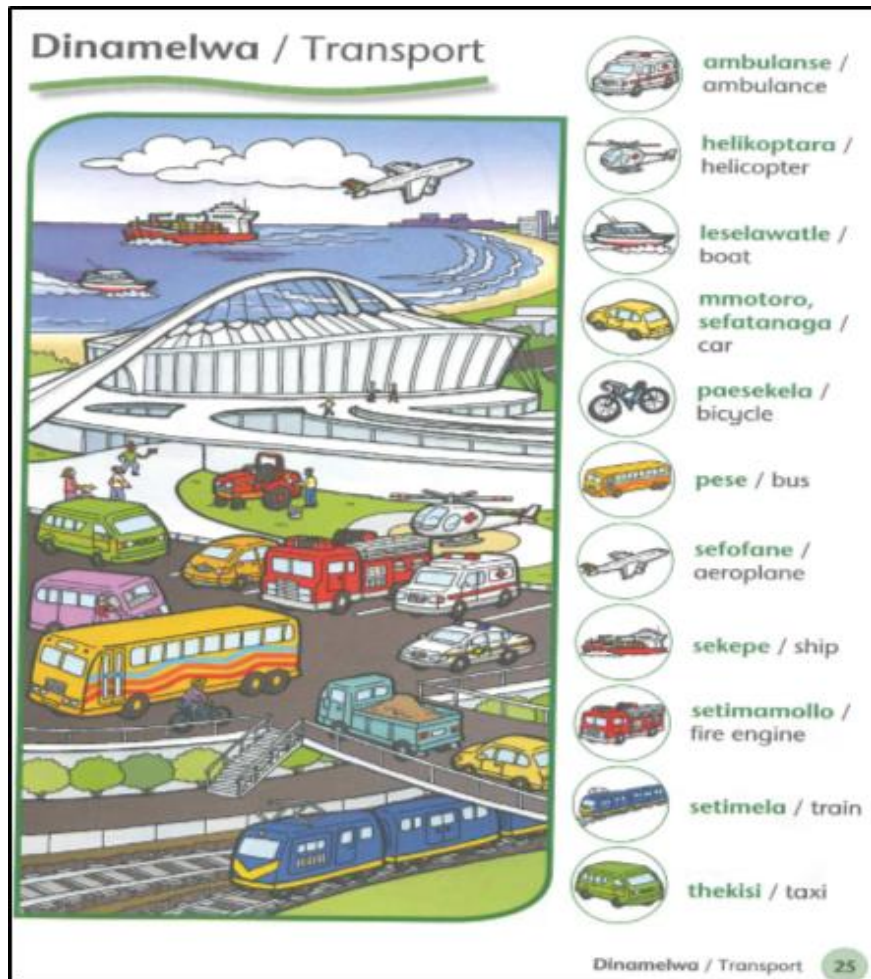


Figure 11: Transport theme in MML dictionary

The representation of different kinds of vehicles within the same space assists users in identifying the similarities and differences between related concepts at a glance: a train is a vehicle that travels on a track, an aeroplane flies in the air, and boats float on water. Although related concepts are treated on the same page in the NLU dictionaries, each lemma and its accompanying illustration is provided without any bigger context. In cases such as these the illustration usually depicts the prototypical member of a class of objects, cf. Putter (1999: 85). Note that the background of each individual illustration is neutral and does not provide any additional visual information that could supplement the semantic information.



Figure 12: Transport theme in NLU dictionaries

A second function of illustrations in children's dictionaries is to expand the user's experiential world, cf. Putter (1999: 99). She argues that users are exposed to illustrations of objects that do not form part of their frame of reference, and by internalizing these concepts, their FoR is expanded. Although we agree in principle, in terms of the discussion regarding the importance of establishing the FoR of the target user during the planning phase, we argue that a balance must be struck between what target users are familiar with and to which they can relate, and exposure to concepts with which they are (still) unfamiliar. A dictionary that limits users' exposure to new concepts by presenting them only with illustrations of objects that fall within their FoR will fail to meet their cognitive needs in that it does not offer users the opportunity to expand their FoR by internalizing new concepts. On the other hand, overwhelming users with illustrations of concepts that are foreign to their FoR can be equally negative. Although Gouws and Tarp (2017) focus their discussion of lexicographic information and data overload on e-dictionaries, we believe that their views can also be applied to paper dictionaries. Furthermore, whereas they use the term 'data overload' to refer to the provision of superfluous data *within* a dictionary article, we would argue for a more general interpretation of the term, i.e. that any element of the dictionary, be it on the microstructural or macrostructural level, can overload the user with data. Confronting users of a children's dictionary with too much (visual) information that falls outside of their FoR amounts to information overload. According to Speier et al. (as cited by Gouws and Tarp 2017: 395), "information overload occurs when the amount of input to a system exceeds its processing capacity". When taking into consideration the level of cognitive development of the target user of the dictionaries

under discussion, flooding them with illustrations of too many foreign concepts increases the cognitive load on the user. Not only do they have to internalize the new concept, they also have to make the link to the lemma as the linguistic sign representing the concept. This could alienate users from the dictionary, and have a negative effect on sustained dictionary use and the establishment of a culture of dictionary use.

It is against this background that one could question the inclusion of a number of lemmas and their accompanying illustrations in the NLU dictionaries. In the thematic section of sport, for example, the lemmas 'skiing', 'ice skating' and 'fencing' are treated, and in the section on entertainment, the lemmas 'piano', 'saxophone' and 'movie, film' appear — the concept 'piano' being illustrated by means of a picture of a grand piano.³ Of course, this relates to the issue of lemma selection — if the visual representation of a particular concept falls outside the FoR of the user, it is probable that the concept itself falls outside the FoR. The lexicographer therefore needs to consider whether a specific concept forms part of the lived experience of a learner in the Foundation Phase and whether the inclusion of such a lemma is justified. One cannot help but question the inclusion of lemmas such as 'urchin', 'barnacle', 'plankton', 'sloth' and 'reindeer' — to name but a few — in the NLU dictionaries.

Lexicographers should furthermore consider including lemmas representing concepts that may be specific to the FoR of speakers of a specific language. African languages in South Africa mostly have a strong regional basis; consequently, certain concepts with which, for example, a Sepedi speaker may be familiar will be outside the FoR of a Sesotho speaker, and vice versa.⁴ Consider, for example, concepts such as *masotša* 'mopani worms' and *lerula* 'marula fruit' which may represent everyday objects to a Sepedi speaker, but with which a Sesotho speaker may be unfamiliar. This would be true for names of plants, animals, birds etc. that are endemic to a particular geographical area. One could imagine that the Xhosa–English dictionary would need to include the names of and accompanying illustrations of common fish species, but that the Tswana–English dictionary in the same series would find the inclusion of such lemmas redundant, since they fall outside the FoR of many young Tswana speakers. The same argument would hold for cultural concepts that are particular to a specific culture and/or language group.

Possible guidelines for illustrations in children's dictionaries

Considering the very specific function of illustrations in children's dictionaries, it is important that these illustrations are designed in collaboration with lexicographers, a point already raised by Gouws et al. (2014: 37). Illustrations in a children's dictionary are different from illustrations in general children's literature, since they are lexicographic devices that function as guiding elements in the dictionary article. The following could be considered as guidelines for selection of illustrations that are functional as lexicographic devices.

Factual correctness

Care must be taken that an illustration is a factually correct representation of a particular concept as signified by the lemma. If there is a mismatch between the concept, the lemma and its illustration, the cognitive information that is presented to the user will be incorrect. In the NLU dictionaries, there are quite a number of lemmas where there is a mismatch between the concept, the lemma and the visual representation. In the thematic section on Entertainment, the lemma 'violin' is treated, but the illustration is that of a cello. In the English–Sepedi version, the English lemmas are mostly aligned with the illustrations, but there are instances of a mismatch between the Sepedi lemma and the illustration. The English lemma 'gemsbuck', for example, is correctly illustrated, but the Sepedi equivalent, i.e. *phala* refers to an impala. In the section on 'Different Homes', the lemma 'log cabin' appears. The lexicographer/translator may have been unfamiliar with the concept of a log cabin, and misinterpreted the English lemma as referring to a cabin that can be locked. The concept of a log cabin was clearly outside of the FoR of the adult lexicographer, once again raising the issue of appropriate lemma selection for a children's dictionary.

Free from 'visual noise'

Illustrations must be free from 'visual noise', i.e. unnecessary and irrelevant detail that could potentially distract users from correctly identifying the concept that is represented by the illustration. Al-Kasimi (1977: 102) uses the term 'preciseness' in this regard: "The dictionary user's attention should be directed only to the feature of the pictorial illustration relevant to the desired concept". The simplicity of illustrations in the JCP dictionary are perhaps the most successful of the three dictionaries in terms of preciseness: there can be little doubt as to what concept the picture is illustrating, cf. the lemmas 'chicken' and 'kettle' in this regard:

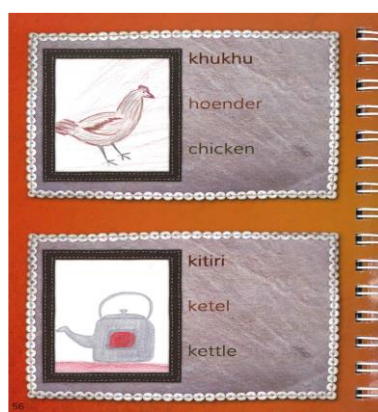


Figure 13: Lemmas 'chicken' and 'kettle' in JCP dictionary

Even so, there is always the possibility of interpretive difference, due to dictionary users' FORs being widely different.

In order to facilitate preciseness, lexicographers have at their disposal a variety of devices that can assist in focusing the user's attention on the feature that is relevant for functional illustration of the concept. These include arrows, colour cues to indicate the most important feature of the illustration, and position cues that, according to Al-Kasimi (1977: 102), imply that the most important portion of the picture should be placed either in the centre, or upper left part of the illustration.

A number of articles in the NLU dictionaries are less successful in this regard. Compare the illustrations for the lemmas 'tropic', 'grassland' and 'pond' in this regard.



Figure 14: Lemmas 'tropic', 'grassland' and 'pond' in NLU dictionaries

In all three examples, the pictures are inadequate representations of the aforementioned concepts. In these examples, an animal is the foregrounded, focal element in the illustration, and the user's intention will likely be distracted by the prominence of the chimpanzee, zebra and duck in the three examples respectively. The illustrations therefore do not adequately support the verbal information provided in the article, since they are open to dual and/or incorrect interpretation.

Free from stereotyping

Saying that children are impressionable beings is stating the obvious. Nevertheless, lexicographers who plan and eventually compile dictionaries with children as target users must be extremely sensitive to enforcing, sometimes inadvertently, any kind of stereotype, be it gender, racial or cultural stereotyping. Depicting the school principal as male and the teacher as female, using male figures to illustrate occupations such as those of engineer, lawyer, dentist, accountant and optometrist entrenches the perception that these occupations belong to the domain of males. In an attempt to circumvent stereotypical

depiction of occupations, the NLU dictionaries use figures that are supposed to be gender neutral; however, their attempts at gender neutrality are scuppered by the fact that these figures display distinctly male characteristics — they are all depicted wearing ties. The only figure that is unmistakably female — dressed in high heels and a flowery dress — is the picture used to illustrate the concept of shopping. For the compilers of the JCP dictionary, avoiding gender stereotyping must have posed a serious challenge, since gender roles are extremely strongly embedded in the cultural community of the Ju|'hansi speakers: men are hunters and women are gatherers. Depicting a hunter as female and someone gathering food from the veld as male would negatively affect the credibility of the cognitive information offered by the dictionary. The lexicographer therefore had no choice but to stick to the gender-based roles in preparing illustrations for the concepts 'hunter' and 'gatherer'.

Illustrating the un-illustratable

Gouws et al. (2014: 37) state categorically that it is not necessary for all lemmas in a children's dictionary to be illustrated with pictures, since some concepts can simply not be illustrated. If an illustration does not contribute to better understanding of the concept being illustrated, it becomes redundant and can even create confusion on the part of the user. The lexicographer should not fall into the trap of trying to include an illustration in each and every dictionary article, simply for the sake of using an illustration. Illustrations are lexicographic devices, therefore the decision as to the inclusion or exclusion of an illustration, the nature of the illustration and its placement should be a cooperative effort of both publisher and lexicographer, and both role players must consider the functionality of all illustrations. Lexicographers should look for innovative alternatives to provide additional semantic information for concepts that are difficult to illustrate due to their abstract nature or for another reason, cf. concepts such as 'manners', 'obedient', 'lazy', 'respect', etc. (cf. the NLU dictionaries' articles in this regard). A further consideration, specifically in the case of dictionaries aimed at children in the foundation phase, should be whether such a concept is indeed relevant in a foundation phase dictionary.

Conclusion

Dictionaries for children speaking an African language should be much more than beautifully illustrated picture books. Their conceptualization and subsequent compilation should be based on sound theoretical principles, as is the case for dictionaries compiled for adult users. In terms of Gouws et al. (2014), they should maintain a sound balance between the selection of the terms, the extent of the treatment, the detail of the distinction, and the target user's skills and existing knowledge. Conceptualization of these dictionaries should be a bottom-up process, calling for a collaborative effort between the target user, the

lexicographer, the publisher, educational experts, illustrators and graphic designers. In their design and compilation, the currently available dictionaries for South African children have not yet made the paradigm shift from a Eurocentric to an Afrocentric approach.

Acknowledgements

Financial support by (a) the South African Centre for Digital Language Resources (SADiLaR) and (b) the National Research Foundation of South Africa (Grant specific unique reference numbers 85763 and 103898) is hereby acknowledged. The Grantholders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the authors, and that the sponsors accept no liability whatsoever in this regard.

Endnotes

1. We acknowledge that a not insignificant number of learners, especially in metropolitan, urban areas attend schools in which the language taught as home language, is not the learner's actual home language. This is especially the case in areas which are linguistically diverse.
2. Statistics provided by the Department of Basic Education in personal e-mail.
3. We do acknowledge that our view on these concepts being outside the FoR of the target user may be subjective and to some extent intuitive, and fully agree that an empirical study is necessary to ascertain whether this is indeed the case.
4. Gauteng would be a notable exception to the regionality of the South African languages, being a province where speakers of many different languages live side by side.

Bibliography

Dictionaries

- Jones, K. and Cwi, T.F. 2014. *Ju/'hoan Tsumkwe Dialect/Prentewoordeboek vir kinders/Children's Picture Dictionary*. Pietermaritzburg: University of KwaZulu-Natal Press.
- Mabule, M. 2010. *Longman Pukuntšū ya sehlopafase: Sepedi/English*. Cape Town: Maskew Miller Longman.
- Mojela, V.M. (Ed.). 2018. *Picture Dictionary: English–Sesotho sa Leboa. Gr R–3: The Official English–Sesotho sa Leboa Foundation Phase CAPS-linked Illustrated Dictionary of the Government of the Republic of South Africa*. South Africa: Sesotho sa Leboa National Lexicography Unit.
- Paizee, D. (Ed.). 2008. *Oxford First Bilingual Dictionary. Setswana + English*. Cape Town: Oxford University Press Southern Africa.

Other

- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Beyer, H.L.** 2014. Explaining Dysfunctional Effects of Lexicographical Communication. *Lexikos* 24: 36-74.
- Cianciolo, Patricia J.** 1981. *Picture Books for Children*. Second edition. Chicago: American Library Association.
- Cignoni, L., E. Lanzetta, L. Pecchia and G. Turrini.** 1996. Children's Aid to a Children's Dictionary. Gellerstam, M. et al. (Eds.). 1996. *EURALEX '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 659-666. Gothenburg: Department of Swedish, Göteborg University.
- 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.** 2013. Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, V. (Ed.). 2013. *Specialised Lexicography. Print and Digital, Specialised Dictionaries, Databases*: 51-62. Lexicographica Series Maior 144. Berlin/Boston: Walter de Gruyter.
- Gouws, R.H.** 2014. Expanding the Notion of Addressing Relations. *Lexicography: Journal of ASIALEX* 1(2): 159-184.
- Gouws, R.H.** 2016. Enkele minder bekende Afrikaanse woordeboekmonumente (A Few Lesser Known Afrikaans Dictionary Monuments). *Tydskrif vir Geesteswetenskappe* 56(2-1): 355-370.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS, AFRICAN SUN MeDIA.
- Gouws, R.H., D.J. Prinsloo and M. Dlali.** 2014. A Series of Foundation Phase Dictionaries for a Multilingual Environment. *Stellenbosch Papers in Linguistics* 43: 23-43.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Prah, K.K.** 2007. *Challenges to the Promotion of Indigenous Languages in South Africa*. Retrieved from <http://www.casas.co.za>.
- Prinsloo, D.J.** 2017. Analyzing Words as a Social Enterprise: Lexicography in Africa with Specific Reference to South Africa. Miller, J. (Ed.). 2017. *Analysing Words as a Social Enterprise: Celebrating 40 Years of the 1975 Helsinki Declaration on Lexicography. Collected Papers from Australex 2015*: 42-59. <https://www.adelaide.edu.au/australex/publications/>.
- Putter, A.-M.** 1999. *Verklaringsmeganismes in kinderverwoordeboeke*. Unpublished M.A. Dissertation. Potchefstroom: Potchefstroom University for Christian Higher Education.
- Tarp, S.** 2011. Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. *Lexikos* 21: 217-231.
- Webb, V.** 1999. Multilingualism in Democratic South Africa: The Over-estimation of Language Policy. *International Journal of Educational Development* 19: 351-366.

Connecting the Dots: Tradition and Disruption in Lexicography*

Sven Tarp, *Centre of Excellence in Language Technology, Ordbogen A/S, Odense, Denmark; Department of Afrikaans and Dutch, University of Stellenbosch, South Africa; International Centre for Lexicography, University of Valladolid, Spain; and Centre for Lexicography, Aarhus University, Denmark (st@cc.au.dk)*

Abstract: This article botanizes in the history of lexicography trying to connect the dots and get a deeper understanding of what is happening to the discipline in the framework of the Fourth Industrial Revolution. The objective is to suggest possible ways out of the present deadlock. History shows that a sudden change of the technological base, like the one we are now experiencing, suggests a total revolution of the discipline in all its major dimensions. In order to be successful, such a revolution requires a mental break with past traditions and habits. As a matter of example, the article focusses on a series of bilingual writing assistants developed by the Danish company Ordbogen A/S and the new challenges posed to lexicography by these and similar tools. It argues that these challenges cannot be solved by means of traditional user research which is retrospective as it unfolds in the framework of an old paradigm. As an alternative, and without excluding other types of user research, the article recommends disruptive thinking by means of brainstorm, immersion, and contemplation and provides some examples on how to proceed. Finally, it problematizes the incipient competition between human and artificial lexicographers and gives a brief account of a possible future redistribution of tasks.

Keywords: FOURTH INDUSTRIAL REVOLUTION, DISRUPTIVE INNOVATION, ARTIFICIAL INTELLIGENCE, BUSINESS MODEL, WRITING ASSISTANTS, ARTICLE STRUCTURE, USER RESEARCH, TEST-DRIVEN DEVELOPMENT, HUMAN LEXICOGRAPHER, ARTIFICIAL LEXICOGRAPHER

Opsomming: Die lê van verbande: Tradisie en ontwingting in die leksikografie. In hierdie artikel word die geskiedenis van die leksikografie geanaliseer in 'n poging om verbande te lê en om 'n beter begrip te verkry van wat aan die gebeur is met hierdie dissipline binne die raamwerk van die Vierde Industriële Rewolusie. Die doelwit is die voorstelling van moontlike metodes om van die bestaande dooiepunt te ontkom. Uit die geskiedenis is dit duidelik dat 'n skielike verandering van die tegnologiese basis, soos die verandering wat ons nou ervaar, 'n totale omwenteling van die dissipline in al sy hoofdimensies suggereer. Om suksesvol te kan wees,

* The main ideas of this article were presented at the 24th Annual International Conference of the African Association for Lexicography (AFRILEX), hosted by the Department of Language and Literature Studies, University of Namibia, Windhoek, Namibia, 26–29 June 2019.

vereis so 'n omwenteling 'n breuk met tradisies en gewoontes van die verlede. As voorbeeld fokus die artikel op 'n reeks tweetalige skryfhulpmiddels wat deur die Deense maatskappy Ordbogen A/S ontwikkel is en op die nuwe uitdagings wat hierdie en soortgelyke hulpmiddels vir die leksikografie inhou. Dit voer aan dat hierdie uitdagings nie deur middel van tradisionele gebruikersnavorsing wat op 'n retrospektiewe manier binne die raamwerk van 'n ou paradigma ontvou, opgelos kan word nie. As alternatief, en sonder om ander tipes navorsing uit te sluit, word ontwrigtende denke in hierdie artikel aanbeveel deur te dinkskrum, jou te verdiep en te bespiegel, en enkele voorbeelde van hoe om te werk te gaan, word verskaf. Ten slotte word die aanvanklike wedywering tussen menslike en kunsmatige leksikograwe uiteengesit en 'n kort verslag van 'n moontlike toekomstige herverdeling van take word gegee.

Slutelwoorde: VIERDE INDUSTRIËLE REVOLUSIE, ONTWRIGHTENDE INNOVASIE, KUNSMATIGE INTELLIGENSIE, SAKEMODEL, SKRYFHULPMIDDELS, ARTIKELSTRUKTUUR, GEBRUIKERSNAVORSING, TOETSGEDREWE ONTWIKKELING, MENSLIKE LEKSIKOGRAAF, KUNSMATIGE LEKSIKOGRAAF

Creativity is just connecting things. When you ask creative people how they did something, they feel a little guilty because they didn't really *do* it, they just *saw* something. It seemed obvious to them after a while. That's because they were able to connect experiences they've had and synthesize new things. And the reason they were able to do that was that they've had more experiences or they have thought more about their experiences than other people.

Unfortunately, that's too rare a commodity. A lot of people in our industry haven't had very diverse experiences. So they don't have enough dots to connect, and they end up with very linear solutions without a broad perspective on the problem. The broader one's understanding of the human experience, the better design we will have.

(Interview with Steve Jobs in *Wired*, Wolf (1996)).

0. Introduction

The history of lexicography encompasses long periods with the accumulation of small and gradual changes within an existing paradigm as well as relatively short periods with abrupt and profound changes within a new paradigm. Today, we are witnessing such a paradigm shift described as a *Cambrian Explosion* by Fuertes-Olivera (2016). The phenomenon is characterized by a turmoil of both old forms that are surviving (printed dictionaries) and new forms that are constantly appearing and disappearing (PDF, CD-ROM, DVD, apps, handheld, web-based dictionaries, etc.). The present turmoil resembles Darwin's "survival of the fittest", especially if other digital reference resources competing with dictionaries are included; see e.g. Frankenberg-García (2018) and Alonso-Ramos and García-Salido (2019). In the final analysis, the turmoil is caused by the introduction and more or less successful application of disruptive technologies which, for their part, are continuously developing and improving.

Current lexicography is developing in the framework of the so-called

Fourth Industrial Revolution. According to Schwab (2015), who coined the term, this new phenomenon "is characterized by a fusion of technologies that is blurring the lines between the physical, digital, and biological spheres". The author, who is also the founder and executive chairman of the World Economic Forum, is emphatic that there is no historical precedent to the speed, scope, and complexity of the current breakthroughs:

When compared with previous industrial revolutions, the Fourth is evolving at an exponential rather than a linear pace. Moreover, it is disrupting almost every industry in every country. And the breadth and depth of these changes herald the transformation of entire systems of production, management, and governance. The possibilities of billions of people connected by mobile devices, with unprecedented processing power, storage capacity, and access to knowledge, are unlimited. (Schwab 2015)

As examples of emerging technology breakthroughs that may multiply these possibilities, Schwab lists "artificial intelligence, robotics, the Internet of Things, autonomous vehicles, 3-D printing, nanotechnology, biotechnology, materials science, energy storage, and quantum computing".

Whether one likes it or not, lexicography has to navigate and find its ways in this disruptive explosion of technological innovations. The present complexity of things does not allow lexicographers to hide their heads in the sand. To a large extent, it is a battle of life and death. It is therefore urgent to take action before the discipline gets caught in deep slumber.

This article will look at the turmoil and the crisis that has crept into lexicography during the past years as a result of the new technological breakthroughs. The crisis will be put into a historical perspective in order to get a more profound understanding of its complexity and main characteristics. A short excursion will be made into history with a special focus on the birth of European lexicography 2 500 years ago as well as the long-term consequences of the irruption of printing technology in European lexicography more than 500 years ago.

The current business model which is becoming increasingly obsolete will then be challenged and an alternative model outlined. The objective is to sustain a transformed lexicography that is fully prepared to be part of the new Industrial Revolution. To that end, the article will discuss what it takes to inject new blood into the lexicographical veins. As a matter of example, it will reflect on the current lexicographical challenges posed by the premium development of an integrated tool that provides instantaneous assistance to second-language writing. In this connection, the timely relevance of most current user research will be disputed and new ways of getting closer to the users and their real needs will be recommended.

Finally, the concept of an artificial lexicographer will be introduced as opposed to the traditional human lexicographer and a possible future distribution of tasks and responsibilities between the two of them will be outlined.

1. Lexicography in crisis

Although there are still regions, like Southern Africa, that continue living in the happy days of printed dictionaries, the general world tendency is now the propagation of web-based dictionaries. The onlinezation poses new challenges to lexicography. More than ten years ago, Rundell (2007) raised what he called the "hardest question", namely "how to fund all this development":

Electronic versions of MLDs have been around for 15 years or so, but none have yet made any money (and they cost a lot to develop). New revenue models need to emerge, and these could include advertising. (Rundell 2007: 50)

Since then, things have only gone downhill. Most publishers of general dictionaries in Western Europe and North America do still not make money from their digital products. MacMillan, for instance, has completely stopped publishing printed dictionaries; cf. Rundell (2014). Instead, it opted for an ad-financed model for its digital products. This solution, which rather looks like a stopgap, entails another unwanted complication, namely the risk of lexicographical data overload; cf. Gouws and Tarp (2017).

The uncomfortable fact is that most users of online dictionaries expect them to be free. They are not ready to pay for this service unless it offers some highly specialized dictionaries that are indispensable for their jobs or studies. In Europe, the unpleasant result is that publishing houses earning money from their online dictionaries can be counted on the fingers of one hand. The few who prosper are mainly lexicographical newcomers who base themselves on a different business model. Most traditional publishing houses seem to be incapable of adapting to the new market conditions. This tendency was also documented by Simonsen (2017) who conducted research into the Danish market. Many of these publishers of high-quality dictionaries have now been forced to close down their business due to dramatically reduced sales, among them famous ones like Longman, Harrap, and Langenscheidt.

Simultaneously, a large number of free dictionaries of dubious quality flourish on the Internet like mushrooms after the rain. This has led to an awkward paradox. On the one hand, modern information-age users need high-quality dictionaries providing quick and reliable information to solve their complex problems and needs. On the other hand, a growing number of these users opt for free-access dictionaries of dubious quality frequently obtaining inadequate and even incorrect information which only adds to their problems.

In some countries, like Spain, no new big general dictionary, either monolingual or bilingual, has seen the light for more than 15 years. Established publishing houses simply do not have money to fund new projects. This reflects a profound crisis which in a certain sense could be described as a crisis of existence for lexicography as we have known it until now. The consequences of all this are potentially disastrous for a society where communication and information is increasingly important.

Rundell's (2007) big challenge therefore remains: Find the money! It is,

however, important not to forget that the lack of an appropriate business model is only the trigger of the current crisis within lexicography. The fundamental cause is the shift of paradigm and the introduction of new disruptive technologies. Hence, the development of a new lexicographical business model must take its point of departure in a profound knowledge of these innovations and their impact on the millennial discipline.

2. A historical vision

Just as in other aspects of life, knowledge of history can prove very useful if one wants to understand the breadth and depth of the current crisis within lexicography. History never repeats itself in a completely identical way, but it nevertheless displays some regularities which, once discovered, can be a great inspiration to understand the present. This is also valid for the relationship between lexicography and technology, a subject that has been treated by many scholars, among them De Schryver (2003), Hanks (2010, 2013), Rundell and Kilgarriff (2011), Nielsen (2013) and Fuertes-Olivera et al. (2018).

Over the years, technology has strongly shaped the development of the five main phases in the practical lexicographical process:

- (1) *determination* of data types to be offered to the future users;
- (2) *retrieval* of raw data from the empirical sources;
- (3) *preparation* of lexicographical data;
- (4) *storing* of data; and
- (5) *presentation* of the lexicographical data in the final product.

Each of these phases has witnessed seismic changes and considerable improvements over the years when new technologies have been introduced. The particular way in which European lexicography was born is inconceivable without the invention of the pen and parchment; cf. Tarp and Gouws (2019). The exponential growth of dictionary output during the past five centuries would not have been possible without the continuous development of the printing and bookbinding technology; cf. Hanks (2010, 2013). The improved quality of lexicographical data during the past fifty years is inseparable from the introduction of digital corpora and the Internet as empirical sources; cf. Hanks (2012), Tarp and Fuertes-Olivera (2016). The lexicographers' present working methods are unthinkable without computers, databases, and e-clouds, etc.

Fuertes-Olivera et al. (2018: 155) have summarized the long-term consequences of the printing technology for lexicography:

Summarily, it can be established that the introduction of the printing technology implied big changes in the production and presentation of the lexicographical product; the empirical basis with the increased use of index cards based on written texts; the design of the dictionary articles with the incorporation of new data categories; the distribution and use of dictionaries; the number of users; the topics treated in dictionaries; and the research areas of scholarly interest. To this

can be added the growing social prestige of lexicographers, some of whom became nationally and internationally famous personalities, as well as the fact that lexicography turned into an increasingly successful business.

The authors conclude that the introduction of the printing technology brought forth an "almost total revolution of the discipline". If we compare all this with the current situation within lexicography, we can see some similarities but also some very interesting differences. There is little doubt that the application of the new disruptive technologies implies an even bigger revolution than the one sparked by the printing technology. As could be expected, it has already led to big changes in the lexicographical product. Fuertes-Olivera et al. (2018: 156), for instance, list "four big transformations" that are going on simultaneously. Among these transformations is the one going from the traditional stand-alone dictionary to a product that is integrated into other information tools, as well as the one going from the dictionary as such to lexicographical data that is handled without appearing in the form of a dictionary. We will later have a closer look at these two phenomena and their possible role as the saviours of lexicography's future. However, we will first look at two other current tendencies.

As we saw above, Fuertes-Olivera et al. (2018) observed that many lexicographers became well-known and famous personalities as a result of printing technology. This was, among other things, due to the fact that many dictionaries were one-man projects, that the printed book format allowed for the authors' names to be put on the cover or front page, and that the dictionaries reached out to a growing number of readers. Today, more or less the opposite is happening. Users who consult online dictionaries will in most cases never see the names of their authors, even if they can be found by means of a link. The brutal fact is that this degrades lexicographers to anonymous data engineers and skilled workers. Obviously, this may affect their self-esteem.

Fuertes-Olivera et al. (2018) also observed that lexicography turned into a successful business after the introduction of printing technology. Today, publishers are struggling to find an appropriate business model and lexicographers are tearing out their hair in frustration when they look for funding for new dictionary projects. The problem in both cases is that an erroneous understanding of the very content of lexicography as a discipline seems to blur the big picture and prevent them from seeing the wood for the trees. A short reflection on the birth of European lexicography will help us to pick up this idea and put it into a future perspective.

3. Back to the roots

Hanks (2013: 507) explains how European lexicography can be traced back to the Classical Greek Period. In the fifth century B.C., it was customary for Greek scribes to insert glosses in manuscript copies in order to explain unusual and obsolete words that appeared in earlier works by Homer and other writers.

Two hundred years later, the glosses were compiled into separate glossaries by scholars at the library in Alexandria (see Figure 1). According to McArthur (1986: 76), historians of lexicography consider that the "origin of the 'dictionary' proper" can be dated back to this practice.

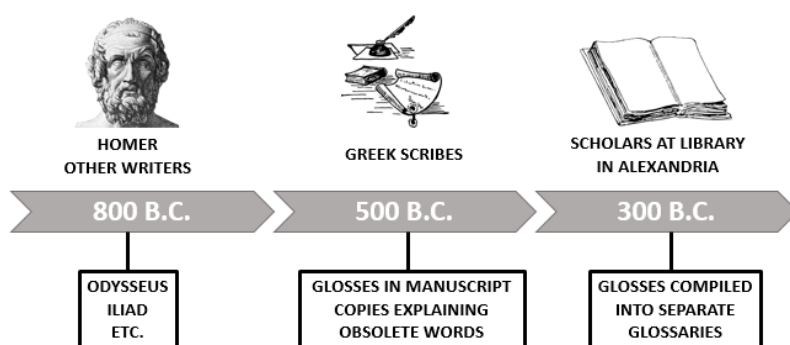


Figure 1: Timeline for the birth of European lexicography

Unfortunately, neither Hanks nor McArthur states clearly when exactly lexicography started. Was it in the fifth or the third century B.C.? This is not a rhetorical question as it has huge consequences for the understanding of the discipline, its content, and its product.

In his famous *Dictionary of the English Language*, Johnson (1755) defines a lexicographer as "a writer of dictionaries" and lexicography as the "art and practice of writing dictionaries". These definitions are repeated in a big number of dictionaries and academic works on lexicography. Nobody denies that the compilation of dictionaries is central to practical lexicography. But is it convincing to exclude other activities from the definition? Why were the scribes' glosses not lexicography? Although the etymology of a word does not necessarily define its modern meaning, it may be relevant to know its origin when working on a timeline. The term "lexicography" is originally Greek and means "writing about the lexicon", precisely what the scribes did. They produced lexicographical data (glosses) addressed to difficult words (glotta) that were later compiled into the glossaries that represent prototype dictionaries in the European tradition.

If the scribes' work cannot be considered part of lexicography, then an increasing amount of work made by 21st-century lexicographers cannot be considered lexicography either. The latter are also producing lexicographical data which in many cases do not end up in dictionaries, whether printed or digital. Their data are stored in lexicographical databases and can be used for multiple purposes. When the Danish company Ordbogen A/S in April 2019 published the first version of its Spanish–English *Write Assistant* (Fisker 2019), its infor-

mation engineers retrieved data from six different digital dictionaries, i.e. six different lexicographical databases, in order to serve this tool. The lexicographers who originally prepared these data have probably no idea of what their data are used for and neither would they be able to recognize them in the new environment. This is not the exception, but the beginning of a new era where publishing houses and other companies are increasingly receiving their revenue from using, handling and selling lexicographical data instead of dictionaries.

It goes without saying that this situation does not add to the lexicographers' self-esteem unless the big picture is grasped and the new paradigm understood in all its complexity. Basically, lexicographers have two options in the long run. Either they accept that their profession comprises far more than the compilation of dictionaries, or they will have to prepare a farewell party for lexicography as a millennial cultural practice because their own work is increasingly presented in forms different from the traditional dictionary.

If they chose the first option, they will undoubtedly find it much easier to go for the money and discern a new business model for their discipline.

4. Upstream in the value chain

As described in Section 2, lexicography has entered a period of financial unsustainability because its hitherto business model has proved obsolete. Only a handful of publishers earn money from their current flagships, i.e. the online dictionaries. This in spite of the fact that these products have tremendous possibilities of improving the traditional dictionary in terms of quantity, quality, updatability, and accessibility. In this light, it is surprising that only few contributions, and even fewer constructive ones, have been published on the subject in the scholarly literature on lexicography. Among the exceptions are Simonsen (2017) and Fuertes-Olivera (2019). The two authors agree that the only way forward is to go upstream in the value chain. The latter writes:

It is necessary to move upstream in the value chain and develop lexicographic services instead of lexicographic products. (Fuertes-Olivera 2019: 25)

This statement points in the right direction although it lacks some accuracy. First, it is not only a question of developing lexicographic services, but also platforms, tools, etc. Second, a lexicographic service is also a lexicographical product, thus a better term for the latter would be "traditional dictionaries" instead of lexicographic products. And third, it cannot be a question of developing these new things "instead of" dictionaries, but in addition to them. Users will still need dictionaries for a long time ahead. The existence of a societal demand for continuously updated high-quality dictionaries cannot be ignored. Thus, the objective of a new business model should be to find financial resources that *also* allow the publishers to sustain their online dictionaries. The dictionaries could then be used by the publishers to promote their different products as well as resources to which users can be referred from the latter.

All this implies that lexicographers and publishers should focus more on new services, digital tools, platforms, etc., (see Figure 2). There seem to be two ways of using their products. The owners of lexicographical data, i.e. databases, can either commercialize these data, or part of them, allowing external service providers, software developers and other stakeholders to make use of them and integrate them in their products, or they themselves can develop new tools, platforms, and services which users are prepared to pay for. In both cases, they should explore all the possibilities and emerging technology breakthroughs mentioned by Schwab (2015) in his reflections on the Fourth Industrial Revolution. This could be robotics, artificial intelligence, etc.

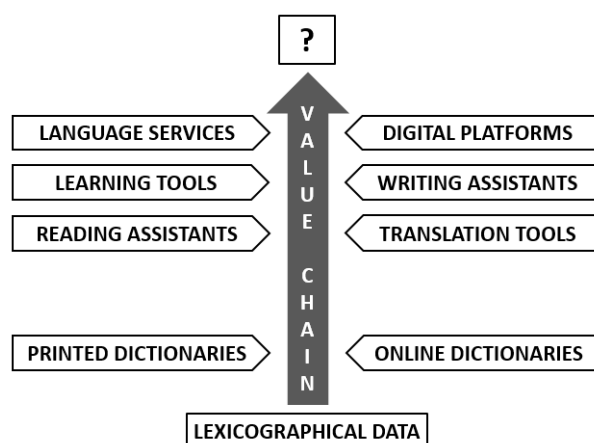


Figure 2: Upstream in the value chain

The move upstream in the value chain cannot be done in a haphazard and amateurish way. The demands of the Fourth Industrial Revolution are enormous, and so are the requirements to future lexicography.

- First of all, it is important to have, or find someone who has, *financial muscles*. The projects will in most cases require considerable investments as the products are expected to be technologically complex and will have to be commercialized on an international scale with many competitors in order to produce revenue.
- Secondly, it is a condition *sine qua non* for success that the work is performed in close *interdisciplinary collaboration* between lexicographers, information engineers, industrial designers, and other relevant specialists. The days with homemade databases are long over. Things have to be done on an industrial scale with the most qualified players.
- Thirdly, a good dose of *creativity* is needed. The collaboration between experts should not be misunderstood as a sum of the different expertise

and skills. An untimely each-of-us-in-our-own-garden attitude would only lead to "linear solutions without a broad perspective" as stated by Steve Jobs at the top of this article. Instead, the collaboration should be carried out as a brainstorm-like interdisciplinary confrontation of ideas that allows the experts to "synthesize new things".

Whatever the visions are, it may prove difficult to concretize a new lexicographical business model without the integration of these three basic requirements. And they will definitely not be the only ones.

5. Recent developments

It may seem strange to start a section dedicated to recent developments in lexicography with another visit to the old library in Alexandria. It will nevertheless turn out to be very useful. When the scholars at the library collected the first glossaries more than two thousand years ago, they introduced two important innovations that may help us to throw light on current and future trends in lexicography. The two inventions are the lexicographical article and the dictionary format, respectively.

The first articles were rather simple. Apart from the gloss, they consisted of a "glotta", the Greek for a difficult word, to which the former was addressed (McArthur 1986: 76). The glotta is equivalent to the lemma in modern lexicography whereas the gloss represents the lexicographical data in embryonic form. This is all it takes to constitute a lexicographical article which later became increasingly sophisticated with much more data that were structured in different ways (microstructure).

The second invention was the dictionary format as a collection of lexicographical articles that are structured according to one or another principle by means of the lemmata (macrostructure).

Even though the old glossaries and their articles may seem far from modern standards, they nevertheless represent prototypes of the dictionary format and overall article structure that have survived until our time. During the past two decades, this tradition has creepingly been challenged as a result of the new digital forms of presentation of the lexicographical product.

The first victim to be sacrificed was the dictionary as a collection of articles with a macrostructure. A modern online dictionary consists of a number of articles. The user can explore whether a certain word has been lemmatized and honoured with its own article. However, in most cases, he or she cannot get an overview of all the articles contained in the dictionary. A few years back, many digital dictionaries could be accessed through a separate alphabetic list which allowed the user to embrace all the lemmata treated. This practice has been abandoned in most online dictionaries today. It implies that the dictionary has lost its character of a collection of articles in the sense that the user cannot get a clear sight of this collection as it was the case with the printed dictionary. In addition, it also means that the articles are no longer displayed to the user in

the framework of a specific structure. Any talk of macrostructures in such dictionaries is therefore pure nonsense.

So far, the second victim has only been partially sacrificed. Online dictionaries accommodate articles with lemma, data, and microstructure. But the former static structure has increasingly been replaced by a dynamic structure that adapts to different user needs in different types of consultation. This implies that the amount and organisation of the displayed lexicographical data are fluctuating. And to this should be added that many dictionary articles have been broken up and require clicks, scrolling down and other techniques to be visualized in their totality.

As we will see in the following chapters, this sacrifice of time-honoured lexicographical traditions can be expected to further accelerate in the nearby future. Yet, the various mutations do not change the fact that we can still talk about lexicographical products. If anything, they should rather be viewed as a natural and necessary adaptation to the new digital environment.

6. Visions and functions of *Write Assistant*

Write Assistant, the series of bilingual writing assistants developed by the Danish company Ordbogen A/S, is a good example of both the new possibilities and the new challenges which lexicography experiences today. Its functionality has been extensively explained by Tarp et al. (2017). Hence, in this section we will only discuss some aspects that are either new or relevant to our topic.

Write Assistant is designed with only one main function as defined by Function Theory, namely to assist its users when writing in a second language; cf. Tarp (2008). The basic idea was conceived in a remote Swedish farm where five people with different backgrounds gathered during a seven-days brainstorm-like session. As such, it is not born out of traditional user research, but of disruptive thinking.

The underpinning philosophy is based on two important *observations*. The first one is that most writing today is performed on smartphones, tablets, and laptops. Writing skills are mainly used in connection with these devices whereas handwriting is increasingly restricted to our personal use. Dictionary consultation to solve text-production problems is therefore almost exclusively done when writing on these devices. First conclusion: Lexicographical assistance should be available directly on smartphones, tablets, and laptops.

The second observation is that people waste too much time when consulting external sources, even if these sources are available on the mentioned devices. This may affect their writing flow, focus and concentration, and in some cases they may even forget what they were writing about after an excursion to external sources. Second conclusion: Lexicographical and other types of assistance should be integrated in other tools and made available directly in the documents and texts which people write in Word, Outlook, Gmail, PowerPoint, Excell, WhatsApp, Messenger, WeChat, Explorer, Safari, Chrome, Firefox, Facebook, Twitter, and a long etcetera.

Write Assistant has therefore been designed as an application that can be downloaded to the user's device and provide instantaneous assistance when the user is writing a text in a foreign language. The tool connects its users to *big data* taken in from two empirical sources, an L2 digital corpus and a lexicographical database. Its driving power is a *language model* that has been trained on the corpus, originally using statistic programming, but now *artificial intelligence* is increasingly being incorporated. As such, *Write Assistant* makes extensive use of lexicographical data that are imported from different sources and even generated in different ways (see Figure 3).

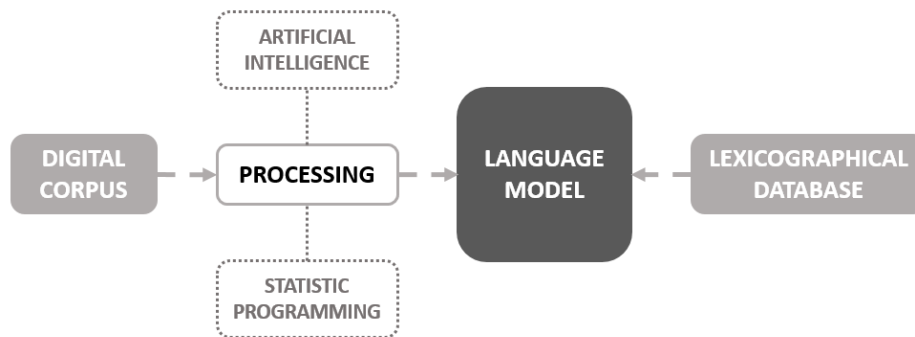


Figure 3: The language model and its empirical sources

When fully developed, the assistance will be provided to the users in three different windows popping up in the document on which they are working, i.e. the suggestion, consultation and alert windows, respectively (see Figure 4).

The suggestion window appears as default on the user's screen and offers *L2-word completions* and *next-words* when the writer types one or more letters or a full word in L2. These suggestions are generated automatically by the language model and are ephemeral in the sense that they can only be recreated if the users write exactly the same sequence of words. In addition, the suggestion window also offers *L2 equivalents* to L1 words typed by the users. These equivalents are fed by the lexicographical database and presented in a prioritized, context-aware order by means of the language model.

If the writer does not know which of the suggested words to use, or how to use it, a simple click or touch will activate the consultation window which allows the writer to access lexicographical data such as meaning, inflection, grammar, synonyms, etc. This window is also, in a certain sense, lexicography's window to the future. The third window, which has still to be designed, is foreseen to provide *alerts* that are only activated when the users write a word which the designers of the tool deem to be linguistically or culturally problematic.

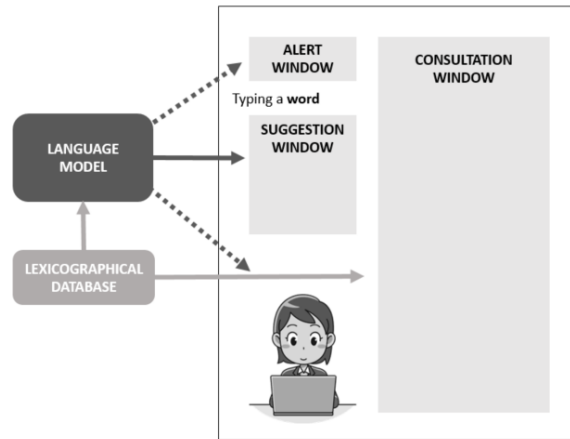


Figure 4: Write Assistant's three data windows

The different data types provided in the three windows pose some conceptual problems to lexicography. It is evident that the data presented in the consultation and alert windows, as well as the L2 equivalents in the suggestion window, are lexicographical as they have been (or will be) prepared by lexicographers and stored in the database. But what about the L2-word completions and next-words provided in the suggestion window? These data are generated automatically by the language model and, as such, they are not the result of a human lexicographer's work. Can they be considered lexicographical data? Is the intervention of a human lexicographer a prerequisite for data to be considered lexicographical?

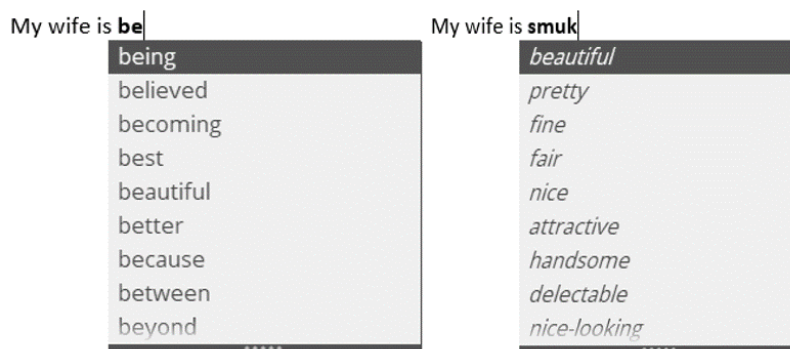


Figure 5: Suggestion windows with L2-word completions and L2 equivalents

Figure 5 shows the suggestion window with two different types of data. The one to the left appears when the writer types "my wife is **be**..." and contains a list of nine likely L2 *completions* of "be", whereas the one to the right provides nine L2 *equivalents* after the writer has typed the L1 word *smuk*. Apart from this, there is no difference in content between the two windows; both contain nine English words listed in a prioritized, context-aware order. The only difference is how the words are generated. In this perspective, Rundell and Kilgarriff (2011: 278) write:

We envisage a change from the current situation, where the corpus software [...] presents data to the lexicographer in [...] intelligently pre-digested form, to a new paradigm where the software selects what it believes to be relevant data and actually populates the appropriate fields in the dictionary database.

The two authors leave no doubt that the data generated in this way are lexicographical although the process is completely automatic and performed without passing through the human eye of the needle. The only difference between their example and the one discussed above is that *Write Assistant* takes the process even further and offers the automatically created data directly to the end-user.

7. Lexicographical challenges

The gradual incorporation of artificial intelligence into *Write Assistant* gives rise to many incognitos. To what extent will deep learning be able to improve the tool? Although success seems to be guaranteed along the main lines, artificial intelligence may not succeed to honour all expectations. Will it, for instance, be able to convert vulgar language into formal language? Will it be optimized to the point where it can convert a first-year student's clumsy text into high-standard academic writing? And if it eventually will be able to do this, how many years will it take until the dream becomes reality?

In spite of these incognitos, in the eye of a lexicographer (there may be other eyes), the lexicographical ingredient represents currently *Write Assistant's* Achilles heel if the vision, as it is, is to turn it into a premium tool. In this perspective, the three different types of window discussed above pose a number of challenges to lexicography, among which can be mentioned:

- Which lexicographical data do writing assistants with the described characteristics require?
- Which are the words requiring linguistic and cultural alerts? And which are the lexicographical data needed to support these alerts?
- Which lexicographical data are required if *Write Assistant*, apart from assisting L2 writing, should also be designed as a learning tool. (It transpires that many learners want to use it with this purpose.)

- How can traditionally prepared lexicographical data interact with data that are automatically generated by means of artificial intelligence?
- How can *Write Assistant* balance *data pushing* and *data pulling* procedures so the users do not lose their responsibility and feel that they have been cornered with the suggestions popping up on their screen? (Within information science, data pushing is defined as a situation where the sender decides on the data to be pushed towards the user, whereas data pulling implies that the receiver can decide what to receive; cf. Gouws (2018). Any decision in this regard may have important consequences for language didactics and learning.)

Many more questions could be asked in this connection. The future lexicographical improvement of *Write Assistant*, for example, will also have big consequences for the lexicographical databases used to sustain it. The experience so far indicates that existing databases are highly deficient and problematic when they are used to feed the tool. Either they do not contain the data required to feed *Write Assistant*; or they do not have them in the necessary quantity; or these data are stored in the database in such a way that they cannot be used properly. As a result, it was necessary to import data from six different digital dictionaries to feed the Spanish–English version of *Write Assistant*, while the German–English version required data from three different sources. Even so, a lot of challenges remain unsolved.

Moreover, once the required data types have been determined and an adequate database designed, another big challenge is posed to lexicography:

- How should the relatively big amount of lexicographical data required to meet the user's consultative needs be presented and structured in the relatively small consultation window without creating data overload with a too long and user-unfriendly access route?

If Shakespeare were still alive, he would have repeated the famous words: "That's the question". In the next section, a possible answer will be discussed.

8. The window to the future

The following is not necessarily what is going to happen. The proposal is based on several open and forthright discussions and exchange of ideas with experts from other fields. So far it is exclusively a proposal developed from the side of lexicography, and it only focusses on the consultation window when it is accessed from the suggestion window, i.e. excluding access from a future alert window (see Figure 4). As such, it reflects a possible "laboratory" or working method that can be used in this and similar projects.

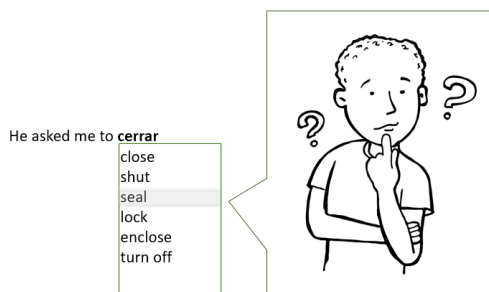


Figure 6: The challenge: How to fill the consultation window?

Figure 6 shows the point of departure. A Spanish user writes "He asked me to" and continues with the Spanish word *cerrar* because he or she is not sure which English word to use. A list of likely English equivalents are immediately furnished in the suggestion window. The writer does not know the meaning of "seal" and therefore activates the consultation window clicking on "seal". Thus, the challenge is how to fill this window with a view of meeting the user's concrete needs. A possible solution can be seen in Figure 6.

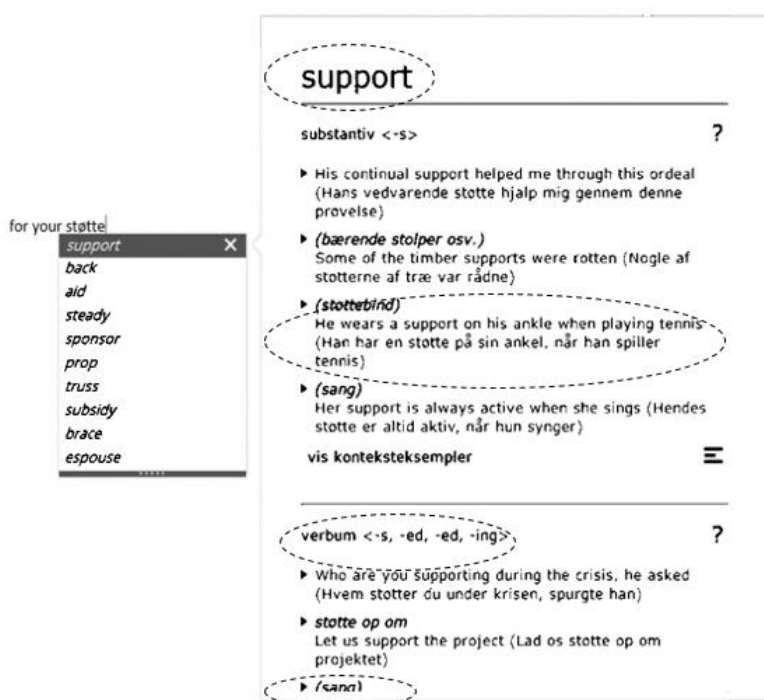


Figure 7: The original Danish–English beta version

Figure 7 is a screenshot from the original beta version of the Danish–English *Write Assistant* where an article from one of Ordbogen's existing online dictionaries has been used to fill the consultation window. This solution has several problems that make it inapt for a tool like *Write Assistant* if we take into consideration that it should be as easy as possible for the user to find what he or she needs. This requires a careful selection of the data to be presented as default. The design of the article in Figure 6 does not help in this regard.

Firstly, the lemma, apart from the big letters, seems to be completely redundant as the user perfectly well knows from which word the article has been accessed. Secondly, the explanations are given in the form of example sentences. This may not be the best solution as it requires a complex and possibly time-consuming mental process to deduce the meaning. Thirdly, the example sentences are provided in both L1 and L2. This is also a space robber that causes the tool to hide several example sentences so they require a further step to be visualized (*vis konteksteksempler*). A better solution would be to furnish an L1 sentence as default with the option to expand it with its English translation. Fourthly, the article offers both part of speech and inflection. These data may also be relevant, but it is probably not the first class of information demanded by the user who, in most cases, is expected to start the consultation process because of comprehension problems. Lastly, the considerable amount of excess data, at least in this consultation phase, occupies space and compels the user to scroll down in order to get more relevant information on meaning.

As mentioned above, it became necessary to import data from six different sources in order to feed the first version of the Spanish–English *Write Assistant*, and even so the result was not satisfactory. Ordbogen A/S is, therefore, preparing a premium English–Spanish–English lexicographical database that is currently being made at the International Centre for Lexicography at the University of Valladolid with Pedro Fuertes-Olivera as the main editor; cf. Fuertes-Olivera et al. (2018).

A print of the word "seal" in this database takes up nine A4 pages, containing three lemmas (one verb and two nouns with different inflection paradigms) with a total of 19 senses and a large number of lexicographical data addressed to each sense (see Figure 8). Thus, the big challenge is now to elegantly put all these data into the consultation window without incurring in data overload and forcing the user to take too many steps before getting the needed information.

In the following, an alternative design of the consultation window will be proposed. It is based on the criterion of relevance (cf. Bothma and Tarp 2014). The most relevant data are provided first (as far as it is possible) and the less relevant have to be accessed through further steps. This implies, among other things, that navigation techniques like scrolling down and sweeping to the sides, when they cannot be avoided, are reduced to a minimum. The guiding principles are "less is more" and "simplicity is the ultimate sophistication". Figure 8 shows the initial attempt in this spirit.

- seal** verbo
seal, seals, sealed, sealing
<https://howjsay.com/search?word=seal> pronunciación
1. cerrar algo, por ejemplo un sobre; se refiere a la actividad física con la que alguien mete un documento en un sobre y lo cierra
valencia
 - someone seals something
 - alguien sella algoNota de uso
 - Verbo transitivo, es decir usado con un complemento directo.sinónimo
 - seal up
 - close upequivalente
 - sellar verbocolocaciones
 - she folded it, sealed the envelope, and ran to the postbox
 - lo dobló, cerró el sobre y corrió al buzón de correos
 2. cerrar algo y colocar un precinto como medida de seguridad, es decir para informar que alguien ha abierto ese algo sin permiso; ese precinto es un sello de cera o algo similar
valencia
 - someone seals something
 - alguien lacra algoNota de uso
 - Verbo transitivo, es decir usado con un complemento directo.sinónimo
 - fasten
 - secureequivalente
 - lacrar verbocolocaciones
 - the civil servant sealed it with the king's seal
 - el funcionario lo lacró con el sello real
 3. cerrar algo, por ejemplo un contenedor, una puerta, una tumba, etc. para que no pueda abrirse; se hace como medida de seguridad
valencia
 - someone seals something
 - alguien sella algoNota de uso
 - Verbo transitivo, es decir usado con un complemento directo.sinónimo
 - shut
 - close
 - stop upcolocaciones
 - the container was sealed and nobody could open it up again
 - the container was sealed and nobody could open it up againcerrar
 - sealed the wood before applying paint
 - sellaron la madera antes de pintarejemplo
 - She merely filled the containers, sealed them with a cork, and pasted on labels.
 - Simplemente cerró los contenedores, los selló con un corcho y puso las etiquetas.
 4. cerrar algo de forma hermética para que no entre ni agua ni aire
valencia
 - someone seals something
 - alguien impermeabiliza algoNota de uso
 - Verbo transitivo, es decir usado con un complemento directo.sinónimo

Figure 8: Extract from English–Spanish database in the Valladolid-Uva project

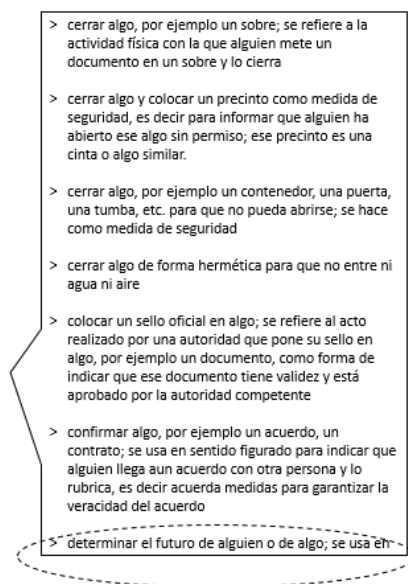


Figure 9: Design of the suggestion window based on relevance

In Figure 9, the definitions of the various senses of the verb "seal" have been imported from the English–Spanish database shown in Figure 8. Here the user gets what is considered most relevant when using *Write Assistant*, namely the meaning of the various senses of "seal". However, the database contains nine senses with the verb "seal" and this implies that the user has to scroll down in order to see the remaining senses. (Some common words in the database have 20, 30 or even more senses). Hence, an alternative solution is shown in the left screenshot in Figure 10. In this case, all the definitions are cut down to only one line allowing for all nine senses of "seal" to be displayed immediately. The idea is that the user in this way can get a preliminary idea of the meaning of each sense, so to say "smell" it. If it "smells" good, it can be expanded with a simple click on ">", whereupon the whole definition will be displayed as shown in the right screenshot in Figure 10.

The solution drafted in Figure 10 creates other problems. If short one-line definitions are not foreseen and prepared for the database, many sentences will be cut in the middle. Although this solution could be recommended as interim, it is nevertheless disturbing and could imply that the user would find it difficult to deduce the preliminary meaning from the abridged definitions to the left in the figure. In the long run, the inclusion of both short and longer definitions into the database will, therefore, have to be planned from scratch as already indicated by Fuertes-Olivera et al. (2018: 159-160).

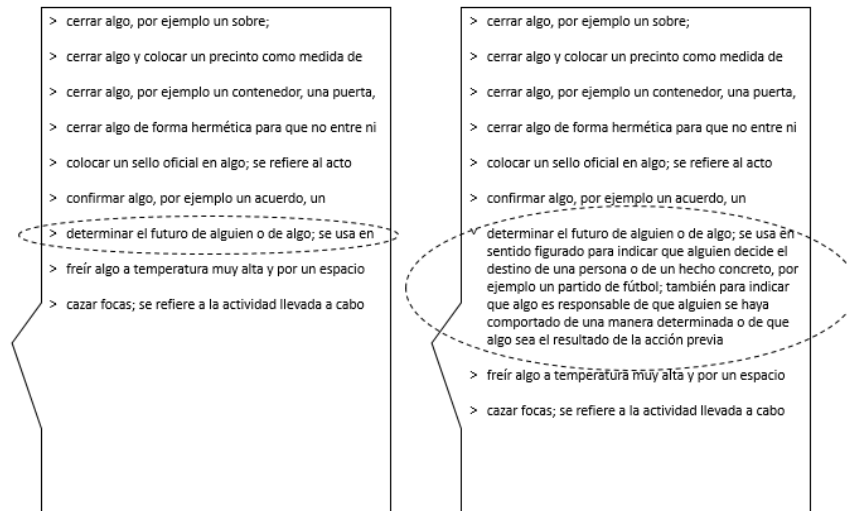


Figure 10: Compact one-line definition and expanded definition

Another inconvenience in Figure 10 is that the expanded definition is placed in the middle of the other one-line definitions. This leaves little room for other data types that could invite the user to a third step in the consultation process. A solution to this problem could be that the remaining one-line definitions are hidden and the expanded one placed at the top of the consultation window as shown in Figure 11.

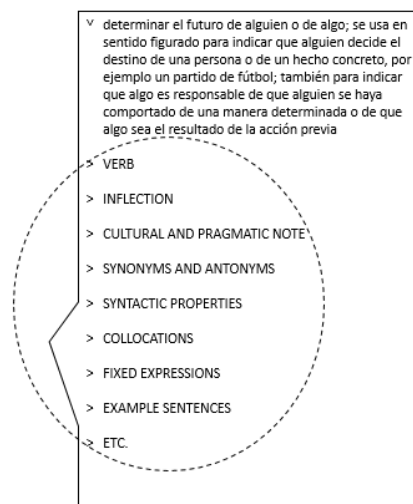


Figure 11: Article with metatexts for further access

The proposal in Figure 11 also includes a list of so-called metatexts that need further elaboration. The idea is that a third click on one of the metatexts allows the user to access additional data that may be needed in a concrete consultation. The order of the metatexts, as well as alternative ways of presenting them, will be discussed in Section 9. As we will see, at this point it does not really matter.

Much more important is it that Figure 11 seems to represent a completely new type of lexicographical "article". First of all, it transpires that all other data, with no exception, are addressed directly to the definition of each sense. What does this mean for the article structure? Has the definition gone through a metamorphosis and converted into a lemma? Or could it be claimed that the real lemma is the word in the suggestion window from where the default article in the consultation window is accessed? But where is then the lemma to the article with L2 words and equivalents listed in the suggestion window? Is it the word — or the letters — typed by the user?

Instead of becoming inquisitive and risking the return of the Cretan Labyrinth, it would be easier to ask the old Greek scribes for an answer. The fact is that the modern "scribes" are now doing something very similar to what their Greek predecessors did before the scholars in Alexandria started compiling glossaries. When they inserted glosses into manuscript copies of old texts, they did it directly in the context where a problem might occur and without the need to invent lemmas and all those frozen article structures that characterized the printed book. They so to say contextualized the lexicographical data; cf. Tarp and Gouws (2019).

The only real difference between the old and the new scribes in this respect is that the latter are doing something much more sophisticated based on disruptive technologies that are light years away from the scribes' pen and parchment.

9. Intermezzo on lexicographical user research

For people engaged in the development of *Write Assistant*, it is obvious that the challenges posed by this and similar disruptive tools cannot be solved by means of traditional user research. In 1982, while building the Macintosh computer, a young Steve Jobs was asked whether he thought it was necessary to do some market research to see what customers wanted. His straightforward answer was:

No, because customers don't know what they want until we have shown them. (Isaacson 2011: 143)

This reflection is still relevant almost forty years later. Current lexicographical user research does not make a dent in the universe. It is generally conducted into already published dictionaries or so-called prototypes with no prospect of being produced due to financial constraints. Its results are most often pub-

lished months or even years after the research has been conducted. The real needs of users as they express themselves *before* dictionaries are consulted are largely ignored. It is therefore of little relevance for the design of a completely new tool aspiring to prove its *raison d'être* in the era of the Fourth Industrial Revolution.

Besides, it should not be forgotten that each dictionary has its own personality. It is problematic to generalize from one type of dictionary to another, from one type of user to another, from one language to another, from one culture to another. Although there are general observations that are valid for the consultation of dictionaries in general, concrete instructions for a concrete dictionary can only be based on research into the usage of this particular dictionary. Anything else is shoddy. General instructions based on general observations only have little practical value and may even produce the opposite result when applied to a concrete dictionary.

The users' immediate reaction to innovations does not have to be the supreme court of truth. When the proposal to remove the lemma from the default consultation window (see Figure 6) was presented, it immediately raised a discussion of whether this radical step should be tested among future users. This position was opposed from the side of lexicography. The argument was that 1) most users are conservative and would prefer what they are used to; 2) *Write Assistant's* main target group is the young generation and they will get used to it within a few hours or days; and 3) the removal of excessive data would benefit the overall design of the consultation window. The situation resembles the one that took place when it was decided to eliminate the cursor arrow keys on the Macintosh keyboard in order to force "old-fashioned users to adapt to point-and-click navigation" with the mouse (Isaacson 2011: 138). The result can be seen today where the mouse has become like a pet for many people.

Regrettably, many lexicographers seem intent on wasting their scarce research time on matters which are fast becoming obsolete, instead of applying their minds to the myriad of challenges posed to their discipline by the Fourth Industrial Revolution. The current technological breakthroughs require above all that more time is dedicated to immersion, reflection, deduction, brainstorm, and interdisciplinary confrontation of ideas.

This recommendation by no means denies the need for user research that can contribute to the confirmation, modification, and/or adjustment of new products. In Figure 10, a number of metatexts in random order were introduced. Two questions immediately arise: How should the metatexts be formulated? And in which order should they be presented? The answers can be achieved by means of a different type of user research, namely the use of iterative processes with *test-driven development* (TDD).

It goes more or less like this: A beta version of *Write Assistant* with metatexts included is provided to a small focus group of 10–15 learners. They are then observed using it during a certain time, e.g. an hour, and subsequently interviewed in order to collect their opinions and suggestions. Upon this basis,

a new version with adjusted metatexts is prepared and tested among a bigger group, e.g. 50 learners, who are also observed and interviewed. This leads to a new adjustment that is tested among an even bigger focus group, and so forth.

Simultaneously, a test is conducted to see which metatexts are most frequently activated by the users so they can be arranged in prioritized order according to their relevance. The same kind of iterative processes are used to test other lexicographically relevant aspects, e.g. how *Write Assistant's* users handle the compact one-line definitions shown in Figure 9, or how the additional lexicographical data accessed through the metatexts could be arranged in the best possible way.

This type of (lexicographical) user research is conducted by professional testers supported by a team of information engineers who implement the adjustments straight away. The whole process normally takes a few weeks until a satisfactory result is achieved. As already mentioned, such a test-driven development of the lexicographical product is in every way much more appropriate in the era of the Fourth Industrial Revolution.

10. Perspectives

If we take a panoramic look at the *Write Assistant* experience as well as other developments taking place in present-day lexicography, we will see the contours of a future transformed discipline characterized by the interaction of the traditional human lexicographers with their modern counterparts, the *artificial lexicographers*. This interaction involves at least four of the five main phases of the lexicographical process introduced in Section 2, i.e. the determination, retrieval, preparation, storing, and presentation of lexicographical data.

Figure 12 illustrates how the artificial lexicographer is encroaching on its human counterpart's traditional domain. We already have *example sentences* and other raw data taken automatically from empirical sources and presented to lexicographers for treatment. As discussed in Section 6, we also have lexicographical data taken from empirical sources, processed and presented *directly* to the users *without the interference* of human lexicographers and *without being stored* in the database. In this respect, a new distinction between *lasting* and *ephemeral* lexicographical data is essential.

With the increasing use of artificial intelligence, we will soon have *partially prepared* data presented to human lexicographers for a final touch, and even *fully prepared* lexicographical data stored directly in databases without the interference of human lexicographers.

The trend is unstoppable. There is little doubt that the introduction of artificial intelligence will accelerate it even more. Hence, we can expect that the part of lexicographical data bypassing the human eye will increase over the next years. All this raises the question of the future role of human lexicographers. To what extent will the human eye be required in the future?

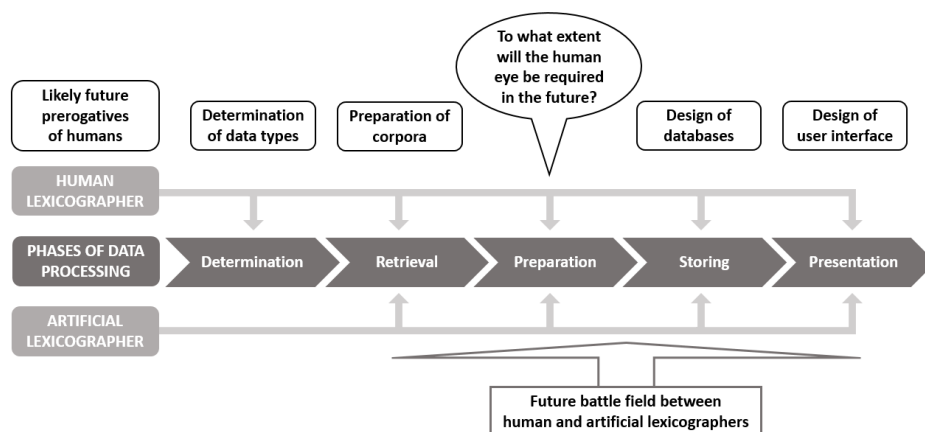


Figure 12: Interaction between human and artificial lexicographer

Figure 12 gives an overall view of the distribution of tasks between the human and the artificial lexicographer in terms of the five phases mentioned above. Any snapshot of this distribution will soon become obsolete due to the current speed of technological breakthroughs. Even so, it seems likely that the human lexicographer, at least for some time ahead, will continue with a number of prerogatives. These prerogatives include the determination of the data types to be presented to the users, the preparation of corpora, as well as the design of databases and user interfaces which cannot be designed competently without the lexicographical criterium. In addition — and until a software solution capable of deducing meaning from the context is developed — a human lexicographer will still be required to separate meaning into senses, write definitions, select equivalents, attach example sentences and other data to the right sense, write cultural and pragmatic notes as well as other explicit data, for instance, on syntax.

Hence, although future surprises are unavoidable, there is no reason for human lexicographers to despair and be anxious about the future.

At the dawn of the computer age, Dreyfus and Dreyfus (1986) challenged those colleagues who were over-optimistic on behalf of the new technology. They opposed the human mind to the machine with their 5-step model for skills acquisition that was topped by the level of genuine, human expertise characterized by intuition, virtuosity, and effortless performance. Since then, much water has flowed under the bridge "blurring the lines between the physical, digital, and biological spheres" (Schwab 2015). This development, however, should not be seen as an invitation to human lexicographers to retire and externalize their skills and knowledge to the machines. It should rather be seen as a wake-up call to constantly outdoing themselves and proving their worth as an indispensable counterpart to the artificial lexicographers. The newcomers do

not deserve to be met with a frontier wall but should be welcomed with qualified integration for the benefit of their joint users.

Acknowledgments

The main ideas of this article were hatched during a research stay in May-June 2019 at the Sino-Danish Sindberg Centre for Lexicography, Translation and Business Communication at Guangdong University of Finance, China. Special thanks to its director, prof. Heming Yong, for inviting me and giving me this rare opportunity to dedicate time to reflection, immersion, and contemplation.

Special thanks to prof. Rufus H. Gouws, Stellenbosch University, and prof. Pedro A. Fuertes-Olivera, University of Valladolid, for their comments and constructive input.

Special thanks to the *Write Assistant* team at Ordbogen A/S for always being willing to explain, discuss and exchange ideas.

Special thanks also to the Spanish Ministry of Economy and Competition for financial support (grant FFI2014-52462-P).

References

- Alonso-Ramos, M. and M. García-Salido. 2019. Testing the Use of a Collocation Retrieval Tool Without Prior Training by Learners of Spanish. *International Journal of Lexicography* 32(4): 1-18. <https://academic.oup.com/ijl/advance-article/doi/10.1093/ijl/ecz016/5525336>. (Accessed 17 July 2019.)
- Bothma, T.J.D. and S. Tarp. 2014. Why Relevance Theory Is Relevant for Lexicography. *Lexicographica* 30: 350-378.
- De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Dreyfus, H. and S. Dreyfus. 1986. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Fisker, K. (Ed.). 2019. *Write Assistant. Spanish-English*. Odense: Ordbogen A/S. Available at: www.writeassistant.com/en/.
- Frankenberg-García, A. 2018. Combining User Needs, Lexicographic Data and Digital Writing Environments. *Language Teaching*: 1-16. <https://doi.org/10.1017/S0261444818000277>. (Accessed 17 July 2019.)
- 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. 2019. Designing and Making Commercially Driven Integrated Dictionary Portals: the Dictionarios Valladolid-Uva. *Lexicography* 6(1): 21-41. doi.org/10.1007/s40607-019-00056-8. (Accessed 17 July 2019.)
- Fuertes-Olivera, P.A., S. Tarp and P. Sepstrup. 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Dictionarios Valladolid-Uva. *Lexikos* 28: 152-176.

- Gouws, R.H.** 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- 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.
- Hanks, P.** 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Keith, A. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.
- Isaacson, W.** 2011. *Steve Jobs. The Exclusive Biography*. New York: Simon & Schuster.
- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton.
- McArthur, T.** 1986. *Worlds of Reference*. Cambridge: Cambridge University Press.
- Nielsen, S.** 2013. The Future of Dictionaries, Dictionaries of the Future. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 355-372. London/New York: Bloomsbury.
- Rundell, M.** 2007. The Dictionary of the Future. Granger, S. (Ed.). 2007. *Optimizing the Role of Language in Technology-enhanced Learning*. Proceedings of the expert workshop organized in Louvain-la-Neuve (Belgium), 4–5 October 2007: 49-51. <https://hal.archives-ouvertes.fr/hal-00197203/document/>. (Accessed 17 July 2019.)
- Rundell, M.** 2014. Macmillan English Dictionary: The End of Print? *Slovenščina 2.0* 2(2): 1-14.
- 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.
- Schwab, K.** 2015. The Fourth Industrial Revolution: What It Means and How to Respond. *Foreign Affairs*. 12th December, 2015. www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution. (Accessed 12 July 2019.)
- Simonsen, H.K.** 2017. Lexicography: What is the Business Model? Kosem, I., C. Tiberius, M. Jakubíček, J. Kallas, S. Krek and V. Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2017 Conference, 19–21 September 2017, Leiden, the Netherlands*: 395-415. Brno: Lexical Computing CZ, 395415. elex.link/elex2017/proceedings-download/. (Accessed 12 July 2019.)
- 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 P.A. Fuertes-Olivera.** 2016. Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-Uva. *Lexikos* 26: 273-296.
- Tarp, S. and R.H. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.
- Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
- Wolf, G.** 1996. Steve Jobs: The Next Insanely Great Thing. *Wired*, 1st February, 1996. www.wired.com/1996/02/jobs-2/. (Accessed 17 July 2019.)

Lexicographical Contextualization and Personalization: A New Perspective

Sven Tarp, *Department of Afrikaans and Dutch, Stellenbosch University, South Africa; Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; International Centre for Lexicography, Universidad de Valladolid, Spain; and Centre for Lexicography, University of Aarhus, Denmark (st@cc.au.dk)*

and

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

Abstract: Contextualization, i.e. to provide solutions to users' information needs directly in the situation or context where these needs occur, played a significant role in the work of the Greek scribes who inserted glosses into manuscript copies of the works of Homer and other earlier writers in order to explain obsolete and unusual words. After the invention of the glossary, a schism developed within lexicography. On the one hand, there was the new compilation of glossaries and dictionaries of a still more complex and sophisticated nature. On the other hand, there was the traditional insertion of glosses into manuscript copies of books from previous periods. Although the advent of dictionaries diminished the use of contextualization procedures they were still adhered to in some publications. This paper discusses the occurrence of contextualization and personalization procedures in different eras and environments and it is shown how these procedures also introduced a lexicographic practice to some extra-dictionary environments. The importance of contextualization and personalization in modern-day lexicography is stressed. Lexicographers often have had unfulfilled dreams of new possibilities within the digital environment. However, the lack of adequate technology has made their dreams impossible — at that stage. Today new technologies and collaboration between lexicography and information science offer numerous new challenges that can be met by lexicographers. It is shown how lexicographical products being integrated into information tools little by little are closing the more than two thousand year old schism in European lexicography, i.e. reuniting contextualization and personalization. Lexicographers, the modern-day scribes, have to endeavour to make the seemingly impossible possible.

Keywords: CONTEXTUALIZATION, E-READER, EXTRA-DICTIONARY, GLOSS, GLOSSARY, INFORMATION SCIENCE, INFORMATION TOOLS, PERSONALIZATION, SCRIBES, WRITING ASSISTANT

Opsomming: Leksikografiese kontekstualisering en verpersoonliking: 'n Nuwe perspektief. Kontekstualisering, dit is om oplossings vir gebruikers se inligtingsbehoefes te verskaf presies in die situasie en konteks waar die behoefte ontstaan, het 'n belangrike rol gespeel in die werk van die Griekse skrywers wat glosse in die manuskripte van kopieë van Homeros

en ander vroeë skrywers gevoeg het om onbekende en ongewone woorde te verduidelik. Na die ontwikkeling van die glossarium het 'n kloof binne leksikografie ontstaan. Enersyds was daar die samestelling van glossaria en woordeboeke van 'n nog meer gesofistikeerde aard. Andersyds was daar die tradisionele invoeging van glosse in ouer manuskripte. Alhoewel die koms van woordeboeke die gebruik van kontekstualisering verminder het, is dié werkswyse steeds in sommige publikasies gebruik. Hierdie artikel bespreek die voorkoms van kontekstualisering en verpersoonliking in verskillende eras en omgewings en daar word aangetoon hoe hierdie werkswyse 'n leksikografiese aard in sommige buite-woordeboekomgewings ingelui het. Die belang van kontekstualisering en verpersoonliking in die moderne leksikografie word benadruk. Leksikograwe het dikwels onvervalde drome gehad van nuwe moontlikhede in die digitale omgewing. Die tekort aan toereikende tegnologie het hulle drome in daardie stadium 'n onmoontlikheid gemaak. Vandag bied nuwe tegnologie en samewerking tussen leksikografie en inligtingswetenskap talle uitdagings wat leksikograwe kan oorkom. Dit word aangetoon hoe leksikografiese produkte in inligtingswerktuie geïntegreer word en geleidelik word die kloof wat reeds meer as tweeduisend jaar in die Europese leksikografie bestaan, oorbrug deur 'n hereniging van kontekstualisering en verpersoonliking. Leksikograwe, die moderne skribas, moet poog om die oënskynlik onmoontlike moontlik te maak.

Slutelwoorde: BUIE-WOORDEBOEK, E-LESER, GLOS, GLOSSARIUM, INLIGTINGSWERKTUIE, INLIGTINGSWETENSKAP, KONTEKSTUALISERING, SKRIBAS, SKRYFHULP, VERPERSOONLIKING

0. Introduction

History is a funny thing. It develops through bumps and jumps, most of the time forwards but frequently also backwards. Sometimes it even tends to repeat itself, in most cases as a tragedy. Other times specific phenomena appear in a specific historical period without us humans paying much or any attention to them for years or even centuries. The phenomena seem to be of little relevance and are therefore left to an anonymous existence until they, in one of those tricky turnabouts in history, suddenly are brought into the limelight although they, as phenomena, are much older than those who promote them as new and revolutionary. The concepts of contextualization and personalization within lexicography are symptomatic of phenomena that have shared such a destiny.

In a lexicographical — and information science — perspective *contextualization means*, in addition to its traditional linguistic sense, *to provide solutions to users' information needs directly in the situation or context where these needs occur*. The modern GPS is an example of such a contextualized service. When a person is driving and has to turn right in order to continue to the fixed destination, the GPS voice gives the instruction to turn right say 100 m before reaching the crossroad. Another driver on their way to the same destination who is 200 m behind the first driver will not get this instruction until a few seconds later. The GPS instructions are contextualized because they are delivered in the exact moment and place where they are needed.

Personalization is somehow related to contextualization but nonetheless different in nature. It means that the provider of information in one way or another knows when a particular user has or is expected to have an information need and also how to prepare the corresponding lexicographical data and adapt them to the user's particular profile so he or she can effectively make use of them in order to meet his or her particular need. It is a matter of course that personalization does not exclude that the needs of different users with more or less the same profile can be met by the same set of lexicographical data. It just means that the data in terms of language, formality, content and design should be prepared and presented in such a way that users can solve their particular problems as smoothly as possible.

In the following article, we will show how contextualization and personalization entered into lexicography more than two thousand years ago and developed over time. We will also discuss how these two interrelated phenomena were separated, especially after the introduction of the printing technology, and almost completely ignored in the academic literature until their recent "resurrection" as a result of the new digital technologies.

1. Where it all began — at least in Europe

Four decades ago, the Egyptian scholar Al-Kasimi (1977: 1) observed that the "major motives behind the rise of lexicography differ from one culture to another" and that each culture therefore develops "dictionaries appropriate to its characteristic demands". In this perspective, and based on a study by Stathi (2006), Hanks (2013: 507) traces the origin of European lexicography back to the Classical Greek Period where "it was customary for Greek scribes to insert glosses into manuscript copies of the works of Homer and other earlier writers" in order to explain "obsolete and unusual words". In contrast to the difficulty experienced by their predecessors in the Middle East who worked on clay tablets, the introduction of pen and parchment made it easy for the Greek scribes to perform this task with a certain elegance. Hanks explains how the glosses, from the third century BC onwards, were "compiled into separate glossaries by scholars at the library in Alexandria" thus giving birth to the dictionary form, as we have known it since then. Historians of lexicography see "the origin of the 'dictionary' proper in this cavalier yet practical way of adding snippets to copied texts" (McArthur 1986: 76).

As it is customary in the mainstream British lexicographical tradition, Hanks (2013) then focusses on some highly relevant linguistic phenomena which he convincingly develops throughout his article. However, if one changes the focus and takes an overall historical approach, another extremely interesting phenomenon will materialize.

Readers of old texts obviously had information needs when they ran into unfamiliar words. These information needs were neither general nor abstract but very much concrete and directly related to both a specific activity (reading)

and a specific place (page, line, position) in the text. When the scribes inserted glosses into manuscript copies of the classical works, they did it in the specific context where the information need first occurred. These glosses, or proto-lexicographical data, were therefore by definition, and from the very beginning, *contextualized data*.

Manuscript books were valuable and only few people could afford to have their own personal copies. This suggests that the majority of literate people who, for their part, only made up a tiny minority of the total population, would have to go to public places (the birth of libraries) in order to read, study and enjoy the works of Homer and other writers. The scribes, who worked at such places and were highly respected, may therefore have known at least some of these readers personally and observed their problems when encountering words that were "obsolete and unusual", maybe even discussing it with them. In this respect, the glosses inserted into the manuscript copies also represented a *personalized service* based on personal acquaintance or even relationship between scribes (proto-lexicographers) and readers (proto-lexicographical users).

Thus, it transpires that European dictionary making was born out of a tradition where the Greek scribes provided personalized and contextualized data to readers whom they knew had (or whom they expected to have) comprehension problems, and therefore also information needs, when reading the works from a previous period.

This interesting, user-friendly and admirable phenomenon began to partially disintegrate only two centuries later when the scholars at the library in Alexandria started compiling separate glossaries. This new invention should not be underestimated as it gave birth to a completely new and successful cultural practice, the making of dictionaries, which has survived for more than two thousand years and satisfied the needs of hundreds of millions of people. However, although being a very practical and useful invention, the glossaries also had a less desirable secondary effect. It meant that the satisfaction of information needs occurring in a specific context was decontextualized and externalized to a different information source. This may be a small problem when the glossary is adapted to a specific book and structured according to the consecutive appearances of the possible problems giving rise to information needs. In most dictionaries, this is not the case. Their macrostructures are organized alphabetically, systematically, or according to other principles that break any direct connection to a specific problem in a specific text. The introduction of separate reference works therefore created a new distance between the occurrence of an information need and its lexicographical solution. This distance grew bigger over time. It also extended the consultation time, complicated the consultation process and increased the risk of not finding an appropriate solution. In this way, the use of contextualized data was abandoned by the branch of lexicography dedicated to dictionary making.

2. Looking back with new eyes

After the invention of the glossary, a schism developed within lexicography. On the one hand, there was the new compilation of glossaries and dictionaries of a still more complex and sophisticated nature. On the other hand, there was the traditional insertion of glosses into manuscript copies of books from previous periods. This situation lasted for almost two thousand years. It continued through the Roman Imperial Period and the three European Middle Ages until the fortunate advent of the printing press in the 15th century. The scribes' beautiful profession died out soon after the introduction of this disruptive technology. And so did the last remains of lexicographical contextualization, at least for the time being.

However, although contextualization was completely abandoned by the compilers of dictionaries, personalization continued to express itself in different ways. In the beginning, the technological and societal conditions for the printed dictionary were still limited. The editions were small and most lexicographers had a fairly good knowledge of their relatively small group of customers and their needs.

Captains, who themselves had experienced communication problems when arriving in foreign seaports, compiled bilingual and multilingual dictionaries to be used by members of their crews as well as fellow captains and seamen. Priests, who had observed learners' problems when studying the Bible, authored special dictionaries explaining difficult words and expressions. Many missionaries in the colonial countries were the first authors of bilingual dictionaries between their own mother tongue and the indigenous languages. Malachy Postlethwayt, the author of *The Universal Dictionary of Trade and Commerce* (1774), had a very clear idea of the needs of his future users, many of which he knew personally. For a detailed discussion of this dictionary, see Tarp and Bothma (2013). The same applies to Samuel Johnson and other well-known and unknown lexicographers from that period.

Even today, you can find many authors who started their lexicographical career on their own initiative after observing unfulfilled user needs, typically within specialized subject fields and small user segments. Most writers of the Scandinavian emigrant's dictionaries, a variant of learner's dictionaries discussed by Pálfi and Tarp (2009), are either emigrants themselves or language teachers who have a very profound knowledge of the problems related to the learning of a new language. The authors are not trained lexicographers but dedicated people who want to assist the emigrants. Their works are seldom a commercial success as they are mostly printed in small editions by small publishing houses dedicated to language didactics.

The situation in terms of the big general dictionaries is different. From the 19th century onwards, a phenomenon that could be described as *lexicographical alienation* began to spread. The editions became bigger and bigger as a result of the growing societal demands and the continuous development of the printing technology. The dictionaries reached out to many more users with different

profiles. The negative consequence of this massification was that users were decontextualized from the specific situation (reading, writing, learning) where their information needs occurred. Instead, they are treated as abstract individuals with abstract needs without paying much attention to their specific needs in different situations. To a large extent, the production of dictionaries became business as usual, although there were beautiful exceptions. In some cases, dictionary projects do not even take their point of departure in detected user needs, but in the aspirations of certain families or companies who want to show that they too are knowledgeable and can produce such highly regarded works.

The result of all this was a still larger distance between the lexicographers and their users who, in the late 20th century, ended up being characterized as the "well-known unknown" (Wiegand 1977: 59). The upsurge and rapid spread of lexicographical user research during the past three to four decades is as an attempt to remedy this alienation. Publishing houses resort to market research by focus groups in order to get a better grasp on their customers' needs and expectations and thereby raise their sales. Lexicographers employ questionnaires and other methods to get a better picture of the well-known unknown, mostly obtaining general knowledge of usage that can only partially be applied in concrete dictionary projects.

However, it is a question whether this kind of remedial action can replace the direct personal contact with users. Tarp and Ruiz Miyares (2013) report on a special Cuban experience where authors of school dictionaries organize caravans and socialize with teachers and learners. In this way, both users and lexicographers get a better knowledge of each other for the benefit of the quality, relevance and proper usage of the dictionaries.

Gouws (2016) discusses how the authors of the *Ju'hoan Tsumkwe* dialect dictionary (Jones et al. 2014) were in personal contact with their target users from a small speech community with an endangered language in order to determine exactly what they needed. Each article in this dictionary contains an illustration made by members of the speech community. These illustrations, respond to real information needs. They help the children to learn the language and grasp the meaning of words unknown to them. The motto was "Hold your people, your language and your culture tightly together".

Similar experiences are rare and highlights the need for a discussion and redefinition of the social role of lexicographers. This redefinition must depart from the need to offer a more personal lexicographical service to users and take into account the current technological possibilities of providing new and better solutions to old problems.

3. The (almost) forgotten reality

Contextualization was an integral part of proto-lexicographic work. In spite of the reduced levels of contextualization and personalization due to the intro-

duction and development of traditional dictionaries some authors of modern-day texts still adhere to the contextualization tradition.¹

According to Evans (1979: 104):

The invention of writing was the most revolutionary of all human inventions, for in one great blow it severed the chains which tied an individual and his limited culture to a finite region of space, to a restricted slice of time ...

Manual writing and the later use of the printing press and digital devices created the opportunities to present more and more data and also to integrate different types of data into a single text in order to enhance the successful retrieval of information.

The increased production of texts compelled the authors to ensure that their users understand the texts and this often demanded that a brief explanation of difficult or lesser known words had to be included. As was the case with glosses provided as annotations in manuscripts this contextualization is in an extra-dictionary environment. Today, in extra-dictionary environments, one finds an increase in assistance given to the target readers of different texts in order to meet their specific information needs. This can for example be found in text books where new terms are introduced and these terms are immediately explained so that the user can acquire the necessary text reception help. In the text book *Kontemporêre Afrikaanse Taalkunde* (Carstens and Bosman 2017), directed at students of Afrikaans linguistics, terms are often introduced and explained to the target users. In their chapter on Pragmatics in this book Van Niekerk and Olivier (2017: 338) make a reference to the Afrikaans terms *lokusie*, *illokusie* and *perlokusie* and immediately explain these terms as follows (the translation has been added):

Die lokusie is die uitspraak van 'n sin met 'n bepaalde verwysing en referensie.
(= Locution is the pronunciation of a sentence with a specific reference)

Die illokusie is die uitspraak van 'n sin met 'n bepaalde bedoeling (bevel, versoek, ens.) (= Illocution is the pronunciation of a sentence with a specific purpose (command, request, etc.)

Die perlokusie is die uitspraak van 'n sin om 'n bepaalde reaksie by die hoorder teweeg te bring ... (= Perlocution is the pronunciation of a sentence to evoke a specific reaction from the listener)

The explanations represent contextualized data which help the reader to understand a term in a specific context. It constitutes a form of specialized lexicography that continues the scribes' tradition of inserting glosses, and where lexicographic procedures are introduced into an extra-dictionary environment.

Contextualized lexicographical data in texts often take the form of footnotes and endnotes. Burm and Van der Merwe (1973) present the well-known Middle Dutch epic poem *Van den Vos Reynarde*. Their publication is directed at Afrikaans students of Middle Dutch who are not familiar with all the Middle

Dutch words and expressions. The verse lines of the poem are numbered and on each page there are footnotes starting with the number of the line and a problematic word or expression in that line. This is followed by a brief treatment — either an Afrikaans translation equivalent, a translation of a line or a part of the line, a brief grammatical note and often a reference to a specific paragraph in a Middle Dutch textbook.

In a South African edition of the Dutch novel *Het Fregatschip Johanna Maria* (Van Schendel 1971) certain words in the text have an endnote number. In the back matter section of this novel the endnotes are arranged by chapter and the problematic word in the text is treated by means of a brief explanation of meaning or an Afrikaans translation equivalent. This edition of Van Schendel's novel was specifically prepared for Afrikaans students studying Dutch literature. The endnotes were written by a professor of Afrikaans and Dutch literature and the contextualization is directed at the specific needs of the student target group.

Another text genre where contextualizing procedures can often be found, is restaurant menus. An entry on the menu is often given in a foreign language, the language of origin of the specific dish, followed by a translation equivalent in the default language of the menu or a brief paraphrase of meaning in this default language. A menu of a South African restaurant includes the following entries:

Bobotie

Baked curried mince with ...

The traditional Afrikaans dish (bobotie) is the main entry in this slot on the menu. Before mentioning the accompanying side dishes a brief paraphrase of meaning ("baked curried mince") is given to help the foreign visitor to the restaurant to understand the meaning of the word *bobotie*. This is yet again an attempt by the compiler of the menu, by means of a procedure of contextualization, to assist the user in retrieving the appropriate information from the entries on the menu.

The restricted environment of a specific text, whether a textbook, novel or menu, and familiarity with the target users of that text makes it easier for authors and compilers of the text to guide their specific readers with respect to specific questions that may arise when the text is read. Contextualized entries like footnotes, endnotes and translation equivalents directed at the explanation of a given word in the text will assist the reader who is not familiar with this word, its meaning or its translation equivalent. This was done in the proto-lexicographic works where words in manuscripts were annotated and these endeavours are still used today.

One potential problem with some extra-dictionary contextualization procedures is the fact that the users do not know whether they are accessing curated data or not. In text books users will regard the paraphrases of meaning or the translation equivalents as equal in quality as the data presented in spe-

cialized dictionaries. Glosses, annotations and lexicographic contributions in general texts, including menus, are not necessarily seen as having the same authority as comparable entries in dictionaries. This is due to the traditional approach that dictionaries are authoritative sources of information. Unfortunately, practical experience has also shown that lexicographic data in extra-dictionary sources often are incorrect translation equivalents or inappropriate paraphrases of meaning.

Contextualization does not only occur in an extra-dictionary environment. Dictionaries are also texts and the language used in a dictionary may include words or expressions that need to be explained for the target user of the dictionary.

In the *Greek-English Lexicon of the New Testament based on Semantic Domains* (Louw and Nida 1988) footnotes are used to give additional information regarding some aspect dealt with in a dictionary article. Figure 1 contains footnotes 14 and 15 for articles 37.97 and 37.98 respectively that present additional contextualized data.

37.97 ἀφορίζω^b: to set aside a person for a particular task or function – ‘to appoint, to set apart for.’¹⁴ ἀφορίσατε δὴ μοι τὸν Βαρναβᾶν καὶ Σαῦλον ‘set apart for me Barnabas and Saul’ Ac 13.2.

37.98 δίδωμι^c: to assign a person to a task as a particular benefit to others – ‘to appoint, to assign (on behalf of).’¹⁵ μετὰ ταῦτα ἔδωκεν κριτάς ‘after this he appointed judges (for them)’ Ac 13.20. It may be possible in some languages to render δίδωμι in Ac 13.20 as ‘he gave them judges,’ but more frequently it is necessary to use a phrase such as ‘he appointed judges to rule over them.’

37.99 παραλαμβάνω^c: to receive an appointment for a particular ministry – ‘to receive an appointment, to receive a task, to be assigned a ministry.’ τὴν διακονίαν ἣν παρέλαβες ἐν κυρίῳ ‘the ministry which you received in the service of the Lord’ Col 4.17.

37.100 μερίζω^d: to assign a particular part or aspect of a function or responsibility – ‘to assign a particular responsibility, to give a particular task to, to appoint a particular part

14 It is possible that ἀφορίζω^b is somewhat more emphatic in meaning than ἀρίζω^b (37.96).
15 δίδωμι^c seems to imply a benevolent purpose in the action of assigning someone to a role.

Figure 1: Footnotes in Louw and Nida (1988)

By adding these footnotes the lexicographers did something similar to what the early scribes have done in their proto-lexicographic work — introduce procedures of contextualization to benefit the target reader.

4. The few visionaries

In the previous sections reference was made to extra-dictionary contextualization. A question is whether this kind of treatment of words and expressions should be regarded as extra-lexicographical contextualization? The perspective adhered to in this paper is that lexicographic work does not necessarily only have to be performed within dictionaries. Where a word occurring in a text is complemented by either a paraphrase of meaning or a translation equivalent this treatment resembles a part of the treatment typically found in dictionaries, i.e. a lexicographic treatment.

The past decades have witnessed a number of visionary ideas and proposals that can form a basis for a renewed attempt at yet again giving contextualization and personalization their rightful place in lexicographic endeavours.

In the proto-lexicographic environment the scribes knew exactly what the problems of the readers of manuscripts would be. They could predict the words that were in need of explanation and they could provide the necessary contextualized solutions to these problems. Today lexicographers are not in that position where they know exactly what kind of explanation to provide. Hanks (1979: 35) said:

The lexicographer is in the impossible position of a man who undertakes to answer peoples' questions, but since he does not know at the time of compilation what questions exactly his public will ask, he has to word his entries so as to answer all possible questions about them.

These answers are not necessarily directed at the specific needs of the target users and do not enhance either contextualization or personalization. Another linguist had insight in this problem and its detrimental effect on dictionaries. Bolinger (1985: 69) said:

Lexicography is an unnatural occupation. It consists in tearing words from their mother context and setting them in rows ... to make them fit side by side, in an order determined not by nature but by some obscure Phoenician sailors ... Half of the lexicographer's labor is spent repairing this damage to an infinitude of natural connections that every word in any language contracts with every other word ...

Bolinger realized the problem but was unable to provide a solution. Some visionary lexicographers did come to the fore with suggestions to counter the problems of decontextualization and to move beyond the unnatural environment of a dictionary towards the provision of personalized data directly in the context where the user experiences an information need.

Bergenholtz (1998: 93) refers to an information tool found in Disney cartoons, i.e. the *Junior Woodchucks' Guidebook*, introduced in the 1950s, that provides answers to almost all the questions of Donald Duck's nephews Huey, Dewey and Louie. This expresses the need for a personalized information tool that could provide one or more users with all kinds of information they might need and that makes provision for quick and easy access to the data.

The ideas expressed by Hanks, Bolinger and Bergenholtz form a sound basis for further innovative work to ensure higher levels of contextualization and personalization in lexicographic environments — whether in dictionaries or in extra-dictionary sources.

The era of visionary dreaming about these new approaches culminated in De Schryver (2003). He presents a comprehensive list of dreams discussed by a number of lexicographers. These dreams focus on the possibilities the digital environment could offer lexicography. They include hopes for different levels of interface for different users, a personal dictionary and the possibility to choose the content and presentation languages. Such dreams form the basis of enhanced future contextualization and personalization possibilities. It is succinctly expressed by Varantola (2002: 31):

I will be shamelessly selfish and ask for the impossible. I will advocate for a dictionary that will always adapt to my needs and always be ready to provide me with exactly the answer that I need and will also agree with. I also expect the dictionary to be able to give satisfactory answers to those questions that I forget to ask.

It is important to note that Varantola regarded her dream as an impossibility. This also applies to many of the other dreams and suggestions that never came to fruition. The lack of the appropriate technical means played a role in impeding the realization of such proposals. Many of the visions, dreams and proposals were expressed without being sure that the day will come that they could be realized. That time has arrived and lexicographers need to formulate plans that could lead to a new lexicographic dispensation.

This paper shows some steps towards this realization.

5. The first lexicographical steps towards a contextualized renaissance

De Schryver (2003) listed more than hundred dreams expressed by a large number of lexicographers in the previous decades. These dreams were, by definition, premature as the corresponding technology was not yet developed. A few of the dreams have subsequently been realised in different tools, at least partially. The rest are still waiting for Godot and many of them will probably share the fate of Vladimir and Estragon.

Since then, lexicographers have not stopped having dreams, or visions, as we prefer to call them.

Gouws (2006), for instance, proposed the idea of a *Mutterwörterbuch*, a

mother dictionary, in the field of bilingual dictionaries with Afrikaans and German as language pair. The idea was that from such a comprehensive bilingual dictionary different users and different user groups would be able to retrieve personalized information directed at their specific punctual needs. Scepticism on the side of a publishing house unfortunately worked against these ideas — also because it would have challenged the then existing technology.

Rundell (2007: 49) emphasizes customization and personalization as "likely new directions" that can unpick "the current globally-marketed one-size-fits-all package", i.e. the standardized dictionary. In the Internet, Google and Wikipedia era, a major challenge in terms of learners' dictionaries is to provide information "which is either not available elsewhere, or not available in an easy-to-use form that takes account of learners' needs (and limitations)" (p. 50). How should this be done? According to the British scholar, the way forward is to make the lexicographical resources less static and more dynamic:

A possible scenario is to see our reference materials as a set of components which customers can mix and match according to their needs. (Rundell 2007: 59)

Tarp (2011) continues in the same vein and proposes a future lexicographical Rolls Royce as a possible way forward to "the individualization of needs satisfaction". This vision is explained in more detail by Tarp (2012) who, among other things, highlights "user profiling, situation description and filtering" as techniques to obtain a more personalized product. He also suggests that "each individual user of a lexicographical e-tool will be given the option to design his or her own master article in terms of the types of data wanted and their arrangement on the screen." (Tarp 2012: 261). The technology required to offer this fancy solution is already there, but publishers do not seem to be very enthusiastic about the idea. The reason is probably that they do not expect users — especially those who only occasionally consult a particular e-tool — to be willing to spend their time designing master articles. In this respect, Tarp (2012) represents the necessary exaggerations born out of an over-optimistic view.

Gouws' (2006) "mother dictionary", Rundell's (2007) "mix and match" and Tarp's (2012) "master article" certainly pointed to a further personalization of the lexicographical product. Nonetheless, the proposals were, to a large extent, of little use as they required a number of steps which most users would probably not take. None of the three authors grasped the main picture. If they had looked back at the Classical Greek Period with new eyes, they would have known that personalization, in order to be successful, has to be combined with a new type of contextualization of the lexicographical product. The way forward is not to ask users to go looking for information in new and complex ways, to mix and match components or to design fancy master articles. The real challenge is rather to develop tools that can provide personalized and customized service to users directly in the context where they experience information needs, and without the users having to take any, or only a few, steps to solve their problems.

Just as it was the case with many of the visionary lexicographers mentioned by De Schryver (2003), this limitation in their visions was most likely due to the fact that the relevant technologies allowing us to think out of the box were not yet fully developed, or developed at all. Today, the technology required to take a big step forward is here, at least to a certain extent.

6. The real solution — at least for now

Currently, there are various examples of lexicographical products being integrated into information tools that little by little are closing the more than two thousand year old schism in European lexicography, i.e. reuniting contextualization and personalization. These advanced information tools are above all designed to assist the reading, writing and translation of texts as well as learning in general. In this chapter, we will briefly discuss tools designed to assist e-reading and L2 writing.

Kindle and other e-readers allow users who have comprehension problems to access integrated dictionaries that may provide assistance with a simple click or touch on the word in question. (For examples and a detailed discussion, see Bothma and Prinsloo 2013). The same holds true for various smart phones and tablets.



Figure 2: Screenshots from an iPhone

In Figure 2, a person reading an article from a South African newspaper on an iPhone does not understand the word "glimpses" and therefore touches the word with a finger. A little box is immediately displayed on the screen, on which a touch on "Look Up" gives access to a number of preselected dictionaries (Figure 3).



Figure 3: Articles from different dictionaries on iPhone

The articles in Figure 3 are taken from four different dictionaries, three monolingual English ones and one biscopal English–Spanish, which have been preselected by the owner of the iPhone. If the user does not feel comfortable with these dictionaries and want to re-saddle, a simple touch on the button "Manage Dictionaries" unfolds a page where the user can delete some of the dictionaries and add others that are more adapted to his or her profile. In this way, the user can personalize the lexicographical service requested from the device.

If the user wants to continue the search process, he or she may access one of the dictionary articles indicated in Figure 3, for instance the one from the *Oxford Dictionary of English*. To the big surprise for users of high-tech tools, a ghost from the past then appears in the form of a traditional dictionary article. The two articles displayed in Figure 4, which in this case contain few senses, also include data types (e.g. pronunciation and etymology) which are completely irrelevant for a user who just wants to understand a specific word. This results in lexicographical data overload as defined by Gouws and Tarp (2017). The user therefore frequently has to scroll down in order to find first the word class and then the sense that are relevant in each consultation, as neither word classes nor senses are prioritized according to the specific context. This prolongs and complicates the search process and goes against the idea of contextualization.

Hence, on the one hand, the user gets access to the lexicographical service directly in the context where the information need occurs, i.e. contextualization. On the other hand, the long road to the required data means that full contextualization is still a challenge to modern lexicography and producers of high-tech tools. Currently, companies like IBM are conducting comprehensive interdisciplinary research in order to develop a tool that can deduce the specific meaning of a word from the context. Unfortunately, and as far as we are informed, they have not yet come up with any convincing results.



Figure 4: Dictionary article from Oxford Dictionary of English on iPhone

More or less the same applies to the writing assistant produced by the Danish company Ordbogen A/S; c.f. Tarp et al. (2017). This tool, branded as Write Assistant, is designed to provide instantaneous assistance to L2 users who may have language problems when writing in English. So far it is available for users with Danish, Spanish, German, French, Italian and Arabic as a mother tongue. In contrast to the monolingual writing assistants which are available on tablets and smart phones, the Danish Write Assistant does not only provide word terminations and predict the next word in the sentence. It also offers L2 equivalents when the users write an L1 word, as well as the possibility to access a dictionary article directly from the suggested words or equivalents (Figure 5).

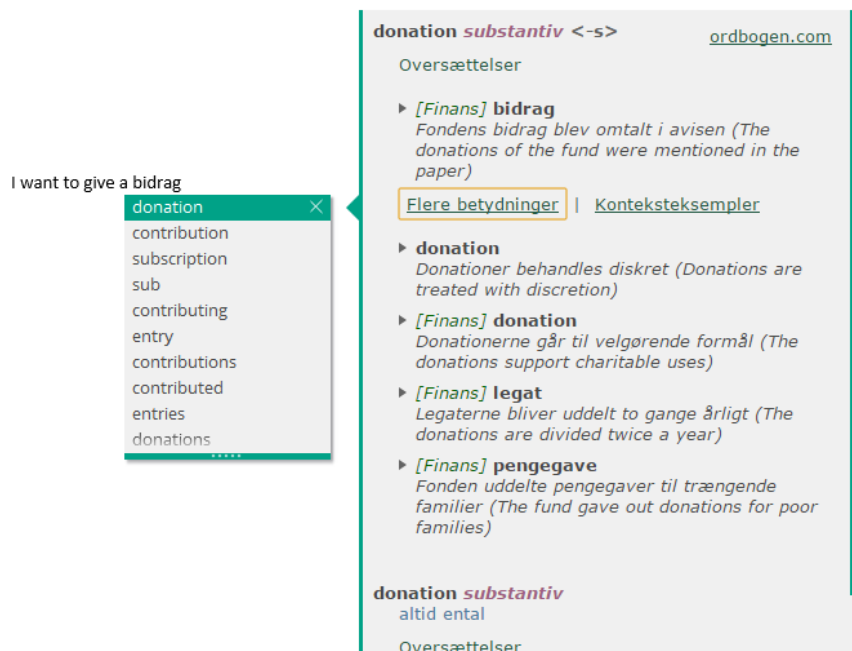


Figure 5: Write Assistant with prioritized equivalents and dictionary article

Instead of expecting an active role from users, which in most cases would be highly unlikely, Write Assistant "observes" its users and, on this basis, chooses the set of lexicographical data most likely to meet their needs in each concrete situation. These data are placed in the immediate vicinity of the word that may pose challenges to the writer. This is shown in the small window in Figure 5, where the English equivalents to the Danish word *bidrag* are placed directly below this word. In addition, the order of the equivalents is prioritized by the tool based on an automatic analysis of the context and the calculation of the

most likely solution in this particular context.

When the word or equivalent suggested by Write Assistant is sufficient to meet the user's needs — for instance, if it is only used as a reminder — the tool has achieved the complete reunification of personalization and contextualization.

However, if the user is not sure which of the suggested words to use, or how to use it, and therefore proceeds to open the big window by touching or clicking on one of the suggested words, a problem similar to the one discussed above appears. As can be seen in Figure 5, the dictionary article, which is a beta version, is still very traditional and therefore not fully adapted to the user's needs in each consultation. With the introduction of artificial intelligence in the autumn of 2019 as well as a much closer interdisciplinary collaboration between lexicographers, designers and information engineers, the provider of the tool expects to eventually uncover the scribes' 2 500 years old secret in the nearby future.

In this respect, it is worth stressing that lexicography is no longer the responsibility of a lone ranger on a white horse. Team work has become extremely important. Collaboration with other fields like information science is a prerequisite for successful lexicographic endeavours in the digital era.

7. Perspectives

The past few years have seen a continuous development of the new disruptive technologies with the introduction of neural networks, artificial intelligence, among others. These technologies are now making their ways into lexicography and the production of information tools in general. As a result, lexicographers are exposed to new horizons.

With contextualization and personalization being reunited we are on the brink of seeing the end of the schism that existed for so long in lexicography. The accumulation of more than two thousand years of dictionary-making experience puts the modern scribes in a much better position to produce contextualized and personalized lexicographical data. These efforts are supported by innovative ways of providing an information merge through the mediation of new technologies. This ascertains yet again that lexicography as a discipline is now closely related to information science.

Modern-day lexicographers are in a position to make some of the unfulfilled dreams of the past a reality. The challenge of the future is to make the impossible possible.

We have work to do.

8. Endnote

1. As indicated earlier in this article we use the term *contextualization* in a way that deviates from the way in which it is often used in lexicography and linguistics. In this paper it refers

to the provision of solutions to users' information needs directly in the situation or context where these needs occur.

9. Literature

- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Bergenholtz, H.** 1998. Das schlaue Buch. Vermittlung von Informationen für textbezogene und text-unabhängige Fragestellungen. Zettersten, A., J.E. Mogensen and V.H. Pedersen (Eds.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexicography May 2–4, 1996 at the University of Copenhagen*: 93-110. Tübingen: Max Niemeyer.
- Bolinger, D.** 1985. Defining the Indefinable. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 69-73. Oxford: Pergamon Press.
- Bothma, T.J.D. and D.J. Prinsloo.** 2013. Automated Dictionary Consultation for Text Reception: A Critical Evaluation of Lexicographic Guidance in Linked Kindle e-Dictionaries. *Lexicographica* 29(1): 165-198.
- Burm, O.J.E. and H.J.J.M. van der Merwe.** 1973. *Van den Vos Reynarde*. Pretoria: J.L. van Schaik.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Evans, C.** 1979. *The Micro Millennium*. New York: Viking.
- Gouws, R.H.** 2006. Die zweisprachige Lexikographie Afrikaans-Deutsch — Eine metalexikographische Herausforderung. Dimova, A., V. Jesenšek and P. Petkov (Eds.). 2006. *Zweisprachige Lexikographie und Deutsch als Fremdsprache*: 49-58. Hildesheim: Georg Olms.
- Gouws, R.H.** 2016. Enkele minder bekende Afrikaanse woordeboekmonumente. *Tydskrif vir Geesteswetenskappe* 56(2-1): 355-370.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Hanks, P.** 1979. To What Extent Does a Dictionary Definition Define? *ITL* 45-46: 32-38.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Allan, K. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*, 503-536. Oxford: Oxford University Press.
- Jones, K.L. et al. (Eds.).** 2014. *Ju/hoan Tsumkwe Dialect/Prentewoordeboek vir kinders/Children's Picture Dictionary*. Pietermaritzburg: University of KwaZulu-Natal Press.
- Louw, J.P. and E. Nida.** 1988. *Greek-English Lexicon of the New Testament Based on Semantic Domains*. New York: United Bible Societies.
- Pálfi, L.-L. and S. Tarp.** 2009. Lernerlexikographie in Skandinavien — Entwicklung, Kritik und Vorschläge. *Lexicographica* 25: 135-154.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer*. Cambridge: Cambridge University Press.
- Postlethwayt, M.** 1774. *The Universal Dictionary of Trade and Commerce: With Large Additions and Improvements, Adapting the Same to the Present State of British Affairs in America, Since the Last Treaty of Peace Made in the Year 1763. With Great Variety of New Remarks and Illustrations Incorporated throughout the Whole: Together with Everything Essential that is Contained in Savary's Dictionary: Also, All the Material Laws of Trade and Navigation Relating to these Kingdoms, and the Customs and Usages to which All Traders are Subject*. The Fourth Edition. London: W. Strahan, J. and F. Rivington, J. Hinton.

- Rundell, M.** 2007. The Dictionary of the Future. Granger, S. (Ed.). 2007. *Optimizing the Role of Language in Technology-enhanced Learning*. Proceedings of the expert workshop organized in Louvain-la-Neuve (Belgium), 4–5 October 2007: 49–51. <https://hal.archives-ouvertes.fr/hal-00197203/document/>. (Consulted 4 July 2019).
- Stathi, E.** 2006. Greek Lexicography, Classical. Brown, K. (Ed.). 2006. *Encyclopedia of Language and Linguistics*. Vol. 5: 145-146. Second Edition. Oxford: Elsevier.
- 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.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28(1): 253-267.
- Tarp, S. and T.J.D. Bothma.** 2013. An Alternative Approach to Enlightenment Age Lexicography: The Universal Dictionary of Trade and Commerce. *Lexicographica* 29(1): 222-284.
- Tarp, S. and L. Ruiz Miyares.** 2013. Cuban School Dictionaries for First-Language Learners: A Shared Experience. *Lexikos* 23: 414-425.
- Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
- Van Niekerk, A. and J. Olivier.** 2017. Pragmatiek. Carstens, W.A.M. and N. Bosman (Eds.). 2017. *Kontemporêre Afrikaanse Taalkunde*: 329-363. Pretoria: Van Schaik.
- Van Schendel, A.** 1971. *Het Fregatschip Johanna Maria*. (South African edition). Pretoria: Academica.
- Varantola, K.** 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility? Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins*: 30-44. Göteborg: EURALEX.
- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, G., H. Henne and H.E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Vienna/Zürich: Bibliographisches Institut.

Focus on the User: Front Matter in Slovenian Dictionaries

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: This article focuses on the front matter texts of three Slovenian monolingual dictionaries: *Dictionary of Standard Slovenian*, *Dictionary of Slovenian Synonyms* and *Dictionary of Legal Terminology*. We asked thirty-two MA students of English at the University of Ljubljana to read the front matter texts and comment on intelligibility of the text, usefulness of the text from the point of view of intended users, language used in the text, examples provided to illustrate the theoretical explanation and length of the text as a whole. According to the majority of the students, the front matter texts of all three dictionaries use terminology known only to linguists. They also frequently comment on the length and complexity of sentences, information overload and the indirect inclusion of examples that illustrate the theory. Apart from being critical of the texts, they point out what they like (the front matter texts in all three dictionaries are well-structured, as they are divided into sections and subsections, thus enabling easy access to relevant information within the text) and give proposals for improvement of the text (rearrangement of various sections, the use of Slovenian terms instead of borrowings, the use of paraphrases instead of terms, shorter sentences and simpler syntax).

Keywords: FRONT MATTER TEXTS, USER GUIDE, GRAPHIC PRESENTATION OF ENTRIES, INTELLIGIBILITY, USEFULNESS, EXAMPLES, LINGUISTIC TERMS, SYNTAX

Opsomming: Fokus op die gebruiker: Voortekste in Sloweense woordeboeke. Hierdie artikel fokus op die voortekste van drie Sloweense eentalige woordeboeke: *Woordeboek van Standaardsloweens*, *Woordeboek van Sloweense Sinonieme* en *Woordeboek van Regsterminologie*. Twee en dertig MA-studente van Engels aan die Universiteit van Ljubljana is gevra om die voortekste te lees en kommentaar te lewer op die verstaanbaarheid van die teks, die bruikbaarheid van die teks vanuit die perspektief van die teikengebruikers, die taal wat in die teks gebruik word, die voorbeelde wat verskaf word om die teoretiese toeligting te illustreer en die lengte van die teks as geheel. Volgens die meerderheid studente word daar in die voortekste van al drie woordeboeke terminologie gebruik wat net aan taalkundiges bekend is. Hulle lewer ook dikwels kommentaar op die lengte en kompleksiteit van sinne, inligtingoorlading en die indirekte insluiting van voorbeelde wat die teorie illustreer. Buiten die kritiek wat hulle op die tekste lewer, dui hulle ook aan waarvan hulle hou (die voortekste in al drie woordeboeke is goed gestruktureer aangesien hulle verdeel is in afdelings en onderafdelings, wat sodoende maklike toegang tot relevante inligting binne die teks

moontlik maak) en maak hulle voorstelle vir die verbetering van die tekste (herorganisering van verskeie onderafdelings, die gebruik van Sloweense terme in plaas van ontlennings, die gebruik van parafrases in plaas van terme, korter sinne en eenvoudiger sintaksis).

Slutelwoorde: VOORTEKSTE, GEBRUIKERSGIDS, GRAFIESE VOORSTELLING VAN INSKRYWINGS, VERSTAANBAARHEID, BRUIKBAARHEID, VOORBEELDE, LINGUISTIESE TERME, SINTAKSIS

1. Front matter: Some theoretical considerations

It is a widely held perception that dictionaries are repositories of words, but they may also contain pictures, symbols, audio files, etc. When users take a dictionary off a shelf or access it on their computers, they do it mostly to look up a word or word combination. A dictionary is the product of any lexicographic work, and the word 'dictionary' itself is powerful, since it implies authority, scholarship and precision (Landau 2001: 6). But what actually is a dictionary? It is not only a reference resource consisting of "an alphabetical list of words with their meanings and parts of speech, and often a guide to accepted pronunciation and syllabification, irregular inflections of words, derived words of different parts of speech, and etymologies" (sense 1.a) or "a similar reference work giving equivalent words in two or more languages" (sense 1.b), as defined in *Collins English Dictionary*, since these definitions focus only on the A-Z section of the dictionary (if the language treated in the dictionary uses an A-Z alphabet), which is by no means the only element of the dictionary structure (cf. also Bergenholtz's definitions of 'dictionary' in Bergenholtz 2012: 30). Admittedly, from the perspective of an average user, this is the most important part, the (only) part they actually consult. However, from the metalexigraphic perspective, the dictionary structure is much more complex, as it refers to the interrelation of dictionary components. The information provided by a dictionary is far from being simple in nature, and if lexicographers want the information included in a dictionary to be comprehensible to users, it should be presented according to certain structural principles. Thus, the dictionary structure includes megastructure, microstructure, access structure, distribution structure and cross-reference structure (Hartmann 2001: 57-62; Svensén 2009: 77-82). The megastructure is the overall framework of the component parts of a dictionary and consists of the macrostructure and the outside matter. The elements of the outside matter are classified depending on their physical position in the dictionary, considering the A-Z section as the central part of the dictionary: the front matter precedes the A-Z section; the middle matter contains items inserted in the A-Z section without forming a constituent part of it; and the back matter follows the A-Z section. The front matter may include the title page, the copyright page and imprint, acknowledgements, the foreword or preface, the table of contents, the list of contributors, the list of abbreviations, labels and codes used in the text and/or illustrations, the pronunciation key,

the user's guide and dictionary grammar. It may also offer essays on certain aspects of language, such as the history of the language, a feature which largely depends on the market at which it is aimed (Hartmann and James 1998: 60; Atkins and Rundell 2008: 176-177). Since this article concentrates on front matter texts only, other elements of the megastructure will not be the focus of attention.

Taking account of the variety of text types or genres a dictionary includes, the most obvious difference is between the text in the outside matter and that of the A-Z section (Bergenholtz, Tarp and Wiegand 1999: 1763). As mentioned above, the division of the outside matter is based on the position of texts in a printed dictionary. Similar to printed dictionaries, online dictionaries also contain a number of outer texts. As Müller-Spitzer (2013: 374) maintains, the term outer features is more appropriate when referring to outside matter in electronic dictionaries, since the elements in the outer domain of online dictionaries do not all belong to the category of text (Klosa and Gouws 2015: 148). In online dictionaries, the outer features, and the user guide in particular, are as important as in printed dictionaries or even more so, since online dictionaries offer a much greater variety of search options than printed dictionaries, and these options have to be explicated if users are to make full use of them (Nielsen 2006: 11). Users also have different expectations when consulting electronic dictionaries as there is a whole range of standard web or mobile practices that are required or that affect the user experience.

In the study whose findings are presented in this article, the focus is on the utility of front matter texts to dictionary users. Front matter texts include one very specific part, i.e., the user guide, which is second only to the A-Z section in terms of importance of the texts in a dictionary (Kirkpatrick 1989; Landau 2001: 149; Nielsen 2006: 1) and is therefore absolutely essential if lexicographers want the dictionary to be an effective tool (Nielsen 2005: 137). As Kirkpatrick (1989: 754) writes in her insightful article on user guides, "[t]he general assumption is that no-one bothers to read the front matter of dictionaries", an assumption, confirmed by the students participating in our small-scale study. She also points out that dictionary users are reluctant to consult the front matter (even less the back matter). This lack of interest in the front matter is at least partly related to the popular belief that all dictionaries are similar in structure, and no specific information on their use is necessary (Kirkpatrick 1989: 754; Nielsen 2006: 7). The types of information that lexicographers should include in a user guide depend on the dictionary compiled and vary from dictionary to dictionary. Nielsen (1995: 170-171) proposes four basic issues to be addressed in a user guide: type of information in specific components; the systematization of this information (grouping, placement and arrangement); the interrelation of information in individual components; and the quickest and easiest access to the information required.

The user guide should be written very carefully by the lexicographer — the sender of the message — who should take into account the linguistic, factual

and cultural competences of the target user — the receiver of the message — who is the primary beneficiary (Nielsen 2006: 11). If it is written for the benefit of lexicographers or academics, priority is given to detail being presented very scientifically, thus adopting a compiler perspective rather than a user perspective. In this case, it is highly unlikely that it would benefit the real user, which means that the quality of information is low (Nielsen 2006: 11).

The aim of our study was not to ask users whether they read the user guides of the dictionaries they consult or whether they have ever read a user guide, but to ask them to read user guides and comment on various aspects of the text. For this purpose, three Slovenian monolingual dictionaries were selected: *Slovar slovenskega knjižnega jezika*, 2nd edition (Dictionary of Standard Slovenian, hereafter referred to as DSS2), *Sinonimni slovar slovenskega jezika* (Dictionary of Slovenian Synonyms, hereafter referred to as DSSyn) and *Pravni terminološki slovar* (Dictionary of Legal Terminology, hereafter referred to as DLT). These three dictionaries belong to different categories according to a dictionary typology: DSS2 is a general monolingual dictionary; the DSSyn is a specialized dictionary dealing with a specific area of language, i.e. synonyms; and the DLT is a terminological dictionary treating the terminology of a special subject field. All three dictionaries were originally conceived and first published as printed dictionaries. Later, they were transferred to an online environment with very minor changes and can now be accessed free of charge via the Fran portal (<https://fran.si/>). Regrettably, only very few existing online dictionaries really use the technical possibilities of the electronic medium to full potential, whether in the conception and preparation of dictionaries or in access to and presentation of the data therein (Fuertes-Olivera and Bergenholtz 2011: 1). Since the emergence of online lexicography has heralded a new culture in dictionary use, lexicographers should take cognizance of this new culture (Gouws 2018: 233). However, the dictionaries included in our study do not make use of all the possibilities offered by the online environment. All three dictionaries under consideration are intended for native speakers of Slovenian; DSS2 and DSSyn can be used for decoding purposes, i.e., they can be consulted for comprehension when reading texts, as well as for encoding purposes, i.e., when writing or editing texts, whereas the DLT is aimed at decoding as well as acquiring knowledge about legal terms.

2. Front matter texts in the dictionaries studied

2.1 Dictionary of Standard Slovenian

The first edition of the DSS was initially published in five volumes over a period of 21 years (from 1970 to 1991). Later, a one-volume publication came out in 1994 with minor corrections (typos and other technical corrections). Since 2000, the Dictionary has been available online with free access. The second edition, whose front matter text is investigated in our study, came out in 2014 and is an

only partially revised version of the first edition.

The front matter of DSS2 consists of as many as 63 pages and includes the following: the copyright page and imprint; list of editors and lexicographers; Prefaces: *Preface to the Second Edition*, *Preface to the First Edition*, *Preface to the Single-Volume Edition*. All these parts together comprise 22 pages. Then follows the part entitled *Introduction*, which is subdivided into the following eight sections with subsections (41 pages altogether):

- *Some General Facts about the Dictionary* (+ two subsections)
- *Type of the Dictionary* (+ two subsections)
- *Structure of the Dictionary* (+ two subsections)
- *Semantic Part of the Dictionary Entry* (+ five subsections)
- *Grammatical Information in the Dictionary Entry* (+ six subsections)
- *Typography*
- *List of abbreviations and labels*
- *Miscellaneous*

The section entitled *Some General Facts about the Dictionary* gives an overview of the history of Slovenian lexicography, including the institutions where lexicographic work was carried out in the past. This lexicographic tradition served as a basis for the preparation of the first edition of the DSS. Tribute is paid to the first Slovenian lexicographers and to individuals who were the driving force behind the creation of the DSS.

The section *Type of the Dictionary* does not pertain to dictionary typology but presents the contents of the dictionary and provides information on the material not included in the dictionary, on the aims of the dictionary as well as on its scope and normativity. Very detailed data on the number of lemmata and sublemmata included in the dictionary are provided. DSS2 contains 97,669 lemmata and 14,181 secondary lemmata (a total of 111,850) and 157,978 senses and subsenses. The subsection *Material Used for the Compilation of the Dictionary* includes information on the sources and explains methods employed in the collection of material.

In the section *Structure of the Dictionary*, the macrostructure of the dictionary is explained and information on sublemmata and cross-references is given, which is followed by a detailed presentation of the structure of the dictionary entry.

The section *Semantic Part of the Dictionary Entry* focuses on the dictionary definitions ranging from the types of definitions to the characteristics and peculiarities of definitions. The arrangement of senses within the entry is explained in detail, followed by the presentation and arrangement of the illustrative material. At the end of the entry, two sections may be included, i.e., the phraseological section and the terminological section. The phraseological section includes various categories of phraseological units, whereas the terminological section contains terms and terminological units. The subsection *Labels* explains what a label is, how a user can interpret the scope of labels depending

on their position in the entry, what types of labels are used in the dictionary and what the peculiarities of labels are. Labels are subdivided into grammatical, semantic, terminological, stylistic, expressive, temporal and frequency labels and special normative labels; each category is thoroughly explained.

The section *Grammatical Information in the Dictionary Entry* provides instructions on how to interpret the first part of the dictionary entry up to and including the part-of-speech label. Then follow some orthographic considerations, phonological rules and the explanation of data presented in square brackets as well as very detailed information on stress by parts of speech.

The section *Typography* discusses typographical devices used in the dictionary, such as different fonts and font sizes, bold type, semibold type and italics. The punctuation marks used in the entries are fully explained and so are some special symbols, such as two slashes (//), one slash (/), dot (•) and diamond (◆).

The section *List of abbreviations and labels* enumerates part-of-speech labels, labels belonging to all categories of labels as well as other abbreviations used in the dictionary, such as *itd.* ('etc.'), *ž* ('f' — feminine gender), *os.* ('person'). The list is extensive and occupies two two-column pages.

The section *Miscellaneous* is subdivided into two subsections: one explains that the dictionary is alphabetically arranged and that foreign diacritical marks exert no influence on the alphabetical arrangement; the other one deals with special signs.

2.2 Dictionary of Slovenian Synonyms

The printed edition of the DSSyn starts with a two-page graphic presentation of the entries in the dictionary. Then follow the title page, the copyright page and imprint, the two-page *Foreword* and the *Structure of the Dictionary of Slovenian Synonyms*, which spans 23 pages. The total number of pages of the front matter texts amounts to 29. The same content can be found in the front matter of the online edition, the only exceptions being a much shorter graphic presentation of the entry and a text entitled *About the Online Edition of the Dictionary of Slovenian Synonyms*, which will be discussed separately.

The section *Structure of the Dictionary of Slovenian Synonyms*, which is of the utmost importance for the dictionary user, is subdivided into nine subsections, some being further subdivided into subsections. First, the content of the DSSyn is dealt with: what synonyms are, explanation of the common semantic meaning of a synonym set, semantic peculiarities of individual synonyms, (co-)hyponyms. Next, the material as well as the lexical and semantic criteria used in the compilation of the DSSyn are discussed at great length. Then follows the section entitled *Basic Arrangement of the Dictionary Text*, where types and various sections of the entry are explained clearly and concisely. The section *Dominant Dictionary Entry: Parts of the Entry Regarding Form and Content* gives a thorough presentation of the entry itself by discussing the lemma and the data on the

lemma itself, the synonym section and the role of individual components of the synonym section. Two further sections deal with the semantic individuality of synonyms in the synonym set and labelling of synonyms. The front matter concludes with a brief discussion of different parts of speech in synonymy and a final section dealing with the Slovenian lexicographic tradition and the DSSyn.

The section *About the Online Edition of the Dictionary of Slovenian Synonyms* included in the online edition only is aimed at explaining the use of possibilities offered by the online environment in comparison with the printed edition of the dictionary. It focuses on the dominant dictionary entry and explains how individual component elements of this type of entry are presented in the online edition. The online edition includes a special icon; by clicking on this icon, the user is shown information on the placing of a certain noun into the specific conceptual group.

2.3 Dictionary of Legal Terminology

The DLT contains front matter of 19 pages: the copyright page and imprint; *From Reviews of the Dictionary*; *Foreword*; *Slovenian Legal Terminology* with two subsections, i.e., *Development of the Slovenian Legal Terminology* and *Contemporary Slovenian Legal Terminology*; *About the Compilation of the Dictionary*; *Structure of the Dictionary*; *Abbreviations and Symbols*. The section *Structure of the Dictionary* is of the utmost importance for the user, especially from the perspective of dictionary use. This section is further subdivided into several subsections: *Type of Dictionary and Basic Information about the Dictionary* (with a list of subfields, from which terms are included in the dictionary); *Target Dictionary Users*; *Methodology*; *Time Span*; *Normativity*; *Structure of a Dictionary Entry*; *Definition*; *Explanation*; *Synonyms*; and *Cross-references*. This central section consists of seven pages of text accompanied by a number of examples taken from the dictionary.

3. Methods of obtaining users' views

Generally speaking, front matter and user guides in particular are intended for the user. Since the focus of our study is on the user, we carried out a small-scale research study among first-year MA students of English at the Department of English, Faculty of Arts, University of Ljubljana. The study was carried out within the framework of the course on lexicography as part of the students' home assignments. Since the front matter texts of the selected dictionaries are very long, we initially planned to concentrate on the user guides within the front matter, but none of the dictionaries studied includes a section with a straightforward title, such as "User guide" or "How to use the dictionary". That is why we decided to take the entire front matter, excluding only Prefaces and Forewords, and we therefore use the term front matter text(s) rather than user guide(s). The front matter texts were subdivided into several parts of approxi-

mately the same length, taking also the thematic component into consideration. Each student could choose which dictionary and which part of the front matter text to read.

The front matter text in DSS2 is the longest and was therefore subdivided into five parts. Four parts were read by three students and one part by four students, which means that sixteen students (including two non-native speakers) read the text. The front matter text in the DSSyn is shorter than that in DSS2 and was split into four parts. Each part was read by three students, i.e., twelve students altogether. The front matter text in the DLT is the shortest of all three front matter texts; that is why the text was divided into two and each part was read by two students, which means that four assignments were handed in. Altogether, thirty-two students participated in the study, two of them being non-native speakers of Slovenian; since the dictionaries investigated are intended for native speakers of Slovenian, the comments of these two students were excluded from the analysis. The comments and opinions of thirty students are presented and discussed in this article.

Even though the students were supposed to read only the selected segment of the front matter text, they were advised to refer to previous parts of the text if they came across unknown terms, since these terms could have been explained earlier in the text. They were asked to pay special attention to and comment on the following:

1. Intelligibility of the text;
2. Usefulness of the text from the point of view of intended users (what is useful, what is not useful);
3. Language used in the text (lexis including terminology and syntax);
4. Examples provided to illustrate the theoretical explanation (especially their presentation);
5. Length of the text as a whole.

They were also encouraged to make any critical comment that might cross their minds while reading the texts, express their opinion about the text as a whole and think of possible suggestions for the improvement of the text to make it more helpful.

4. Users' views and comments

4.1 Dictionary of Standard Slovenian

4.1.1 Terminology and syntax

Students reading the beginning of the front matter text think it is written in simple language and the terminology used is in accordance with the terminology taught and learnt in grammar school. One of the students says: "The language in the introduction is simple and intelligible. The author conveys a clear

message and does not use any unnecessary or subjective descriptions." Several students point out that terminology used in the text is explained (e.g., the explanation of the term *glava*, the term referring to the first part of the dictionary entry up to and including the part-of-speech label). Similarly, students generally like the beginnings of each section, as the text systematically defines terms used in that particular section and a few general facts are explained.

As they progress through the text, students start to be more critical, especially about excess (detailed) information, which is of no practical use for the general user, and about the large number of terms from different areas of linguistics with which they are not familiar. According to the majority of the students reading different segments of DSS2's front matter text, the terminology used is known only to experts in the field of linguistics and is thus far too difficult and inappropriate for an average user (e.g., *ekspiratorni naglas* — expiratory accent). In their opinion, this is the main reason that it is a real struggle for the user to plough through the text. However, some students point out that there are cases where the use of terms cannot be avoided (e.g., the explanation of a complex classification of labels).

Students also frequently comment on the length and complexity of sentences, which they mostly find too long and too complex; consequently, they are difficult to process and understand. However, one student remarks that, although the sentences are quite long, Slovenians are used to long sentences. He/She thinks that it is the abundant use of terminology that makes these long sentences difficult to comprehend.

One student clearly expresses his/her dissatisfaction with the style in which the entire text is written by saying that the text should be readily intelligible to all dictionary users, i.e., to linguists and non-linguists alike, but it seems that the information is given in a manner understandable only to linguists. He/she goes on commenting on the substantial number of (linguistic) terms and complexity of syntactical structures, which gives the impression that the lexicographers focus on the aesthetic value of the text rather than on coherence and intelligibility. Consequently, the cognitive effort needed to understand the text is unnecessarily much greater, which also results in the need for much more time to read the text. His/Her suggestion is to use short sentences and to repeat the same structures. He/She admits that the text would not be as artistically accomplished as it is. However, the text is not a work of art; on the contrary, its primary function is to provide instructions for the use of the dictionary. Therefore, the aesthetic side of the text is irrelevant, or in other words, functionality should be given priority.

4.1.2 Amount of information

Students mostly agree that the explanations found under various (sub)headings are too detailed, since an average user would not even bother to read the front matter text and would thus simply "ignore" a lot of information

included in the dictionary. Information overload causes numerous problems, especially when reading the text for the first time; on second reading, the text is slightly easier to understand. One student gives an example from the section on the scope of labels used in the dictionary, where the lexicographers explain in great detail how the differing positions of the labels exert influence on their scope. Given the many possible combinations, the user needs to read the text several times to memorize all the details. On the other hand, one student claims that the wealth of information is beneficial to somebody dealing with the language or somebody interested in lexicography; therefore, it is better to have too much information (TMI). Although being generally critical of the inclusion of TMI, students point out that the front matter text is well-structured, as it is divided into numbered paragraphs; the paragraph numbers are used to refer to other parts of the text, thus enabling easy access to other relevant information within the text.

Students believe that several sections of the front matter text are not needed and would probably be skipped by most users seeking to retrieve only the basic information relating to the actual use of the dictionary. Examples of what students consider inessential sections include: *Some General Facts about the Dictionary*, which contains a comprehensive explanation about the history and compilation of the DSS; *Characteristics of Definitions* and *Types of Definitions*. According to students, these sections might be of interest to an expert rather than to an average user and could be left out without detriment to the user.

4.1.3 Examples

Some students express criticism of the examples. One of them points out that the examples are provided only "indirectly". What this student refers to is that only the lemma is given as an example. This means that the reader of the front matter text needs to look up that specific lemma in the A-Z section and find the part of the entry exemplifying the theory. In many instances, this process might be not only time-consuming and "extremely impractical" but also quite demanding. Another problem indicated by quite a few students is that the choice of examples is not always appropriate, since the lemmata given are words unknown to an average native speaker (e.g., *bisus* 'byssus', *bantam* 'bantam(weight)'), which causes additional problems because the user does not always know what an example illustrates. On the other hand, students point out that there are more than enough examples, but they are not helpful if the user does not look up the word in the A-Z section. They agree that it would be much better to have fewer examples and suggest including (at least part of) the entry, with a clear indication which part of the entry is referred to in the theoretical explanation. One student also says that, although examples provided are relevant, they are of little or no help if the reader does not understand the theoretical explanation, which, in turn, is much too scientific.

Generally, it can be claimed that the students like the examples, since these can help in understanding the theory. One student commenting on the

explanation of labels says, for example, that the text contains short sentences and all labels are exemplified, which leads the user quickly to the desired information. The examples make it easier for the user to understand the message lexicographers are trying to convey. Students also noticed the typography of examples (i.e., italics), which they find useful, since it indicates clearly what is theory and what is the exemplification of the theory.

4.1.4 Proposals for improvement of the text

Students suggest the rearrangement of various sections in the front matter text. Sections that are absolutely essential for users to study if their look-up operations are to be effective, such as practical guidelines instructing the user how to use the dictionary, should be moved to the beginning of the front matter. Apart from that, they should be typographically distinguished from other sections of the front matter in order to immediately attract the user's attention. The students also propose the inclusion of a visual presentation of the dictionary entry with explanations of individual entry components, since they consider it useful and beneficial.

As for the use of linguistic terms, students prefer a Slovenian term to a foreign one (e.g., *protipomenka* instead of *antonim* 'antonym'). In the front matter text, many terms are explained, but there are still several terms that remain unexplained; to resolve this inconsistency, students speak in favour of a more consistent inclusion of explanations of terms. Another suggestion they make is to use paraphrases (e.g., 'the last sound of the word') rather than terms (e.g., *izglasje*). One of the students clearly states that the text is written far too scientifically; therefore, he/she suggests that if lexicographers believe so much terminology is necessary in the front matter text, it could prove helpful to compile a glossary of linguistic terms used in the front matter text with clear explanations of the concept behind the terms. The glossary, according to the student, should be part of the front matter.

4.2 Dictionary of Slovenian Synonyms

4.2.1 Terminology and syntax

Throughout the front matter text, very specific terminology is used; however, some students maintain that, despite this, the text is relatively comprehensible. Most of the terms are explained, and if the user/reader is interested in the topic and reads the text carefully and slowly, he/she can understand the gist of what the text is about. In some parts of the text, they found terms that are not explained and can be a "big mystery" for someone unfamiliar with linguistic terminology (e.g., denotative meaning). One student puts it nicely that the complexity of the lexis is "a hindrance to the average user as it encumbers the reading, forcing the reader to spend more time and effort on a segment which should be handy and useful".

A parallel can be drawn between the observations of the students reading the front matter text of DSS2 and those reading the DSSyn regarding the complexity of syntax used. The front matter text of the DSSyn is also syntactically demanding (sentences that are too long and too complex), which is likely to discourage users from reading it before they start to use the dictionary.

4.2.2 Examples

Theoretical explanations are exemplified, but — similar to DSS2 — the examples are mostly provided in the form of lemmata and only occasionally in the form of full entries or excerpts from entries. Students generally consider the examples helpful, but make some critical remarks about the typography of examples. They note that examples differ typographically from the explanations, the problem being that some examples are bolded and some italicized (cf. also Farina, Vrbinc and Vrbinc 2019). They find the system of marking the examples in two different ways confusing and suggest the use of the same typography for all examples. Another proposal made by students concerns the consistent inclusion of full entries or at least the relevant part of the entry instead of lemmata. This suggestion is in agreement with that of the students who read the front matter text of DSS2.

4.2.3 What the students like

Several students emphasize that the text is visually well-organized, as it is divided into coherent paragraphs that are mostly not too long. The main text is accompanied by keywords printed in red on the left- or right-hand side of the text in an extra column and positioned at the beginning of the corresponding paragraphs. These keywords guide the reader through the text and make it easier for the user to find a particular topic. Instead of reading the entire text, readers can simply look for the appropriate key word(s) and read only those segments they are interested in rather than the whole text.

Two other points are made by students regarding segments of text they find useful. One of these is the section dealing with the use of the synonyms provided in the dictionary, since it explains how synonyms can be incorporated in other contexts. The other one refers to the inclusion of pronunciation accompanying foreign words whose pronunciation is not predictable for native speakers of Slovenian.

4.2.4 Proposals for improvement of the text

Apart from some proposals targeting the inclusion and treatment of examples (cf. 4.2.2), one of the main issues addressed by the students concerns the use of (too) many technical terms that are not intelligible to English majors, let alone to average users who are non-linguists. To resolve this, one student who thinks

that the terms should be explained suggests providing explanations by means of footnotes. In this way, the additional explanations would be excluded from the text proper, thus making the text more readable and understandable. A suggestion made by one student is to have two parallel front matter texts: a simplified version intended for a general user and a more detailed and technical version for "more curious and demanding users", a suggestion echoing a point made by Nielsen (1995: 169).

Students reading the part of the front matter text that explains labels used in the DSSyn seem to dislike the way labels are presented and explained. Labels are grouped according to categories expressing diasystematic information, which poses problems to an average user unfamiliar with the existence of the theoretical classification of labels. Not being familiar with the classification of labels, it is next to impossible for the user to locate a label he/she is trying to find. To settle this issue, students suggest listing the labels in a table of labels and arranging them alphabetically, together with appropriate examples in the form of entries.

Students frequently refer to the problem of the length of the text and point out that there is plenty of material in the front matter that could easily be omitted without detriment to users. Generally, they like the structure of the text with its subdivision into sections and subsections, as well as good examples illustrating the theoretical explanations. Nevertheless, they think the text is difficult to read and consequently not comprehensible to the average reader. Most of the information provided is relevant and useful for users of this dictionary, but under one condition only: If they manage to understand the point made in a particular part of the text. Their common observation is that, as far as the text is concerned, "much can be done to improve its intelligibility". Given the unintelligibility and length of the text, several students would stop reading it as soon as they found the part that would help them to use the dictionary. To improve access to and retrieval of information needed, one student suggests that it would be useful to have "an index of a certain kind that would enable the user to find the necessary information quickly". Several students agree that the text has several superfluous paragraphs: e.g., the origins of synonymy and sources used for the compilation of the dictionary, since "on the entire page and a half, they only mention the DSS as their source and material".

One student makes an interesting remark about the entire text: "It seems like a textbook (e.g., a list of types of synonyms) rather than a dictionary. While reading the text, one has the impression that they have to remember all the details as if they were going to be examined at school the next day".

4.3 Dictionary of Legal Terminology

4.3.1 Terminology and syntax

With the exception of some terms (e.g., *kodificirati* 'codify', *tezaver* 'thesaurus', *ekscerptor* 'excerptor') that are not explained and would therefore constitute

quite a challenge to some users, terminology used in the front matter text is not too difficult to comprehend. Students, however, critically observe that if terms are used, they are not all equally well-defined. They list legal terms lacking explanations (e.g., *sodna praksa* 'case law'; *običajno pravo* 'customary law'; *kanonsko pravo* 'canon law'; *pravna doktrina* 'legal doctrine') as well as terms that they find adequately explained (e.g., *terminološki dogovor* 'terminological agreement'). Since the dictionary is primarily intended for (semi-)experts in law and for lay-people interested in this subject field, the legal terminology used in the front matter text "needs no explanation".

Regarding syntax, students reading the front matter text in the DLT come to the same conclusion as their peers reading the front matter texts in DSS2 and DSSyn: The syntax is difficult, the main problem being the length of sentences which are "hard for a reader to follow". One student notes that on several occasions, the authors of this front matter use "incredibly convoluted syntactic structures to form sentences".

4.3.2 What the students like

The students mostly like the structure of the front matter text, which is divided into sections and subsections enabling readers to focus on the specific item of information they need. The division of the text is a useful feature, since an average user is unlikely to read the front matter text from beginning to end.

Students note that the exact number of lemmata included and treated in the dictionary is provided (i.e., 10,094 lemmata), a piece of information they really find interesting, since it is rarely encountered in general dictionaries nowadays. Other interesting features identified by the students are as follows: a clear explanation of the type of dictionary; a detailed list of areas of law from which the legal terms included in the dictionary are taken; and the list of sources (printed and electronic ones, dictionaries and parallel corpora as well as terminological databases) with Slovenian legal terminology.

4.3.3 Proposals for improvement of the text

The section *About the Compilation of the Dictionary* includes a list of people who contributed to the dictionary; it occupies as many as one and a half pages. Students believe that most people reading the front matter would simply not bother reading the names, so the names of contributors would "go unnoticed". They suggest moving the list of contributors either somewhere else in the front matter or even to the back matter, since it interrupts the flow of the text describing how the dictionary was compiled.

Overall, students consider the text interesting, though probably too detailed for an average user. They agree that there is nothing wrong with the length of the text as long as it is organized clearly and divided into sections. As opposed to some students who like the structure of the text (see 4.3.2), one stu-

dent suggests that the user would benefit more if the front matter text was organized more clearly and divided into more sections. The same student also proposes that paragraphs should be broken into shorter ones (some paragraphs extend over half a page), since this, combined with more sections, would contribute to improved readability of the text. One of the students nicely summarizes the problem of length and TMI as follows: "The text — albeit relatively lengthy — offers thorough explanations of every constituent part of the dictionary. What it lacks in compactness, it makes up for in attention to detail."

5. Discussion and conclusion

Students commenting on the same front matter text and even on the same segment of the front matter text often express contradictory views on the use of terminology. All agree that the texts abound in linguistic terms most users would not understand. It is, however, interesting to note that some students say that all or most of the terms are explained, while others claim terms are not explained at all. There may be several explanations for these contradictory claims. Some students may simply ignore the explanations provided, or terms are explained in other segments of the front matter text they were not supposed to read. Prior to their task, students were warned about this and were told to try to find unknown terms in the previous parts of the text. The most likely reason, however, is that they could not remember the explanations for such a great number of terms used in the text and believed that they had not previously come across a specific term, when in fact, the term had been explained earlier in the text.

Students reading the front matter text in the DLT generally agree that terminology is not abundantly used and, if it is, the terms are explained. In DSS2 and DSSyn, however, the abundant use of terminology is one of the greatest challenges faced by any reader of the texts. This problem affects not only the average dictionary user but even the tiny minority of users with linguistic education. One student who clearly expresses his/her dissatisfaction with the comprehensibility of the text claims that a reader of the text should have at least a BA in Slovenian to fully comprehend the text. Interestingly, another student who also had comprehension problems went one step further and asked his/her friend who has a BA in Slovenian and who is currently an MA student of Slovenian whether he/she was familiar with the terms used in the segment of the text dealing with accent (tonemic accent, dynamic accent, etc.) and his/her colleague responded it was the first time he/she had ever heard of them. He/She concludes that, if the front matter text is intended neither for a general user nor for an expert, the question remains: Who is it actually for?

The length of the front matter texts studied is also a problem commented upon by a number of students, including those reading the front matter text in the DLT, the shortest front matter text of all three dictionaries. Our students often say that they would never have read the front matter if it had not been for

an assignment because it is too long, and one of them concludes with resignation: "No wonder no one bothers to read the front matter". Obviously, users want to find the necessary information immediately and not wade through information of no interest; in other words, they do not want TMI. One conclusion that can be drawn from the students' criticism is that lexicographers should consider including only the essential information in the front matter text — all texts that are not absolutely necessary should either be omitted altogether or placed separately and not as part of the user guide. As Varantola (2003: 233) admirably puts it, "user instructions are hands-on sections that should be highly sensitive of user behaviour and strategies and thus be written with the user's 'discovery procedure' in mind".

Without a shadow of a doubt, the front matter texts in DSS2 and DSSyn in particular were not written with a user in mind — they were written for an expert in linguistics or possibly even for the compilers themselves! This is in line with Varantola (2003: 233), who explicitly states that "[w]hat is often wrong about the user manual sections in dictionaries is that they are written from the compiler's point of view". Nielsen (2006: 10) points out that it is difficult to know whether the reason behind the user's failure to read the front matter is "directly linked to their perception of the quality and information value of those texts seen from the user's point of view". He goes on by saying that, if users believe the texts are of low quality and consequently decide not to waste time reading them, the texts "do not serve any purpose". In our case, we definitely cannot say that the texts are poorly written — on the contrary, as many of our students observe, the texts are precise, informative and well written, but, unfortunately, far too difficult for an average user to understand. So Nielsen's claim about such texts not serving any purpose certainly applies to the front matter texts in the three dictionaries studied, especially to those in DSS2 and DSSyn. As has already been pointed out in Section 1, the sender of the user guide is the lexicographer and the receiver is the user. We cannot but agree with Nielsen (2006: 10), who says that "it is important that lexicographers keep this in mind when they write the user guide so that they do not produce a text that looks as if it has been written for their own benefit or that of their academic colleagues, but instead a text that appears to have been written for the benefit of the reader/user". The user reading the front matter text cannot pose questions to the lexicographers who wrote the text, which means that these texts are clearly an example of one-way communication (Nielsen 2005: 141; Nielsen 2006: 10).

Interestingly, if we focus on the style of writing used in the front matter texts under consideration, we clearly notice that the texts are written as a comprehensive description of individual information categories rather than providing information directly to the user. In no part of the texts studied do lexicographers address the user personally by using the second person, which is "a clear signal to the user that the text is addressed to him or her" (Nielsen 2006: 17). Nielsen's recommendation is completely ignored in the front matter texts

under consideration.

One of the positive features identified and highlighted by the students is the systematic nature of all the front matter texts they read. In particular, they refer to the neat arrangement of material in numbered paragraphs in DSS2 and the use of the red keywords in the special column in the DSSyn. In this way, the lexicographers have the possibility to make cross-references to other paragraphs, since the inclusion of cross-references whenever relevant assists use and understanding. For the users, it is easy to follow the cross-references and find the specific place referred to.

Another issue that needs to be addressed is that none of the front matter texts under consideration includes a graphic presentation of the entry (the only exception being the graphic presentation of a single entry in the DSSyn; cf. also 2.2), an element that is almost a must if we are to visually convey the structure and layout of the dictionary entry. The graphic presentation of the entry should be accompanied by explanations focusing on various dictionary functions. It is true that a graphic presentation might be "less satisfying linguistically and philosophically", but it definitely "takes significantly less time to absorb than elegant, discursive prose" (Kirkpatrick 1989: 756). Not only would visually oriented users greatly benefit from this, but it is also a way to present different information categories in the most user-friendly way, which enables the users to retrieve the necessary information as quickly as possible. This is why graphic presentation is essential in both printed and online dictionaries.

Our students were forthcoming in explaining the obstacles they faced when confronted with the task of reading the front matter texts and commenting on various aspects of the text. The perceptions of these advanced users — English majors — are valuable to those seeking to improve the front matter texts, or more specifically, the user guides within the front matter, in Slovenian dictionaries, as well as those writing user guides of any other dictionary regardless of the language(s) included or treated in the dictionary. While it is definitely a challenge to write a text that sufficiently explains the different aspects of dictionary functions in language that the intended user is likely to understand, we cannot but agree with Nielsen (2005: 145), who says that if lexicographers write user guides perceived by the user as beneficial and necessary, "this may herald a new era for the user's guide".

References

Dictionaries

Collins English Dictionary: <https://www.collinsdictionary.com/dictionary/english>. Accessed on 6 March 2019.

Pravni terminološki slovar (Dictionary of Legal Terminology): <https://fran.si/iskanje?FilteredDictionaryIds=211&View=1&Query=%2A>. Accessed on 12 February 2019.

Sinonimni slovar slovenskega jezika (Dictionary of Slovenian Synonyms): <https://fran.si/iskanje?FilteredDictionaryIds=208&View=1&Query=%2A>. Accessed on 21 February 2019.

Slovar slovenskega knjižnega jezika (Dictionary of Standard Slovenian): <https://fran.si/iskanje?FilteredDictionaryIds=133&View=1&Query=%2A>. Accessed on 16 January 2019.

Other literature

Atkins, Beryl T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Bergenholtz, Henning. 2012. What is a Dictionary? *Lexikos* 22: 20-30.

Bergenholtz, Henning, Sven Tarp and Herbert Ernst Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, Lothar, Hartwig Kalverkämper, Herbert Ernst Wiegand, Christian Galinski and Werner Hüllen (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologie-wissenschaft / Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Vol. 2: 1762-1832*. Berlin/New York: Walter de Gruyter.

Farina, Donna M.T.Cr., Marjeta Vrbinc and Alenka Vrbinc. 2019 [in press]. Problems in Online Dictionary Use for Advanced Slovenian Learners of English. *International Journal of Lexicography*. doi: 10.1093/ijl/ecz017.

Fuertes-Olivera, Pedro A. and Henning Bergenholtz. 2011. Introduction: The Construction of Internet Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 1-16. London/New York: Continuum.

Gouws, Rufus. 2018. A Dynamic Lexicographic Practice for Diverse Users and Changing Technologies. Mesthrie, Rajend and David Bradley (Eds.). 2018. *The Dynamics of Language. Plenary and Focus Lectures from the 20th International Congress of Linguists*: 231-244. Cape Town/Claremont: UCT Press.

Hartmann, Reinhard Rudolf Karl. 2001. *Teaching and Researching Lexicography*. Harlow, Essex: Pearson Education.

Hartmann, Reinhard Rudolf Karl and Gregory James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.

Kirkpatrick, Betty. 1989. User's Guides in Dictionaries. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography. Volume 1*: 754-761. Berlin/New York: Walter de Gruyter.

Klosa, Annette and Rufus H. Gouws. 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.

Landau, Sidney I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.

Müller-Spitzer, Carolin. 2013. Textual Structures in Electronic Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 367-381. Berlin: De Gruyter.

Nielsen, Sandro. 1995. Dictionary Components. Bergenholtz, Henning and Sven Tarp (Eds.). 1995. *Manual of Specialised Lexicography*: 167-187. Amsterdam/Philadelphia: John Benjamins.

- Nielsen, Sandro.** 2005. User's Guides. Barz, Irmhild, Henning Bergenholtz and Jarmo Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 137-146. Frankfurt am Main: Peter Lang.
- Nielsen, Sandro.** 2006. A Functional Approach to User Guides. *Dictionaries: Journal of the Dictionary Society of North America* 27: 1-20. doi:10.1353/dic.2006.0006.
- Svensén, Bo.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Varantola, Krista.** 2003. Linguistic Corpora (Databases) and the Compilation of Dictionaries. Van Sterkenburg, Piet (Ed.). 2003. *A Practical Guide to Lexicography*: 228-239. Amsterdam/Philadelphia: John Benjamins.

Monitoring Academic Studies of Turkish Lexicography: A Bibliometric Study of 84 Years*

Ferdi Bozkurt, *Turkish Language and Literature Department, Open Education Faculty, Anadolu University, Eskişehir, Turkey*
(ferdib@anadolu.edu.tr)

Abstract: The aim of this article is to report the main trends of the research texts written in Turkish in the field of Turkish lexicography. This report will also provide an opportunity to discuss the main trends in Turkish lexicography and the areas that need to be developed. The study is a descriptive research and literature review that is limited to the texts written in the Turkish language on Turkish linguistics during the years 1932–2016. The Turkish Academic Network and Information Center (ULAKBİM), Google Scholar, the Turkey Council of Higher Education Thesis Center, the Turkey National Collective Catalog, National Library of Turkey, EBSCOhost, ProQuest Dissertations & Theses Global databases and 68 academic journals published on Turkish linguistics were searched for the study. The database contains 1001 texts. Findings related to these texts show that studies on lexicography have increased in recent years. Moreover, the findings show that there is a lack of some text types.

Keywords: TURKISH LEXICOGRAPHY, LEXICOGRAPHIC TEXTS, BOOKS, MASTER'S THESES, DOCTORAL DISSERTATIONS, REVIEWS, NEWS, RESEARCH TRENDS, BIBLIOMETRICS, CITATIONS, SCIENTIFIC IMPACT

Opsomming: Monitering van akademiese studies oor die Turkse leksikografie: 'n Bibliometriese studie van 84 jaar. Die doel van hierdie artikel is om verslag te doen oor die hooftendense in die navorsingstekste geskryf in Turks op die gebied van die Turkse leksikografie. Hierdie verslag sal ook die geleentheid bied om die hooftendense in die Turkse leksikografie en die areas wat ontwikkeling benodig, te bespreek. Dié navorsing is 'n beskrywende navorsings- en literatuurverslag wat beperk is tot die tekste wat in Turks oor die Turkse linguïstiek gedurende die jare 1932–2016 verskyn het. Die Turkse Akademiese Netwerk- en Inligtingsentrum (ULAKBİM), Google Scholar, die Turkse Raad van Hoër Onderwys se Tesissentrum, die Turkse Nasionale Gesamentlike Katalogus, die Nasionale Biblioteek van Turkye, EBSCOhost, ProQuest Dissertations & Theses Global se databasisse en 68 akademiese joernale wat in die Turkse linguïstiek gepubliseer word, is vir hierdie studie deursoek. Die databasis bevat 1001 tekste. Bevindings

* An earlier version of this article was presented at the 12th International Conference of the Asian Association for Lexicography (ASIALEX 2018), which was hosted by the King Mongkut's Institute of Technology Ladkrabang in Krabi, Thailand, 8–10 June 2018.

rakende hierdie tekste dui daarop dat studies in die leksikografie in die afgelope jare toegeneem het. Bowenal toon die bevindings dat daar 'n gebrek aan sommige tekssoorte is.

Sleutelwoorde: TURKSE LEKSIKOGRAFIE, LEKSIKOGRAFIESE TEKSTE, BOEKE, MEESTERSPROEFSKRIFTE, DOKTORALE VERHANDELINGS, RESENSIES, NUUS, NAVORSINGSTENDENSE, BIBLIOMETRIE, SITATE, WETENSKAPLIKE IMPAK

1. Introduction

In the digital age we are in, more and more information is being produced every day. Information is expanding in a variety of fields, such as education, culture, art, technology and science, and is spreading faster than ever before this century. The case is also inevitably the same in linguistics. More information is produced in the field of lexicography, where some researchers are still debating whether it is a branch of science itself or a sub-branch of linguistics (Atkins and Rundell 2008; Bergenholtz and Gouws 2012; Landau 1984; Tarp 2012; Zgusta 1971). As a result of academic studies in lexicography, texts like books, articles, presentations, reviews, etc. have emerged.

1.1 Lexicographic tradition of the Turkish language

The first known dictionary study for the Turkish language began with Mahmut Kashgar. He completed *Divânu Lüğati't-Türk* (Dictionary of Turkish Languages) in February 1074 (Ercilasun and Akkoyunlu 2014). *Divânu Lüğati't-Türk* accepted that the basis of Turkish lexicography accords well with the principles of Arabic lexicography.

All dictionaries prepared in the Ottoman period from 1074 to 1901 were translated from Arabic and Persian. And in these dictionaries, headwords were of Arabic and Persian origin. These dictionaries were prepared according to the principles of the Arabic lexicographic tradition (Yavuzarslan 2017). One of the main reasons for this situation was the need for language learning among different nations as a result of the increase in commercial, scientific and religious interaction of Muslim Turks with Arabs and Persians.

In the Turkish lexicographic tradition, the most typical example of the transition from Arabic lexicography to Western lexicography is the *Kamus-ı Türkî* published in 1901. The lexicographic concept in *Kâmûs-ı Türkî* is based especially on the works of the late 1800s, the last period of the Ottoman Empire. Some researchers (Dilaçar 1953; Gökçe 1998; Parlâtır 1995) studied bilingual or annotated dictionaries of this period. While the dictionaries before *Kâmûs-ı Türkî* contained only Arabic or Persian headwords, Turkish headwords were used for the first time in *Kâmûs-ı Türkî*.

The dictionaries compiled before the Republic of Turkey emerged as the result of individual efforts. Normally, dictionaries compiled by means of group work are regarded as more scientific. However dictionaries such as *Lehçetü'l-lügat* (1796) compiled by Şeyhülislam Esad Efendi and *Kâmûs-ı Türkî* (1901) compiled by Şemseddin Sami are accepted masterpieces of Turkish lexicography.

Mustafa Kemal Atatürk made great strides in language studies as in many other fields of science after he had founded the Republic of Turkey on October 29, 1923. Atatürk, by both creating official institutions and participating personally in Turkish linguistic studies, coined new Turkish geometry terms to replace the geometry terms of Arabic–Persian origin from the Ottoman tradition. He also wrote a geometry book using the new terminology. Moreover, he provided government support for Turkish linguistic studies. The Turkish Language Association, which was established on July 12, 1932, triggered linguistic studies such as morphology, syntax, lexicology, phonetics, semantics, and etymology. Since linguistic studies were largely supported financially by the Republic of Turkey, there was a significant increase in scientific studies. As a result of these studies, books, doctoral dissertations, articles, presentations, bulletins, and scientific research projects were produced.

1.2 Bibliometrics

According to Day and Gastel (1995: ix) "the goal of scientific research is publication". Publications as a result of scientific research are therefore absolutely important. The mathematical and statistical review of the publications by means of scientific methodology, in other words, by determining their quantitative distribution by using statistical analysis, will contribute to determine the direction of development in lexicographic studies.

It is important to illustrate the impact of scientific studies such as articles, master's theses, doctoral dissertations and presentations written in the field of science.

The most important contributions of illustrating this scientific trend is to:

- indicate the distribution of these texts according to their types and citations,
- show the effects of the institutions contributing to the creation of the texts and the characteristics of the journals that provide the texts to the researchers, and
- identify the researchers producing the texts and the types of their scientific texts that contribute to the field.

It was Alan Pritchard who first introduced the term *bibliometrics* into the scientific literature in 1969. Pritchard (1969: 348) described bibliometrics as the "application of mathematics and statistical methods to books and other media of com-

munication". This explanation was extended by Potter (1981: 5) who redefined it as: "Bibliometrics is, simply put, the study and measurement of the publication patterns of all forms of written communication and their authors." Therefore, the relationships between the authors and the texts are also included in the concept. With the help of bibliometrics, it is possible to quantify the number and distribution of contributions in the literature. This approach is very useful to determine the development of knowledge in the literature.

Hulme (1923) believed that "statistical treatment must show the existence of phases of activity and retardation in each science and might even indicate approximately the period when its ultimate boundaries would be reached". This issue is very important for academic professionals in terms of the formation of the literature in their field of study because it is necessary to examine previous studies and conduct field research for today's scientific approaches. In order to see the scientific impact of the studies conducted in different fields, very important measurement platforms such as the Science Citation Index (SCI), established in 1955 by Eugene Garfield (1955), were created. These indexes facilitate the acquisition of information on the effectiveness of scientific publications and have become systematic over time.

Nowadays, bibliometric studies are carried out on a wide range of disciplines and it is possible to encounter bibliometric studies on linguistics. For instance, Lei and Liu (2018) conducted a study which covers 42 journals of applied linguistics mentioned in the Social Science Citation Index (SSCI). Mohsen et al. (2017) conducted a study on linguistic publications which appeared during the years 2005–2014 and which were included in the Thomson Reuters' Social Science Citation Index (SSCI). Arik (2015) reviewed the articles published during 1900–2013 and mentioned in the Social Science Citation Index (SSCI), and discussed the characteristics of the linguistic articles included in the Arts and Humanities Citation Index (A and HCI) between 1975 and 2013. Lei and Liao (2017) examined the linguistic publications by researchers from China, Hong Kong, Taiwan, and Macau between 2003 and 2012, and they identified high-impact and popular journals by applying the bibliometrics method regionally.

1.3 Bibliometrics in lexicography

In the field of lexicography, De Schryver (2009a; 2009b) conducted analysis studies on internationally important journals such as the *International Journal of Lexicography* and *Lexikos*. In the first of his articles, De Schryver (2009a) built a corpus from the studies published in the *International Journal of Lexicography* (IJL), which is one of the most important journals in the field of lexicography, and conducted a detailed content analysis as well as a citation analysis accordingly. He made comparisons between the important journals of linguistics (*Linguistics* and *Applied Linguistics*) and lexicography (*Lexikos* and *Dictionary*). In his second article, De Schryver (2009b) conducted research on *Lexikos*, making a

statistical analysis of this journal covering 18 years. Both of these studies are very important in terms of presenting contributions to the field of lexicography.

The number of bibliographic studies in the field of Turkish lexicography are rather limited. In this field, generally, bibliographic studies about the prepared dictionaries and the reviews of these dictionaries are made (Eminoğlu 2010; Kotan 2017). The study of Yıkmaş and Sazak (2017) identified 31 book chapters, 106 theses, 188 presentations, 23 reviews and 219 articles related to lexicography. The study of Yıkmaş and Sazak (2017) is limited to the years 2000 and 2016. These two researchers divided the texts into text types such as books, book chapters, doctoral dissertations, master's theses, presentations, reviews and articles.

In the mentioned studies, it is not possible to find a large literature on Turkish lexicography. In addition, the impact and type of the publications, and cooperative ways among the authors were not discussed statistically. The original aspect of this study lies in the fact that it is the first bibliometric study on Turkish lexicography. One of the strengths of the study is the bibliometric analysis of lexicography at the level of a particular country through all available scientific texts produced in the field of Turkish lexicography.

The main objectives of the current study are to:

- find out the authorship pattern,
- calculate the degree of collaboration,
- determine distributions of the institutional contributions to the texts,
- find out the self citation ratio of an author,
- determine the most cited texts, and
- determine the most productive authors.

2. Data and methodology

This study involves texts produced between the years 1932 and 2016. The reason for adopting 1932 as the starting year is that the Turkish Language Institution was established in that year.

It was necessary to create a list of keywords and a set of criteria for the data set used in the study because there are thousands of Turkish articles, presentations and theses produced in the field of morphology, syntax, phonetics, phonology, psycholinguistics, sociolinguistics, computational linguistics, historical linguistics, applied linguistics, etc.

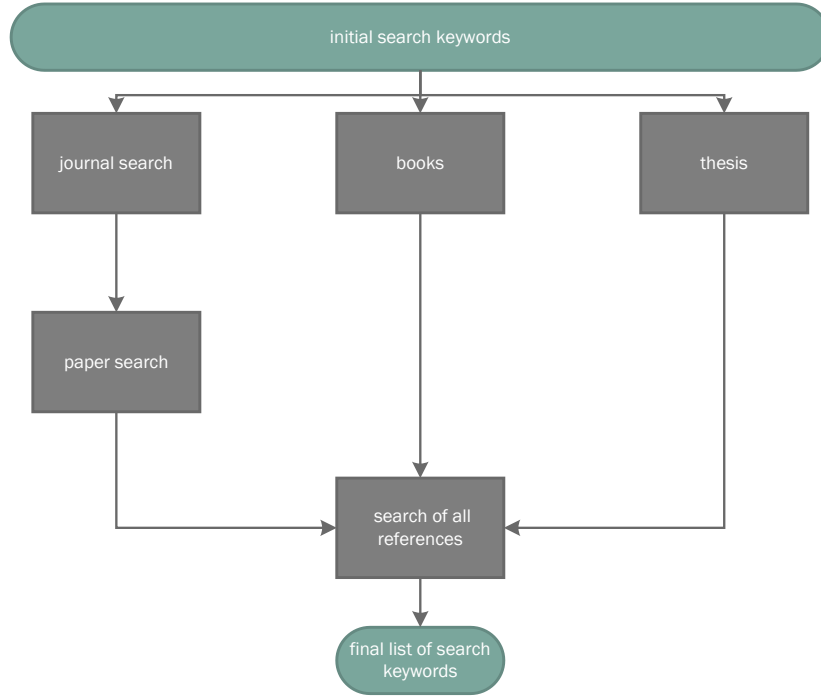


Figure 1: Determining the final search keyword list

In order to differentiate the texts related to lexicography, firstly, *sözlük* (lexicon) and *sözlük bilim* (lexicology) were determined as initial search keywords. These initial search keywords were queried in 68 journals, written in Turkish and published in the field of linguistics and grammar. Afterwards, queries were made in Google Scholar with these initial keywords. In the obtained texts, it was found that the Turkish word *sözlük* (dictionary) and the Arabic word *lügat* (dictionary), and *sözlükbilim* and *leksikografi* were used synonymously. The word *sözlük bilim* (lexicography) was used with different spelling variants such as *sözlükbilim*, *sözlükbilimi* and *sözlük bilimi*. In addition, the texts in the reference list sections of the articles encountered during the queries made with initial keywords were also examined. Thus the following were accepted as final search keywords: *sözlük* (dictionary), *lügat* (dictionary, an old usage), *sözlükbilim* (lexicography), *sözlük bilim* (lexicography), *sözlükbilimi* (lexicography), *sözlük bilimi* (lexicography), *sözlükçülük* (synonymous with lexicography), *leksikografi* (lexicography).

As a result of these inquiries, a process sequence was followed to create the final keyword list.

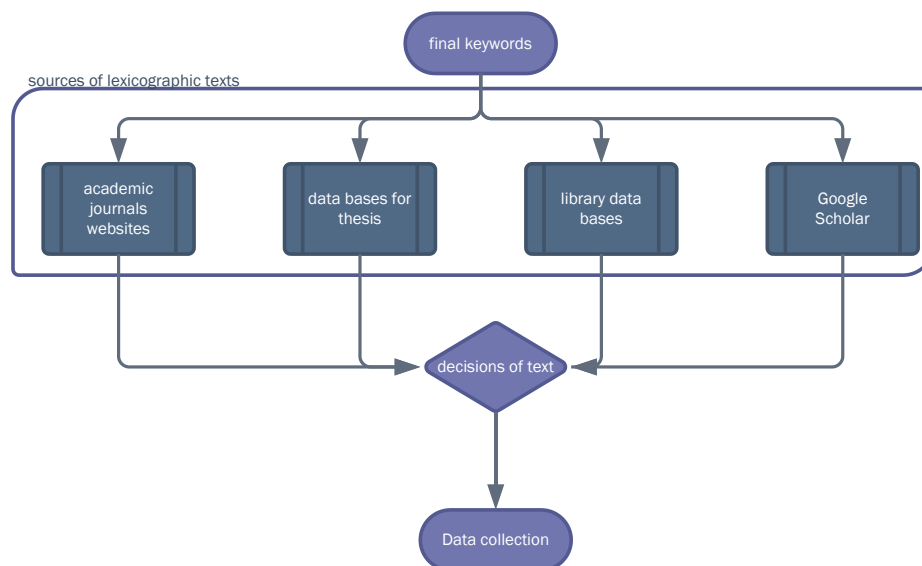


Figure 2: Selecting the target texts

In order to investigate the articles, the Turkish Academic Network and Information Center (ULAKBİM)², Google Scholar³, EBSCOhost⁴, the Turkey National Collective Catalog⁵, the National Library of Turkey⁶, websites of 76 journals publishing articles on linguistics and Turcology, and the reference list of determined articles were searched.

The Turkey Council of Higher Education Thesis Center⁷ and the ProQuest Dissertations & Theses Global⁸ were scanned for master's theses and doctoral dissertations.

The books were determined after analyzing the reference lists of the articles found in Google Scholar and in the libraries of Anadolu University and Eskişehir Osmangazi University.

Approximately 2,873 scientific texts were obtained following the searches. However, in most of the texts, although one of the final keywords was *sözlük* (dictionary), it was found that the text was not related to lexicography when it was examined. Another case is that in the Turkish abstract and keywords list of 124 articles whose main texts were written in English, German and French, dictionary and lexicography words were found; however, these texts were removed from the database of this study since the database would be composed of articles written only in Turkish. Then, the lexicographic text decision stage shown in Figure 2 was in progress. The researcher decided whether the content of the texts was related to lexicography or not by examining the content of the texts one by one.

Table 1: Text types determined for the current study

Text Type	Number of Texts	%
Doctoral dissertations	12	1.20
Master's theses	39	3.90
Published oral presentations	301	30.07
News	21	2.10
Books	3	0.30
Articles	475	47.45
Reviews	150	14.99
Total	1001	100.00

As a result of the last stage, the texts in Table 1 have formed the database of this research.

3. Data analysis and interpretation

Frequency and percentage values of publication years are shown in Table 2:

Table 2: Number of publications by year

Year	Frequency	%	Year	Frequency	%	Year	Frequency	%
1934	1	0.10	1973	2	0.20	1996	9	0.90
1935	2	0.20	1974	2	0.20	1997	5	0.50
1936	3	0.30	1975	13	1.30	1998	15	1.50
1939	2	0.20	1976	5	0.50	1999	31	3.10
1942	1	0.10	1977	6	0.60	2000	16	1.60
1952	2	0.20	1979	1	0.10	2001	5	0.50
1953	3	0.30	1980	2	0.20	2002	18	1.80
1954	6	0.60	1981	2	0.20	2003	8	0.80
1956	2	0.20	1982	3	0.30	2004	22	2.20
1957	1	0.10	1983	1	0.10	2005	12	1.20
1959	3	0.30	1984	1	0.10	2006	21	2.10
1960	3	0.30	1985	2	0.20	2007	63	6.29
1961	1	0.10	1986	2	0.20	2008	84	8.39
1962	2	0.20	1987	2	0.20	2009	111	11.09
1963	1	0.10	1989	1	0.10	2010	59	5.89
1965	2	0.20	1990	1	0.10	2011	68	6.79
1967	2	0.20	1991	4	0.40	2012	46	4.60
1968	1	0.10	1992	1	0.10	2013	70	6.99
1969	2	0.20	1993	1	0.10	2014	52	5.19
1970	1	0.10	1994	5	0.50	2015	61	6.09
1971	4	0.40	1995	8	0.80	2016	98	9.79
1972	17	1.70						
			Total			1001 100.00		

Table 2 shows the distribution of texts in the field of Turkish lexicography by year. As can be seen in the table, the years when most texts about lexicography were produced in Turkey are the following: 2009 (111 texts), 2016 (98 texts), 2008 (84 texts), 2013 (70 texts), 2011 (68 texts). In some years (1937, 1938, 1940, 1941) no texts were produced. However, texts have been produced annually and uninterruptedly in the field of lexicography since 1967. This study covers the years between 1932 and 2016, the average number of texts being 11.9 during these years. As shown in Table 2, only one text was produced in some years (1968, 1983, 1984 etc.). In addition, as can be seen in the table, an increase has been observed in the number of texts related to lexicography over the last 20 years.

3.1 Lexicographic research texts

3.1.1 Doctoral dissertations

By limiting the years to 1932–2016 in the Turkey Council of Higher Education Thesis Center, 252 master's theses and 50 doctoral dissertations were found as a result of the searches in which *lûgat* (dictionary) and *sözlük* (dictionary, an old usage) were the keywords. During the search, it was found that 25 doctoral dissertations were not accessible. Master's theses and doctoral dissertations prepared in the field of history, geography and Islamic sciences and not related to lexicography were emerging as a result of the search. The reason for this was the dictionary sections in the form of indexes added by the researcher to the last parts of the theses. In addition, especially in the theses prepared for the transfer of old texts written in Arabic letters to modern Turkish, researchers added dictionary sections to the last sections of the studies. This resulted in high numbers of search results. On the other hand, it was observed that terms such as *sözlük* (dictionary) and *lûgat* (dictionary, an old usage) were used in grammar studies on texts. In the Turkish Council of Higher Education Thesis Center, because these terms were used in some studies written in English, German, French, Persian and Spanish and conducted in the field of Western and Eastern languages, the number of search results were extremely high. Since only theses written in Turkish are in the scope of the current study, theses other than Turkish were excluded. Moreover, in the Turkish Council of Higher Education Thesis Center, only the abstracts of the studies conducted in the previous years, such as 1956, 1962, 1985, could be seen and the full texts could not be accessed. While selecting the master's theses and doctoral dissertations in the field of lexicography, the keywords in the research scope were queried. Of 302 graduate studies, 230 provided with full-text access were read and examined by the researcher individually, and 51 master's theses and doctoral dissertations to be included in the research were determined.

Table 3: Doctoral dissertations by year

Year	Frequency	%
1993	1	8.33
2000	1	8.33
2007	1	8.33
2009	2	16.67
2010	3	25.00
2011	2	16.67
2013	1	8.33
2016	1	8.33
Total	12	100.00

As can be seen in Table 3, the first doctoral dissertation that met the criteria for the database of the current study was conducted in 1993. The highest number of doctoral dissertations produced were in the year 2010 (3 doctoral dissertations).

Table 4: Universities and institutes producing doctoral dissertations

University	Institute	Department	Number	Percentage value of total doctoral dissertations
Ankara University	Institute of Social Sciences	Eastern Linguistics and Literature (Arabic Language and Literature)	1	8,33
	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
Atatürk University	Institute of Social Sciences	Department of Basic Islamic Studies	1	8,33
	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
Çukurova University	Institute of Educational Sciences	Department of German Language and Literature	1	8,33
	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
Dicle University	Institute of Educational Sciences	Department of Social Fields Education (Turkish Language Education)	1	8,33
Eskişehir Osmangazi University	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
Gazi University	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
İstanbul University	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
Marmara University	Institute of Social Sciences	Department of Turkish Language and Literature	1	8,33
Yeditepe University	Institute of Social Sciences	Department of Anthropology	1	8,33

Table 4 shows the institutes and departments where doctoral dissertations were conducted. According to this, the maximum number of doctoral dissertations in the field of lexicography were written in Ankara University (2), Atatürk University (2) and Çukurova University (2). When the characteristics and types of the institutes are considered, 10 of the 12 doctoral theses conducted in the field of lexicography were written in the Institute of Social Sci-

ences, and only two of them were written in the Institute of Educational Sciences. It was found that the maximum number of doctoral dissertations (total 7) were written in the Department of Turkish Language and Literature.

Table 5: Doctoral dissertation advisor

Name of Dissertation Advisor	Number of PhD Dissertations Supervised	Total Percentage Value of Doctoral Dissertations
İsmail Hakkı Aksoyak	2	16.6
Erdoğan Boz	1	8,33
Feza Tansuğ	1	8,33
Gülden Sağol Yüksekaya	1	8,33
Leyla Uzun & Rahmi Er	1	8,33
Mustafa Özkan	1	8,33
Sadettin Özçelik	1	8,33
Şükrü Haluk Akalın	1	8,33
Süleyman Tülücü	1	8,33
Tahir Balcı	1	8,33
Turgut Karabey	1	8,33
TOTAL	12	100.00

As can be seen in Table 5, 12 doctoral dissertations were written under the supervision of 11 different academics. İsmail Hakkı Aksoyak is the academic who supervised the most doctoral dissertations. When the number of supervisors of doctoral dissertations are considered, it can be seen that 11 doctoral dissertations were conducted with one supervisor and only one doctoral dissertation was conducted with two supervisors.

3.1.2 Master's theses

Table 6: Master's theses by year

Year	Frequency	%
1996	1	2.56
1997	1	2.56
2005	1	2.56
2006	2	5.13
2007	6	15.38
2008	1	2.56
2009	3	7.69
2010	4	10.26
2011	4	10.26
2012	5	12.82
2013	4	10.26
2014	2	5.13
2015	5	12.82
Total	39	100.00

Master's theses seem to have increased in recent years as shown in Table 6. The year in which master's theses were the most frequently conducted is 2007. Especially since 2006, it is seen that master's theses have been conducted more than once. Since 2009, there has been a steady increase in the number of master's theses.

Table 7: Universities and institutes producing master's theses

University	Institute	Department	Number of Master's Theses	Total Percentage Value of Master's Theses
Ankara University	Institute of Social Sciences	Department of Linguistics	3	7,69
		Eastern Linguistics and Literature (Arabic Language and Literature)	1	2,56
		Department of Turkish Language and Literature	1	2,56
Eskişehir Osmangazi University	Institute of Social Sciences	Department of Turkish Language and Literature	5	12,82
Marmara University	Institute of Educational Sciences	Department of Turkish Language Education	4	10,25
Sakarya University	Institute of Educational Sciences	Department of Turkish Language Education	2	5,12
	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Afyon Kocatepe University	Institute of Educational Sciences	Department of Turkish Language Education	1	2,56
	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Gazi University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
	Institute of Science and Technology	Department of Computer Engineering	1	2,56
Yıldız Technical University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
	Institute of Social Sciences	Department of Translation and Interpretation (Department of French Language)	1	2,56
İstanbul University	Institute of Social Sciences	Department of Turkish Studies	1	2,56
İstanbul Aydın University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Atatürk University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Dokuz Eylül University	Institute of Educational Sciences	Department of Teaching Turkish as a Foreign Language	1	2,56
Ege University	Institute of Social Sciences	Turkish World Studies (Turkish Languages and Dialects)	1	2,56
Fatih Sultan Mehmet Vakıf University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Fırat University	Institute of Educational Sciences	Department of Turkish Language Education	1	2,56
Kahramanmaraş Sütçü İmam University	Institute of Social Sciences	Eastern Linguistics and Literature	1	2,56
Mersin University	Institute of Social Sciences	Department of German Language and Literature	1	2,56
Nevşehir University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Ondokuzmayıs University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Pamukkale University	Institute of Educational Sciences	Department of Turkish Language Education	1	2,56
Selçuk University	Institute of Educational Sciences	Department of Turkish Language Education	1	2,56
Trakya University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56
Yüzüncü Yıl University	Institute of Social Sciences	Department of Basic Islamic Studies	1	2,56
Anadolu University	Institute of Social Sciences	Department of Turkish Language and Literature	1	2,56

Table 7 shows at which universities and institutes master's theses were written. According to this, the universities with the highest number of master's theses are as follows: Ankara University 5, Eskişehir Osmangazi University 5, and Marmara University 4. When the type of institutes were considered, the highest number of master's theses (27 theses) were found at the Institute of Social Sciences. 11 master's theses were written at the Institute of Educational Sciences, and only 1 master's thesis was written at the Institute of Science and Technology. With regard to the characteristics of the institute's master's theses that were conducted, it is remarkable that the number of theses in the Institute of Science and Technology are scarce.

When the departments where master's theses were written, are considered, the maximum number of master's theses (16 theses) were written in the Department of Turkish Language and Literature. 10 master's theses were written in the Department of Turkish Language Education and 3 master's theses were written in the Department of Linguistics. The Department of Turkish Language and Literature is clearly in the forefront in the number of master's theses in the field of lexicography. Therefore, master's theses in the field of Turkish lexicography are mostly conducted in Turkish language and literature.

Table 8: Master's theses supervisors

Name of the Master's Theses Supervisor	Number of Theses Supervised	Total Percentage Value of Master's Theses
Erdoğan Boz	5	12.82
Engin Yılmaz	2	5.12
Mustafa Sinan Kaçalın	2	5.12
Ahmet Akçataş	1	2.56
Ahmet Günşen	1	2.56
Ahmet Turan Sinan	1	2.56
Asuman Akay Ahmed	1	2.56
Avni Gözütok	1	2.56
Aybars Erözden	1	2.56
Cem Bozşahin	1	2.56
Dilek Ataizi	1	2.56
Ertuğrul Efeoğlu	1	2.56
Faruk Çiftçi	1	2.56
Fikret Turan	1	2.56
Günay Karaağaç	1	2.56
Hacer Karacan	1	2.56
Halil İbrahim Usta	1	2.56
Halit Biltekin	1	2.56
Hülya Kasapoğlu Çengel	1	2.56
İbrahim Ethem Özkan	1	2.56

Latif Beyreli	1	2.56
M. Faruk Toprak	1	2.56
M. Mehdi Ergüzel	1	2.56
Mehmet Gürlek	1	2.56
Nadir Engin Uzun	1	2.56
Nurettin Öztürk	1	2.56
Salih Demirbilek	1	2.56
Sıla Ay	1	2.56
Veli Doğan Günay	1	2.56
Vural Ülkü	1	2.56
Yakup Civelek	1	2.56
Zehra Göre	1	2.56
Zeki Kaymaz	1	2.56
Total	39	100.00

33 different academics supervised 39 master's theses. Academics supervising the most master's theses are as follows: Erdoğan Boz 5, Engin Yılmaz 2, Mustafa Sinan Kaçalin 2. As regards the study area of the supervisors, it was found that the supervisors are mostly researchers in the field of Turkish language and literature. In addition, it can be seen that all of the master's theses are conducted with only one supervisor. Although the higher education system in Turkey allows more than one master's thesis supervisor, more than one supervisor was not preferred in conducting master's theses.

3.1.3 Published oral presentations

Table 9: Published oral presentations by year

Year	Frequency	%	Year	Frequency	%
1969	1	0.33	2007	37	12.29
1972	2	0.66	2008	60	19.93
1985	2	0.66	2009	36	11.96
1999	9	2.99	2010	23	7.64
2000	4	1.33	2011	7	2.33
2002	1	0.33	2012	9	2.99
2003	1	0.33	2013	25	8.31
2004	8	2.66	2015	1	0.33
2005	1	0.33	2016	72	23.92
2006	2	0.66	Total	301	100.00

As can be seen from the total number of published oral presentations these are one of the most produced text types compared to other text types. Table 9 shows that the oldest oral presentation in the database was presented in 1969.

Although many scientific activities such as symposiums, conferences and congresses were held since the establishment of the Turkish Language Institute in 1932, Turkish researchers did not choose to present their studies on lexicography in the form of oral presentations. With the beginning of the 2000s, it is observed that there is an increase in the preference for oral presentations in texts related to lexicography. Although the number of presentations have increased in the last decade, in some years the number of presentations have been quite low compared to other years. One of the most important essential factors for the increasing number of presentations is lexicographic symposiums organised in Turkey in recent years.

Table 10: Frequency and percentages of published oral presentations with regard to the number of authors

Number of authors	Frequency	%
Single author	273	90.69
Two authors	23	7.6
Three authors	3	0.99
Four authors	1	0.33
Total	301	100.00

Of the published oral presentations, 273 (90.69%) are by one author, 23 (7.6%) are by two authors, 3 (0.99%) are by three authors and 1 (0.33%) is by four authors. As shown by these data, oral presentations prepared by single authors are common. The number of oral presentations with two authors are low. It can be claimed that teamwork is not preferred in these studies.

Table 11: Congresses and symposiums producing the most oral presentations

#	Name of the congress/symposium	Number of Presentations
1	II. Uluslararası Sözlükbilimi Sempozyumu (2nd International Lexicography Symposium)	39
2	Uluslararası Türkiyat Araştırmaları Bilgi Şöleni Bildirileri (International Symposium on Turkish Studies)	36
3	Uluslararası Kaşgarlı Mahmud Sempozyumu (International Kashgar Mahmoud Symposium)	34
4	I. Uluslararası Sözlükbilimi Sempozyumu (1st International Lexicography Symposium)	29
5	Kaşgarlı Mahmut ve Türk Dünyasının Dili, Edebiyatı, Kültürü ve Tarihi (Kashgarlı Mahmut and Language, Literature, Culture and History of the Turkish World)	22
6	Türkiye’de ve Dünya’da Sözlük Yazımı ve Araştırmaları Uluslararası Sempozyumu (International Symposium of Dictionary Compiling and Research in Turkey and the World)	33
7	IV. Uluslararası Türk Dili Kurultayı Bildirileri (4th International Turkish Language Congress)	21

8	Uluslararası Sözlükbilimi Sempozyumu (International Lexicography Symposium)	8
9	VI. Uluslararası Dünya Dili Türkçe Sempozyumu (6th International World Language Turkish)	8
10	IV. Uluslararası Türk Dili Kurultayı (4th International Turkish Language Congress)	7
11	VI. Uluslararası Türk Dili Kurultayı Bildirileri (6th International Turkish Language Congress)	7
12	III. Uluslararası Dünya Dili Türkçe Sempozyumu (3rd International World Language Turkish)	6
13	IV. Uluslararası Dünya Dili Türkçe Sempozyumu (4th International World Language Turkish)	10
14	V. Uluslararası Türk Dili Kurultayı Bildirileri (5th International Turkish Language Congress)	6
15	III. Uluslararası Büyük Türk Dil Kurultayı (3rd International Turkish Language Congress)	4

Papers on lexicography were presented in 301 different categories. The reason why the categories were so diverse was that papers on lexicography were presented in symposiums related to grammar and linguistics in the years before 2014. Before the 2000s, symposiums exclusively on lexicography were almost none. The First International Lexicography Symposium was held by Sakarya University on 26–27 November, 2014. The Second was held by İstanbul University on 3–4 November, 2015, and the Third was held on 3–4 November 2016 by Eskişehir Osmangazi University. Since then, it has been decided to organise the Lexicography Symposium every two years. These symposiums have increased the number of oral presentations on lexicography in Turkey.

3.1.4 News

News is a type of text including neologism examples, suggestions for foreign words and/or promotions of new publications about lexicography/dictionaries. News is published in the *Journal of the Turkish Language Association*.

Table 12: News by year

Year	Frequency	%	Year	Frequency	%
1952	1	4.76	1999	2	9.52
1953	1	4.76	2004	1	4.76
1959	2	9.52	2007	2	9.52
1960	1	4.76	2009	1	4.76
1962	1	4.76	2011	2	9.52
1971	1	4.76	2012	2	9.52
1987	1	4.76	2013	1	4.76
1989	1	4.76	2014	1	4.76
			Total	21	100.00

The number of news items related to lexicography are produced irregularly by year. News text type is published annually. However, it is insignificant numerically for this study.

3.1.5 Books

Table 13: Books by year

Year	Frequency	%
2007	1	33.3
2011	1	33.3
2016	1	33.3
Total	3	100.00

In the field of Turkish lexicography, the number of books are insufficient. There is still no lexicographic handbook on Turkish lexicography, such as those by Atkins and Rundell 2008, Durkin 2016, Jackson 2013, Svensén 2009, or Zgusta 1971.

3.1.6 Articles

Table 14: Articles by year

Year	Frequency	%	Year	Frequency	%	Year	Frequency	%
1934	1	0.21	1974	1	0.21	2001	4	0.84
1939	1	0.21	1977	1	0.21	2002	11	2.32
1942	1	0.21	1981	1	0.21	2003	6	1.26
1952	1	0.21	1983	1	0.21	2004	11	2.32
1953	2	0.42	1984	1	0.21	2005	5	1.05
1954	1	0.21	1986	2	0.42	2006	14	2.95
1956	1	0.21	1987	1	0.21	2007	14	2.95
1957	1	0.21	1990	1	0.21	2008	18	3.79
1959	1	0.21	1991	2	0.42	2009	64	13.47
1960	1	0.21	1994	3	0.63	2010	22	4.63
1961	1	0.21	1995	3	0.63	2011	46	9.68
1965	1	0.21	1996	3	0.63	2012	24	5.05
1970	1	0.21	1997	4	0.84	2013	36	7.58
1971	2	0.42	1998	15	3.16	2014	39	8.21
1972	9	1.89	1999	16	3.37	2015	48	10.11
1973	1	0.21	2000	9	1.89	2016	23	4.84
Total						475	100.00	

As can be seen in Table 14, the most produced text type in the field of Turkish lexicography is articles. The ratio of the total text type of the articles among the types of texts produced in this area is 47.4%. 475 articles were produced in a total of 48 years. In the field of Turkish lexicography, an average of 9.8 articles is produced annually. The number of articles have shown an upward trend in the last 10 years. The year the most articles (64 articles) were published is 2009, followed by 2011 (46 articles). When the years are taken into consideration, there is a marked increase towards the present.

Table 15: Frequency and percentage of the number of authors

Number of authors	Frequency	%
Single author	437	92
Two authors	36	7.57
Three authors	2	0.42
Total	475	100.00

Table 15 provides information on the number of authors. Of the articles in the database, 437 (92%) are by single authors, 36 (7.57%) are by two authors, and 2 (0.42%) are by three authors. Contrary to published oral presentations, there are no articles by 4 authors in Turkish lexicography. The number of articles by multiple authors are close to the number of oral presentations. As shown by these data, the studies with multiple authors in the field of Turkish lexicography are mainly composed by two authors. It can be said that articles prepared as teamwork are not much preferred.

In Turkey journal publishing only in the field of lexicography has not been realized yet. Articles related to lexicography are published in academic journals on linguistics and grammar research.

Table 16: Academic journals publishing the most articles in the field of lexicography and the number of articles published

Name of the journal	Published Number of Articles	%
Turkish Studies	124	26.1
Türk Dili Dil ve Edebiyat Dergisi	73	15.3
Kebikeç Dergisi	28	5.8
Littera Turca	28	5.8
Dil ve Edebiyat Araştırmaları Dergisi	24	5.1
Ankara Üniversitesi Dil ve Tarih Coğrafya Fakültesi Dergisi	15	3.1
International Journal of Language Academy	14	2.9

Selçuk Üniversitesi Türkiyat Araştırmaları Dergisi	12	2.5
Uluslararası TEKE Dergisi	11	2.3
Ankara Üniversitesi Türkiyat Araştırmaları Enstitüsü Dergisi	9	1.8
Belleten	8	1.6
Türk Dünyası Dil ve Edebiyat Dergisi	8	1.6
Türkbilig Türkoloji Araştırmaları Dergisi	8	1.6
Diyalektolog	7	1.4
Ege Üniversitesi Türk Dünyası İncelemeleri Dergisi	6	1.2
Millî Folklor Dergisi	6	1.2
Türklük Bilimi Araştırmaları Dergisi	6	1.2
Türük Dergisi	6	1.2
İstanbul Üniversitesi Türk Dili ve Edebiyatı Dergisi	5	1
Bilig Türk Dünyası Sosyal Bilimler Dergisi	4	0.8

Table 16 shows the first 20 journals in which articles related to Turkish lexicography are published, and the number of articles related to lexicography published in these journals, and their ratio to the total number of journals.

The total number of journals in the database of the study which published articles in the field of lexicography are 76. Of the 475 articles in the database, 402 were published by the first 20 journals shown in the table. The journals with the highest number of articles on Turkish lexicography are as follows: 1. *Turkish Studies*: 124 articles, 2. *Türk Dili Dil ve Edebiyat Dergisi* (Turkish Language and Literature Journal): 73 articles, 3. *Kebikeç Dergisi* (Kebikeç Journal): 28 articles.

Out of 31 journals, especially those issued by social science institutes in Turkey (e.g. *Kahramanmaraş Sütçü İmam University Journal of Social Sciences*, *Eskişehir Osmangazi University Journal of Social Sciences*, *Istanbul University Journal of Social Sciences*), only one journal published an article on lexicography.

3.1.7 Reviews

Newly published dictionaries are generally introduced in texts in the review category.

Table 17: Reviews by year

Year	Frequency	%	Year	Frequency	%	Year	Frequency	%
1935	2	1.33	1975	13	8.67	2002	6	4.00
1936	3	2.00	1976	5	3.33	2003	1	0.67
1939	1	0.67	1977	5	3.33	2004	2	1.33

1954	5	3.33	1979	1	0.67	2005	5	3.33
1956	1	0.67	1980	2	1.33	2006	3	2.00
1960	1	0.67	1981	1	0.67	2007	2	1.33
1962	1	0.67	1982	3	2.00	2008	5	3.33
1963	1	0.67	1991	2	1.33	2009	5	3.33
1965	1	0.67	1992	1	0.67	2010	7	4.67
1967	2	1.33	1994	2	1.33	2011	6	4.00
1968	1	0.67	1995	5	3.33	2012	6	4.00
1969	1	0.67	1996	5	3.33	2013	3	2.00
1971	1	0.67	1999	4	2.67	2014	10	6.67
1972	6	4.00	2000	2	1.33	2015	7	4.67
1973	1	0.67	2001	1	0.67	2016	1	0.67
1974	1	0.67	Total				150	100.00

Ordered by year, the number of reviews are close to each other. Although the number of reviews have increased for some years, they continue with an average number of 3.2%. When the number of dictionaries published between the years 1935 and 2016 in Turkey are considered, it is remarkable that there are so few reviews.

3.2 Researchers and citations

In this section, the quantitative status of the authors contributing to lexicography and the citations of these authors' works in the database of the current study are discussed.

3.2.1 The contributors, and their contributions

The programs in which the doctoral dissertations and master's theses are written have been mentioned above. Graduate theses are mostly conducted in the Turkish language and literature departments. Among the lexicographers who produced the most texts, we see the names of those who are supervisors of master's theses and doctoral dissertations.

Table 18: Most productive researchers

#	Name of researcher	Master's theses	Doctoral dissertations	Published oral presentations	News	Books	Articles	Reviews	Total
1.	Erdoğan Boz			10		1	13		24
2.	Turkish Language Institute				6			14	20
3.	Tuncer Gülensoy			4			3	5	12

4.	Adem Aydemir			9		9
5.	Mehmet Ölmez			2	6	8
6.	İsmail Parlatur	3		4	1	8
7.	Ali Püsküllüoğlu				8	8
8.	Galip Güner			7		7
9.	Nail Tan		6			6
10.	Fatih Doğru	1	2	3		6
11.	Paşa Yavuzarslan		1	4		5
12.	Nuh Doğan		3	2		5
13.	Sami N. Özerdim				5	5
14.	Zuhal Kargı Ölmez		1	4		5
15.	Bülent Özkan		1	4		5
16.	Akartürk Karahan		1	2	2	5
17.	Aysu Ata		2	3		5
18.	Atabey Kılıç			5		5
19.	Hasan Eren		1	2	2	5
20.	Ferdi Bozkurt	1	2	2		5

The twenty researchers who are the most productive in Turkish lexicography are shown in Table 18. In the database of the current study, there are 678 researchers who produced texts related to Turkish lexicography. The ratio of the total number of texts to the number of researchers is 1.4%. The three most productive researchers in Turkish lexicography are Erdoğan Boz, the Turkish Language Association (studies related to this lexicographic association are sometimes published under the corporate name) and Tuncer Gülensoy. The first 20 researchers produced the most texts: 15.7% of the total texts. The authors who published the most articles are Erdoğan Boz (13 articles), Adem Aydemir (7 articles) and Galip Güner (7 articles). The authors who published the most oral presentations are Erdoğan Boz (10 papers), Tuncer Gülensoy (3 papers), Nuh Doğan and İsmail Parlatur (3 papers each).

When the different types of writing characteristics of researchers related to lexicography are considered, another remarkable fact emerges. There is no researcher who has produced texts in all text types: master theses, doctoral dissertations, published oral presentations, news, books, articles and reviews. Erdoğan Boz, Ferdi Bozkurt and Fatih Doğru were each responsible for three different text types.

3.2.2 Most highly cited publications

As mentioned above, the total number of journals that are in the Turkish Lexicography Corpus database and that publish articles in the field of lexicology are 76. However, among these journals, only the *Bilig Journal of Turkish World Social Sciences* and *National Folklore* are indexed in the Science Citation Index. Only 4 articles in the database created for Turkish lexicography were published in this journal. The remaining journals are indexed in international databases/

resources such as EBSCO, MLA and INDEX COPERNICUS, and in national databases/resources such as ULAKBIM and ARASTIRMAX, while some journals are not indexed in any databases/resources at all. In this study, citation data are obtained from Google Scholar.

3.2.2.1 Doctoral dissertations

Table 19: The most cited doctoral dissertations

#	Author	Year	Text	Total citations	Percentage of total citations
1.	Ahmet Hilmi İmamoğlu	1993	Farsça-Türkçe manzum sözlükler ve Şahidi'nin Sözlüğü (inceleme-metin)	6	37.50
2.	Ahmet Dönger	2009	Sözlük Kullanma Eğitiminin Yabancı Dil Olarak Almanca Öğrenimine Etkisi	3	18.75
3.	Özer Şenödeyici	2011	Naili Divanı Sözlüğü (Bağlamli Dizin ve İşlevsel Sözlük)	3	12.50
4.	Furkan Öztürk	2007	Bakı Divanı Sözlüğü (Bağlamli Dizin ve İşlevsel Sözlük)	2	12.50
5.	Resul Özavşar	2013	Tarama Sözlüğü ve Türkçe Sözlük'e Göre Anlam Değişmeleri	1	6.25
6.	Dilek Herkmen	2009	Divânü Lugatî't-Türk'te Fil Yapımı	1	6.25
Total				16	100.00

Only 6 of the 12 doctoral dissertations have been cited, 6 doctoral dissertations having received a total of 16 citations. The most cited dissertation is *Farsça-Türkçe Manzum Sözlükler ve Şahidi'nin Sözlüğü* (inceleme-metin) (Persian-Turkish Poetry Dictionaries and the Dictionary of Shahidi) (examination-text)).

3.2.2.2 Master's theses

Table 20: The most cited master's theses

#	Author	Year	Text	Total citations	Percentage of total citations
1.	Emrah Özcan	2006	Başlangıç Düzeyi Yabancı Dil Olarak Türkçe Öğretimi için Sözlükçe Çalışması	8	12.6
2.	Burak İbrahim Sevindi	2013	Türkçe Metinlerde Denetimli ve Sözlük Tabanlı Duygu Analizi Yaklaşımlarının Karşılaştırılması	7	11.1
3.	Bilge Gökter	2010	Sözlükbilim Temelinde Türkiye'de Ağız Sözlükçülüğü (İlke ve Yöntemler)	5	7.9
4.	Ziya Doğan Koreli	2007	Eylem ve Ad Olarak Kullanılan ve Türkçe Sözlükte Bulunmayan Öbekler ve Bunların Türkçenin Yabancı Dil Olarak Öğretimindeki Yeri	5	7.9

5.	Ceren Oğuz	2011	Kâmus-ı Türkî'den Türkçe Sözlük'e Anlam Değişmeleri: Adlar (K-Z)	5	7.9
6.	Fatih Doğru	2012	Kâmus-ı Türkî'den Türkçe Sözlük'e Anlam Değişmeleri: Eylemler	5	7.9
7.	İlke Küçük	2012	Kâmus-ı Türkî'den Türkçe Sözlüğe Anlam Değişmeleri- Adlar (A-K)	5	7.9
8.	Ceren Berber	2012	Türkçenin Yabancı Dil Olarak Öğretiminde Yardımcı Kaynak Hazırlama: Sözlük Modeli Önçalışması	3	4.7
9.	Musa Salan	2010	Et-Tuhfetü'z -Zekiyye Fı'l-Lugati't-Türkiyye'de Fiil	3	4.7
10.	Necati Kaya	2007	Okul Sözlüklerinin Değerlendirilmesi (1945-2005)	3	4.7
11.	Hatice Korkmaz	2007	Divanu Lugati't Türk'teki Atasözlerinin Anlambilimsel Açından İncelenmesi	3	4.7
12.	Mehmet Nuri Alpak	2006	Arap Dilinde Sözlük Çalışmaları ve Nazım Efendi'nin Tercümanı'l-Lügat Adlı Eserinin İncelenmesi	3	4.7
13.	Ali Çiçek	1997	1928 tarihine kadar yazılmış Fransızca-Türkçe, Türkçe-Fransızca ve çok dilli sözlükler üzerine bir dil araştırması	2	3.1
14.	Selim Tiryakiol	2013	Dil Öğretimi Geleneğimizde Manzum Sözlükler (Tuhfe-i Asım Örneği)	2	3.1
15.	Serper Acar	2009	Necati Bey Divanı Sözlüğü (Bağlamlı Dizin Ve İşlevsel Sözlük)	2	3.1
16.	Efsun Bilgin	2015	Türkçe Sözlük'te Özel Ad Kaynaklı Sözler	1	1.5
17.	Pervin Sayan	2005	11-14 Yaş İçin Hazırlanan Okul Sözlüklerinde Tanımlar Üzerine Bir Araştırma	1	1.5

22 of the 39 master's theses have never received any citations, while 17 master's theses have received a total of 63 citations. The most cited master's thesis is *Başlangıç Düzeyi Yabancı Dil Olarak Türkçe Öğretimi için Sözlükçe Çalışması* (A Glossary Study for Teaching Turkish as a Foreign Language) written by Emrah Özcan.

3.2.2.3 Articles

Table 21: The 30 most cited articles

#	Author	Year	Text	Total citations	Ratio of citations to total citations
1.	Ali Göçer	2009	Türkçe Eğitiminde Öğrencilerin Söz Varlığını Geliştirme Etkinlikleri ve Sözlük Kullanımı	51	3.62%
2.	Ahmet Kocaman	1998	Dilbilim, Sözlük, Sözlükçülük	41	2.91%
3.	Atabey Kılıç	2007	Türkçe-Farsça Manzum Sözlüklerden Tuhfe-i Vehbi (Metin)	33	2.34%

4.	Nurettin Demir	1999	Ağız Sözlükçülüğü	24	1.70%
5.	Deniz Melanlioğlu	2013	Ortaokul Öğrencilerinin Sözlük Kullanma Alışkanlıkları: Nitel Bir Araştırma	22	1.56%
6.	Paşa Yavuzarslan	2004	Türk Sözlükçülük Geleneği Açısından Osmanlı Dönemi Sözlükleri Ve Şemseddin Sâmî'nin Kâmûs-ı Türkî'si	22	1.56%
7.	Gülsel Sev	2004	Divanü Lugatit Türk'te İnkilemeler	21	1.49%
8.	Havva Yaman	2010	İlköğretim İkinci Kademe Öğrencilerinin Sözlük Kullanma Alışkanlıkları Üzerine Bir Değerlendirme	21	1.49%
9.	Doğan Aksan	1998	Türklerde Sözlükçülük, Bugün Türkiye'de Sözlük	19	1.35%
10.	Nadir İlhan	2009	Sözlük Hazırlama İlkeleri, Çeşitleri ve Özellikleri	19	1.35%
11.	Ali Göçer	2010	Türk Dili ile İlgili Sözlüklere Genel Bir Bakış Günümüz İlköğretim Sözlükleri	18	1.28%
12.	Şükrü Haluk Akalın	2010	Sözcük Bilimi ve Sözlükçülük	18	1.28%
13.	Zeynep Korkmaz	2000	Türkiye'de Ağız Sözlükleri	18	1.28%
14.	Zeynep Korkmaz	2000	Türkiye'de Ağız Sözlükleri	18	1.28%
15.	Özen Yaylagül	2010	Divanü Lugatit-Türk'te Yer Alan Atasözlerindeki Metaforlar	17	1.20%
16.	Zekeriya Bingöl	2006	Sözlük ve Sözlükçülük Üzerine Bir Araştırma	17	1.20%
17.	Kerime Üstünova	1998	Dede Korkut Destanlarında İki Sözcüklü Yüklem	16	1.13%
18.	Bayram Çetinkaya	2009	Eşdizimli Sözlükler	13	0.92%
19.	Harun Tolasa	1986	18. yy'da Yazılmış Bir Divan Edebiyatı Terimleri Sözlüğü Müstakimzade'nin İstilahatü's-Şi'riye'si	13	0.92%
20.	Süer Eker	2009	Divanü Lugatit-Türk ve İran Dillerinden Kopyalar Üzerine I	13	0.92%

Table 21 shows the 20 most cited articles in Turkish lexicography. All 475 articles have received a total of 1406 citations. 238 articles haven't been cited, while 237 articles have been cited. 46 articles have received only 1 citation. When the total number of citations to articles are divided by the number of cited articles, the average number of citations are 5.06. When the total number of citations are divided by the number of all articles in the database, the average number of articles are 2.70. The most cited article is "Türkçe Eğitiminde Öğrencilerin Söz Varlığını Geliştirme Etkinlikleri ve Sözlük Kullanımı" ("Activities for Developing

the Vocabulary and Dictionary Usage of Students in Turkish Education") by Ali Göçer.

Table 22: Dispersion of the 30 most cited articles by the year produced

Year of publication of articles	Number of articles	Number of citations	Year ratio by number of articles
2009	7	132	18,8
1998	4	88	22
2010	5	86	17,2
2000	3	48	16
2004	2	43	21,5
2013	2	34	17
2007	1	33	33
2002	2	25	12,5
1999	1	24	24
2006	1	17	17
1986	1	13	13
1970	1	12	12
2003	1	12	12
2005	1	12	12
2012	1	12	12

Table 22 shows the number of the most cited articles and their publication years. When the publication years of the articles are considered, the articles published in 2009 were cited most, followed by those of the years 1998 and 2010.

3.2.2.4 Academic journals

Table 23: The number of citations of the journals

#	Name of journal	Number of citations	Ratio of citations to total citations
1.	Turkish Studies Language/Literature	356	25.32%
2.	Türk Dili Dil ve Edebiyat Dergisi	196	13.94%
3.	Kebikeç İnsan Bilimleri için Kaynak Araştırmaları Dergisi	164	11.66%
4.	Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Türkoloji Dergisi	83	5.90%
5.	Türklük Bilimi Araştırmalar Dergisi-TÜBAR	54	3.84%
6.	Selçuk Üniversitesi Türkiyat Araştırmaları Dergisi	41	2.91%

7.	Milli Folklor Dergisi	29	2.06%
8.	Uluslararası Türkçe Edebiyat Kültür Eğitim	28	1.99%
9.	A.Ü. Türkiyat Araştırmaları Enstitüsü Dergisi	28	1.99%
10.	Diyalektolog	27	1.92%
11.	Dil Araştırmaları	26	1.84%
12.	Central Asian Studies	25	1.77%
13.	Dil ve Edebiyat Araştırmaları Dergisi	25	1.77%
14.	Bilig Türk Dünyası Sosyal Bilimler Dergisi	22	1.56%
15.	Modern Türklük Araştırmaları Dergisi	19	1.35%
16.	International Journal of Language Academy	19	1.35%
17.	Zeitschrift für die Welt der Türken	17	1.20%
18.	İstanbul Üniversitesi Türk Dili ve Edebiyatı Dergisi	16	1.13%
19.	Türkbilig Türkoloji Araştırmaları Dergisi	13	0.92%
20.	Fırat Üniversitesi Sosyal Bilimler Dergisi	13	0.92%

The table above shows the most frequently cited journals dealing with Turkish lexicography. 475 articles in the database have been published in 76 different journals. The articles in the journals have received a total of 1406 citations. Of the 474 articles in the study, 238 haven't received any citations, while 236 have been cited. Accordingly, in Turkish lexicography the most effective academic journals in terms of citation are *Turkish Studies*, *Türk Dili Dil ve Edebiyat Dergisi* (Turkish Language and Literature Journal), and *Kebikeç İnsan Bilimleri için Kaynak Araştırmaları Dergisi* (Kebikeç Journal of Human Resources Research).

4. Conclusion

For this bibliometric study, a database of 1001 texts produced between 1932 and 2016 in the field of Turkish lexicography has been created. Important information about the characteristics of the texts in the database has been put forward. The ratio of texts produced in these fields by type, their distribution by year, the distribution of doctoral dissertations and master's theses by year and institute, and the distribution of researchers supervising these studies have been indicated. In addition, the researchers producing the most texts in the field of lexicography, the degree of collaboration among the authors, institutions producing the most texts in the field of lexicography, the number of personal citations, diversity of authors' text production, and the most cited texts have been specified. Information such as which academic journals are more effective and which scientific activities include the most texts about lexicography have been indicated.

One of the general and most important results is that there has been an increase in nearly all text types in recent years. Researchers have given more space to studies on lexicography. One of the biggest problems regarding Turkish lexicography is the lack of books on theoretical lexicography.

Another important deficiency is the lack of collaborative texts produced by researchers in this field. Projects undertaken by multiple authors as team-work may increase the number of lexicographic texts.

A general remark can be the scarcity of doctoral dissertations, the writing of which is a longer and more extensive process than that required for the master's degree, and whose academic quality is important. That the number of symposiums that are the meeting place of researchers giving oral presentations have increased is a positive development. The symposium environments bring together researchers who work in the field of lexicography and make positive contributions to it. In addition, the published conference proceedings resulting from these symposiums bring together the texts related to the field, increase the accessibility of the texts and awareness of the researchers.

In Turkey, there is no academic journal publishing articles that discuss issues related only to lexicography. When the number of articles are considered, it can be concluded that there are enough articles for publishing an academic journal in the Turkish language exclusively on lexicography. The existence of a journal that publishes only on lexicography will provide a common platform for researchers and will facilitate access to related articles.

Acknowledgements

The texts used in the current study were created by utilizing the list of open-access texts given on the website of the project funded by the Scientific Research Projects Coordination Unit of Eskişehir Osmangazi University (<http://tsd.ogu.edu.tr/index.php?s=yayinlar>). The author of this study also belongs to the researchers of the project (Project number 2016-019056).

Endnotes

1. <https://ulakbim.tubitak.gov.tr/> accessed February 2, 2019.
2. <https://scholar.google.com.tr> accessed February 2, 2019.
3. <http://web.a.ebscohost.com/ehost/search/basic?vid=0&sid=e97c426e-c7c7-4ee5-a9c1-bd6f8d1b1129%40sessionmgr4008> accessed March 2, 2019.
4. <http://www.toplukatalog.gov.tr/> accessed May 2, 2019.
5. <https://kasif.mkutup.gov.tr/> accessed February 22, 2019.
6. <https://tez.yok.gov.tr/UlusalTezMerkezi/> accessed February 4, 2019.
7. <https://www.proquest.com/products-services/pqdtglobal.html> accessed January 7, 2019.

References

- Arik, E. 2015. A Bibliometric Analysis of Linguistics in Web of Science. *Journal of Scientometric Research* 4(1): 20-28.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*: Oxford/New York: Oxford University Press.

- Bergenholtz, H. and R.H. Gouws.** 2012. What is Lexicography? *Lexikos* 22: 31-42.
- Day, R.A. and B. Gastel.** 1995. How to Write and Publish a Scientific Paper. Fourth edition. Cambridge: Cambridge University Press.
- De Schryver, G.-M.** 2009a. Bibliometrics in Lexicography. *International Journal of Lexicography* 22(4): 423-465.
- De Schryver, G.-M.** 2009b. *Lexikos* at Eighteen: An Analysis. *Lexikos* 19: 372-403.
- Dilaçar, A.** 1953. Türkiye Türkçesi Sözlüklerinden Seçme Eserler. *Türk Dili* 22: 677-680.
- Durkin, P. (Ed.).** 2016. *The Oxford Handbook of Lexicography*. New York: Oxford University Press.
- Eminoğlu, E.** 2010. *Türk dilinin sözlükleri ve sözlükçülük kaynakçası. Vol. 1*. Sivas: Asitan Yayıncılık.
- Ercilasun, A.B. and Z. Akkoyunlu.** 2014. *Kâşgarlı Mahmud Dîvânü Lugâti't-Türk Giriş-Metin-Çeviri-Notlar-Dizin*. Ankara: Türk Dil Kurumu Yayınları.
- Garfield, E.** 1955. Citation Indexes for Science. *Science* 122(3159): 108-111.
- Gökçe, A.** 1998. *Türkiye Türkçesinin tarihî sözlükleri. Volume 6*. Ankara: Kebikeç Yayınları.
- Hulme, E.W.** 1923. *Statistical Bibliography in Relation to the Growth of Modern Civilization*. Two Lectures Delivered at the University of Cambridge in May, 1922.
- Jackson, H.** 2013. *Lexicography: An Introduction*. London: Routledge.
- Kotan, H.** 2017. Erzurum Yazma Eserler Kütüphanesinde Bulunan El Yazması Sözlükler. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 21(4): 1491-1509.
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography*. New York: The Scribner Press.
- Lei, L. and S. Liao.** 2017. Publications in Linguistics Journals from Mainland China, Hong Kong, Taiwan, and Macau (2003–2012): A Bibliometric Analysis. *Journal of Quantitative Linguistics* 24(1): 54-64.
- Lei, L. and D. Liu.** 2018. Research Trends in Applied Linguistics from 2005 to 2016: A Bibliometric Analysis and its Implications. *Applied Linguistics* 40(3): 540-561.
- Mohsen, M.A., H.-Z. Fu and Y.-S. Ho.** 2017. A Bibliometric Analysis of Linguistics Publications in the Web of Science. *Journal of Scientometric Research* 6(2): 109-118.
- Parlatır, İ.** 1995. Türkçe Sözlük Çalışmaları ve Sorunlarımız. *Türk Dili Dil ve Edebiyat Dergisi* 517: 3-19.
- Potter, W.G.** 1981. Introduction. *Library Trends* 30(1): 5-7.
- Pritchard, A.** 1969. Statistical Bibliography or Bibliometrics. *Journal of Documentation* 25(4): 348-349.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Tarp, S.** 2012. Do We Need a (New) Theory of Lexicography? *Lexikos* 22: 321-332.
- Yavuzarslan, P.** 2017. Türk Sözlükçülük Geleneği Açısından Osmanlı Dönemi Sözlükleri ve Şemseddin Sâmî'nin Kâmûs-ı Türkî'si. *DTCF Dergisi* 44(2): 185-202.
- Yıkılmış, S. and Ş. Sazak.** 2017. Türk sözlükbilimi bibliyografyası üzerine bir deneme (2000–2016). Boz, E. et al. 2017. *III. Uluslararası Sözlükbilimi Sempozyumu, Eskişehir, 3–4 November 2016*: 780-825. Eskişehir: Eskişehir Osmangazi University.
- Zgusta, L.** 1971. *Manual of Lexicography*. Janua Linguarum. Series Maior 39. Prague: Academia/The Hague/Paris: Mouton.

La Création du Néologisme Français dans le Dictionnaire Général Bilingue Arabe–Français Moderne

Ibada Hilal, *Département de Langue et Littérature Françaises,
Université de Jordanie, Amman, Jordanie (ibada01@yahoo.fr)*

Résumé: Le présent article a pour objet d'étudier pourquoi et comment le dictionnaire général bilingue arabe–français moderne pourrait provoquer des néologismes français comme correspondants à des entrées et sous-entrées arabes étiquetées d'un domaine scientifique ou technique. Pour ce faire, la présente étude se propose de travailler sur la lettre [th] du dictionnaire général bilingue arabe–français *Abdelnour moderne* daté de l'année 2002. Ce dictionnaire est considéré comme l'un des dictionnaires généraux bilingues arabe–français les plus utilisés dans le monde arabe.

Mots-clés: NÉOLOGISME FRANÇAIS, DICTIONNAIRE GÉNÉRAL BILINGUE ARABE–FRANÇAIS MODERNE, UNITÉ LEXICALE, PALÉOLOGISME, CORPUS, ÉCART CULTUREL

Abstract: Creation of French Neologism in the Modern Arabic–French General Bilingual Dictionary. The purpose of this article is to investigate why and how the modern Arabic–French general bilingual dictionary could induce French neologisms corresponding to labeled Arabic entries and subentries of a scientific or technical domain. To do this, the present study proposes to work on the letter [th] of the Arabic–French general bilingual dictionary *Abdelnour Modern* (2002). This dictionary is considered as one of the most used Arabic–French general bilingual dictionaries in the Arab World.

Keywords: FRENCH NEOLOGISM, MODERN ARABIC–FRENCH GENERAL BILINGUAL DICTIONARY, LEXICAL UNIT, PALEOLOGISM, CORPUS, CULTURE GAP

1. Introduction

Le néologisme est le produit linguistique qui résulte du processus de la néologie. Cette dernière est, par ailleurs, «*le processus de formation de nouvelles unités lexicales. Selon les frontières qu'on veut assigner à la néologie, on se contentera de rendre compte des mots nouveaux, ou l'on englobera dans l'étude toutes les nouvelles unités de signification (mots nouveaux et nouvelles combinaisons ou expressions)*» (Dubois 1994: 322). Notons, que la nouveauté en question se détermine par rapport à une période précise de la vie de la communauté linguistique (Guilbert 1973: 12).

Il est bien connu, d'ailleurs, qu'un nombre non négligeable de néologismes arabes apparaît dans le dictionnaire général bilingue¹ français–arabe moderne. Ces néologismes sont des correspondants² arabes donnés à des entrées³ ou sous-entrées françaises étant originairement des termes scientifiques et techniques⁴. En effet, la terminologie scientifique et technique française, qui est en évolution constante, pénètre de plus en plus dans la langue générale et entre par la suite dans le dictionnaire général. Le correspondant arabe auxdites entrées et sous-entrées françaises n'existant pas encore en langue arabe, le lexicographe bilingue français–arabe fait, le plus souvent, appel à la création du correspondant arabe. Le dictionnaire général bilingue français–arabe *Abdelnour détaillé*, à titre d'exemple, va même jusqu'à consacrer un astérisque (*) au correspondant arabe qui constitue un néologisme créé par le lexicographe lui-même⁵. De toute évidence, une partie importante de ces néologismes sont proposés comme correspondants arabes à des entrées ou sous-entrées françaises étiquetées d'un domaine donné.

Ainsi, si le dictionnaire général bilingue français–arabe moderne joue un rôle dans la création de néologismes arabes et pour la raison primordiale que la culture française représentée par la langue française prévaut à l'époque moderne contre la culture arabe représentée par la langue arabe, le dictionnaire général bilingue moderne qui part, cette fois-ci, de l'arabe vers le français pourrait-il, de la même manière, provoquer des néologismes français en proposant un correspondant français non existant en langue française à une entrée ou sous-entrée arabe étiquetée d'un domaine technique ou scientifique? Si oui: quelles en sont les raisons, quels sont les types du produit néologique émergé et quels sont les types de néologie repérés ainsi que les procédés de formation des néologismes en question?

Pour répondre à ces questions, nous avons choisi de travailler sur le dictionnaire général bilingue arabe–français *Abdelnour moderne*. Ce dictionnaire est élaboré par *Jabbour Abdelnour* et édité en 2002 par *Dar El-Ilm Lilmalayin* à Beyrouth. Ledit dictionnaire prétend confronter la langue arabe moderne à la langue française moderne. Il est, d'ailleurs, l'un des dictionnaires généraux bilingues arabe–français les plus utilisés dans le monde arabe; ce dictionnaire étant exhaustif (*Abdelnour 2002: 8*) et pratique à consulter.

Cela dit, comme notre travail ne vise pas l'exhaustivité, nous nous sommes décidée d'étudier les correspondants français des entrées et sous-entrées arabes étiquetées d'un domaine technique ou scientifique de la lettre ت [th] dudit dictionnaire. Cette dernière est constituée d'un nombre de pages assez raisonnable à étudier: 16 pages. Ces 16 pages renferment, en outre, un bon nombre d'entrées et sous-entrées étiquetées d'un domaine scientifique et technique, en l'occurrence 170. Celles-ci sont réparties entre plusieurs domaines. Parmi ces domaines il y a notamment la philosophie, le commerce, le droit et la justice, les mathématiques, la chimie, la biologie, la botanique, l'entomologie, la médecine et l'astronomie.

Pour repérer les éventuels néologismes français⁶, nous avons cherché les correspondants français desdites 170 entrées et sous-entrées arabes dans le dic-

tionnaire général monolingue; celui-ci demeurant le facteur de reconnaissance des néologismes par rapport à la langue générale (Sablayrolles 2000: 254-255). En effet, ces derniers ne devraient normalement pas figurer dans un dictionnaire qui représente la langue de l'usage tel que le dictionnaire général monolingue. Cependant, vu les imperfections du dictionnaire, nous avons recouru à deux dictionnaires généraux monolingues français modernes à la fois, i.e. *Le Petit Robert* et *Le Petit Larousse*.

Pour pouvoir vérifier les néologismes dans le dictionnaire général monolingue par rapport à l'année 2002; celle-ci étant l'année de l'édition du dictionnaire *Abdelnour moderne* en question, les éditions des deux dictionnaires français choisis sont évidemment datées de cette année même, i.e. 2002.

Nous avons ainsi retenu les néologismes non trouvés dans aucun des deux dictionnaires généraux concernés. Et contrairement à toute attente, nous avons bien trouvé des néologismes français donnés comme correspondants à des entrées et sous-entrées arabes. Nous en avons, en effet, trouvé quatre. Malgré la médiocrité du nombre des néologismes trouvés, la simple existence de néologismes français donnés comme correspondants à des entrées et sous-entrées arabes étiquetées stimulent forcément la curiosité; la culture française représentée par la langue française prévalant à l'époque moderne contre la culture arabe représentée par la langue arabe!

Les quatre néologismes relevés sont, ainsi, les suivants: *fièvre octane* donné comme correspondant à la sous-entrée arabe حمى ثمانية [hummā thumāniyya] de même que *pennisète*, *sénébière* et *thapsie* successivement donnés comme correspondants aux entrées ثمام [thumām], ثفاء [thuffā'] et ثافسيا [thāfsyā]. Toutes les trois entrées appartiennent au domaine botanique alors que la sous-entrée relève du médical. Peut-être, dans la lettre ث [th], le taux déjà élevé d'entrées et sous-entrées étiquetées du domaine botanique par rapport à celles qui sont étiquetées de tout autre domaine scientifique et technique ($\approx 25\%$) a-t-il conduit à un tel résultat.

2. Analyse

2.1 Types du produit néologique émergé

Bien que le néologisme ne corresponde pas seulement à une nouvelle unité lexicale mais aussi à une nouvelle collocation (Gaudin et Guespin 2000: 228) ou combinatoire syntaxique (Mel'čuk et al. 1995: 117), tous les quatre néologismes retenus par notre étude sont des unités lexicales. Cela montre vraisemblablement qu'on a beaucoup plus besoin de créer des néologismes pour évaluer et nommer des objets — concrets et abstraits — que pour exprimer des idées. En effet, l'unité lexicale sert à dénommer des objets, des qualités, des comportements et actions et des états. Les combinatoires syntaxique et lexicale servent, par contre, à imposer certaines contraintes s'exerçant sur l'association des unités en discours.

Trois des quatre néologismes sont, par ailleurs, des unités lexicales simples et un seul néologisme constitue une unité lexicale complexe. En effet, chacune des unités lexicales; *pennisète*, *sénebière* et *thapsie*, correspond à ce qu'on appelle traditionnellement le mot, c'est-à-dire un élément linguistique significatif entre deux blancs. Le quatrième néologisme; *fièvre octane*, est par contre composé de deux éléments lexicaux autonomes: *fièvre* et *octane*. Ceux-ci se comportent comme une unité du point de vue de la classe de mots (la catégorie grammaticale), en l'occurrence, substantive. L'unité lexicale complexe dépasse généralement le mot traditionnel pour correspondre au syntagme lexicalisé.

Tous les quatre néologismes sont, d'ailleurs, des substantifs. Cela serait dû au fait qu'on a plus besoin de dénommer des objets concrets et abstraits (concepts) — exprimés par le substantif — que des comportements et actions — exprimés par le verbe — ou des qualités — exprimées par l'adjectif — ou des états — exprimés par l'adverbe (Sablayrolles 2000: 319).

2.2 Types de néologie repérés et procédés de formation des néologismes en question

Les quatre néologismes concernés sont créés par matrices internes⁷, c'est-à-dire produits dans le même système linguistique, en l'occurrence, le français. Tous les quatre néologismes seraient, en effet, empruntés à une forme ancienne de la langue française; le français du XIXe siècle. Lesdits néologismes constitueraient, ainsi, des emprunts internes, c'est-à-dire des emprunts non pas à d'autres systèmes linguistiques, à des langues étrangères. L'emprunt interne en jeu est effectué entre deux formes de la langue française; une est plus ancienne que l'autre. Ce type d'emprunts internes est, d'ailleurs, appelé le paléologisme (Sablayrolles 2000: 235).

De fait, aucun des quatre néologismes en question ne figure non plus dans le dictionnaire spécialisé daté toujours de l'année 2002. Ils ne figurent même pas dans le dictionnaire général ou spécialisé daté des XXIe ou XXe siècles. Cependant et contrairement à toute attente, nous avons trouvé lesdits néologismes dans des dictionnaires datés du XIXe siècle. Ainsi, *fièvre octane*, *thapsie* et *sénebière* figurent dans le fameux *Littre*⁸, qui est un dictionnaire général monolingue élaboré par *Emile Littré* et publié en 1889 par *Hachette*, alors que *pennisète* figure dans le *Complément du dictionnaire de l'Académie Française*⁹, qui est un dictionnaire spécialisé monolingue¹⁰ élaboré sous la direction de *Louis Barré* en 1842. Ce dernier néologisme ne figure même pas dans *Le Littré*.

Comme le néologisme *pennisète* ne figure que dans un dictionnaire spécialisé, il ne s'agit donc pas seulement de paléologisme créé par emprunt interne mais aussi de néologisme sémantique créé par changement de sens; plus précisément par extension de sens (Pruvost et Sablayrolles 2003: 112). En effet, ledit néologisme serait déterminologisé (passé de la langue spécialisée à la langue générale). Et pour qu'il soit déterminologisé, il doit forcément avoir subi une extension sémantique.

2.3 Raisons d'émergence du néologisme repéré

Deux raisons principales seraient l'origine de l'apparition des quatre néologismes en jeu:

- (1) L'absence de corpus lors de l'élaboration du dictionnaire dans le monde arabe

Aucun des deux entrées arabes; ثافسيا [thāfsyā] à laquelle correspond le néologisme *thapsie* ou حمى ثمانية [ḥummā thumāniyya] à laquelle correspond le néologisme *fièvre octave*, n'auraient existé en arabe en 2002. Elles ne figurent, en effet, dans aucun dictionnaire monolingue arabe général ni spécialisé de ladite année; le dictionnaire monolingue étant le représentant de la langue à un moment donné.

A notre grande surprise, pourtant, nous avons trouvé ثافسيا [thāfsyā] dans un dictionnaire monolingue scientifique ancien datée de l'année 1832, en l'occurrence, الشرح المغني [Al-sharḥ al-mughnī]. حمى ثمانية [ḥummā thumāniyya], qui ne figure même pas dans le dictionnaire ancien, figure par contre dans un dictionnaire bilingue français–arabe moderne. Il s'agit du dictionnaire *Al-kamel al-kabir* édité en 2001¹¹.

Aussi, pour élaborer sa nomenclature, le dictionnaire bilingue arabe–français moderne se base-t-il, entre autres, sur des dictionnaires anciens et le dictionnaire bilingue français–arabe moderne. Il ne se base surtout pas sur un corpus, composé d'énoncés réels, qui représente la langue à un moment donné.

Par conséquent et comme les correspondants français des entrées ثافسيا [thāfsyā] et حمى ثمانية [ḥummā thumāniyya] n'existaient pas en français de 2002, le lexicographe bilingue arabe–français a fait appel à la néologie.

- (2) L'écart culturel entre les deux langues concernées dans le dictionnaire bilingue en jeu, en l'occurrence, l'arabe et le français

En effet, comme «*élaborer un dictionnaire bilingue, c'est mettre en relation deux cultures*» (Pruvost 2002: 8) et que «*dans le cas des deux langues, le français et l'arabe, il existe des différences culturelles importantes*» (Cherifi 2009: 238); les deux civilisations arabe et française étant distinctes, ledit dictionnaire bilingue a recouru à la néologie pour donner un correspondant français à l'entrée arabe qui existait normalement en arabe en 2002 mais dont le correspondant français n'existait pas en français de l'année même.

C'est le cas des entrées arabes; ثُفَاء [thuffā'] à laquelle correspond le néologisme *sénebière* et ثُمَام [thumām] à laquelle correspond le néologisme *pennisète*, existant normalement en langue arabe en 2002. Nous avons, effectivement, trouvé toutes les deux entrées dans le dictionnaire général monolingue arabe de l'année 2002. Toutes les deux figurent, à titre d'exemple, dans المنجد في اللغة العربية المعاصرة [Al-munjed fī al-lughā al-'arabiyya al-mu'āšira] de l'année 2001¹². Ce dernier est un dictionnaire général monolingue arabe prétendant traiter la langue arabe moderne. ثُمَام [thumām] figure, de plus, dans المعجم العربي الأساسي [Al-

mu'jam al-'arabī al-'asāsī] de l'année 1999 et l'année 2003¹³. Ledit dictionnaire est un dictionnaire général arabe prétendant traiter la langue arabe essentielle.

Les correspondants français des deux entrées arabes en question n'existaient, par contre, pas en langue française en 2002. En effet, les deux genres de plantes que ثُفَاء [thuffā'] et ثُمَام [thumām] désignent n'étaient à priori pas connus dans la culture française en 2002.

Ainsi, face à cet écart linguistique dû à un écart culturel concernant une réalité culturelle spécifique, en l'occurrence botanique, le lexicographe bilingue arabe-français a fait appel à la néologie.

Rappelons tout de même, qu'en français du XIXe siècle, on trouvait des correspondants aux deux dites entrées arabes. En effet, les deux plantes désignées par ثُفَاء [thuffā'] et ثُمَام [thumām]¹⁴ auraient été connues dans la culture française au XIXe siècle bien qu'elles ne le soient plus en 2002. C'est peut-être la raison pour laquelle le lexicographe bilingue se serait servi du paléologisme au lieu de l'emprunt pur et simple ou de procédés d'explication.

Il est à signaler, finalement, que le lexicographe bilingue propose parfois des gloses définitionnelles à côté du néologisme français pour rendre la signification de celui-ci plus claire. Cependant, ces petites gloses ne figurent pas systématiquement à côté de chaque néologisme. De petites définitions sont, ainsi, proposées pour les deux néologismes *thapsie* et *sénebière* alors que les deux autres néologismes *fièvre octane* et *pennisète* apparaissent seuls sans le moindre commentaire.

3. Conclusion

Tel que le dictionnaire général bilingue français-arabe moderne, le dictionnaire général bilingue arabe-français moderne peut provoquer des néologismes comme correspondants à des entrées et sous-entrées étiquetées d'un domaine technique ou scientifique. Cependant la raison diffère. Pour le dictionnaire français-arabe, c'est le fait que la culture française prévaut à l'époque moderne contre la culture arabe à cause du développement scientifique et technique constante. Pour le dictionnaire arabe-français, c'est le manque de corpus d'une part et d'autre part l'écart culturel concernant une réalité spécifique connue par la culture arabe et non pas par la culture française; les deux civilisations arabe et française étant distinctes par nature de tout temps. La sénebière et le pennisète n'étaient, par exemple, pas connus dans la culture française en 2002 alors qu'ils l'étaient par la culture arabe. La sénebière s'utilisait en médecine douce et en cuisine alors que le pennisète dans l'élevage de volaille.

Tous les néologismes repérés sont des unités lexicales substantives comme, normalement, on a le plus besoin de créer des néologismes pour évaluer et dénommer des objets concrets et abstraits.

De plus, lesdites unités lexicales, ayant déjà fait partie d'une forme ancienne de la langue française, constituent des paléologismes créés par processus d'emprunt interne.

Cela dit, comme nous travaillons seulement sur une seule lettre (ث [th]), notre étude ne peut faire l'objet d'une généralité. Il serait, ainsi, intéressant d'étudier d'autres lettres du même dictionnaire en question et d'étudier d'autres dictionnaires généraux bilingues arabe–français modernes.

Notes

1. Un dictionnaire général bilingue est un dictionnaire qui n'est pas spécialisé dans un domaine particulier et qui «confronte deux systèmes linguistiques et notamment deux systèmes lexicaux» (Szende 1996: 116).
2. Dans notre travail, nous utilisons le terme *correspondant* par opposition à *équivalent*; ce premier s'employant dans la langue et l'autre dans le discours.
3. Les entrées d'un dictionnaire se définissent comme des mots ou des syntagmes à définir, généralement en caractères plus gras, suivis d'un ensemble de renseignements et terminés par un blanc.
4. Qui appartient à des domaines de spécialité; non pas à la langue générale.
5. Or, normalement, «un mot n'entre dans le dictionnaire que parce que les lexicographes ont jugé qu'il s'était déjà bien implanté dans l'usage. Ceux-ci ne font qu'entériner cet usage, ils ne le créent pas» (Sablayrolles 2006: 143).
6. Notons que le dictionnaire bilingue en question, en l'occurrence *Abdelnour moderne*, ne consacre pas de signe distinctif au correspondant qui constitue un néologisme créé par le lexicographe lui-même.
7. Les termes utilisés dans le présent travail qui concernent la typologie des néologismes sont empruntés à J. Tournier et J.-F. Sablayrolles. La typologie proposée par ces derniers est fondée sur les procédés de formation des néologismes et serait la plus détaillée, cohérente et hiérarchisée des typologies.
8. *Fièvre octane* figure sous l'entrée *octane* et est défini par: «fièvre intermittente qui revient tous les huit jours». *thapsie* et *sénébière* figurent comme entrées avec les définitions respectives suivantes: «nom d'un genre d'ombellifères à fruit oblong, échancré aux deux extrémités, muni de quatre ailes membraneuses» et «genre de plantes formé par une espèce de *lepidium* et une de *cochlearia*, crucifères».
9. *Pennisète* figure comme entrée et est défini par: «genre de plantes graminées».
10. C'est un complément du *Dictionnaire de l'Académie Française pour les sciences, les arts et les techniques*.
11. Il est à noter que *حمى ثمانية* [humma thumāniyya] figure dans le dictionnaire bilingue français–arabe comme correspondant à la sous-entrée *fièvre octane*. Ceci démontre, évidemment, que même le dictionnaire bilingue français–arabe ne se base pas sur un corpus représentant la langue à un moment donné. En effet, et comme nous l'avons montré plus haut dans cette étude, *fièvre octane* n'existait pas en français de 2001. De plus, peut-être ledit dictionnaire a-t-il fait la traduction littérale de *fièvre octane* pour arriver au correspondant arabe *حمى ثمانية* [humma thumāniyya].
12. C'est l'année d'édition la plus proche de l'année concernée par la présente étude; 2002.
13. De même, ce sont les deux années d'édition les plus proches de l'année concernée par la présente étude; 2002.

14. ثُفَاء [thuffā] désigne un genre de plantes crucifères. ثُمَام [thumām] désigne un genre de plantes graminées.

Références Bibliographiques

Dictionnaires

- Abdelnour, J.** 2002. *Abdelnour dictionnaire moderne arabe-français*. Beyrouth: Dar El-Ilm Lilmalayin.
- Abdelnour, J.** 2005. *Abdelnour dictionnaire détaillé français-arabe*. Beyrouth: Dar El-Ilm Lilmalayin.
- ALECSO.** 1999. *Al-mu'jam al-'arabi al-'asāsi*. Paris: Larousse.
- ALECSO.** 2003. *Al-mu'jam al-'arabi al-'asāsi*. Paris: Larousse.
- Barré, L.** 1842. *Complément du dictionnaire de l'Académie Française*. Paris: Didot.
- Dār al-mashreq.** 2001. *Al-munjed ft al-lughā al-'arabiyya al-mu'āšira*. Beyrouth: Dār al-mashreq.
- Dubois, J.** 1994. *Dictionnaire de linguistique et des sciences du langage*. Paris: Larousse.
- Kazrūfi, S.** 1832. *Al-sharḥ al-mughnī*. Inde: Calcutta.
- Larousse.** 2002. *Le Petit Larousse*. Paris: Larousse.
- Littré, E.** 1889. *Le Littré dictionnaire de la langue française*. Paris: Hachette.
- Reda, Y.** 2001. *Al-kamel al-kabir dictionnaire français-arabe*. Beyrouth: Librairie du Liban.
- Robert, P.** 2002. *Le Petit Robert*. Paris: Dictionnaires Le Robert.

Autres Références

- Cherifi, N.** 2009. «L'écart culturel dans les dictionnaires bilingues français-arabe, arabe-français». *Ela* 154: 237-248.
- Gaudin, F. et L. Guespin.** 2000. *Initiation à la lexicologie française: de la néologie aux dictionnaires*. Bruxelles, Belgique: Editions Duculot.
- Guilbert, L.** 1973. «Théorie du néologisme». *Cahiers de l'Association internationale des études françaises* 25(1): 9-29.
- Mel'čuk, I. et al.** 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve, Belgique: Editions Duculot.
- Pruvost, J.** 2002. Traduire l'écart culturel dans les dictionnaires bilingues? *Traduire* 2. Paris: Les belles lettres.
- Pruvost, J. et J.-F. Sablayrolles.** 2003. *Les néologismes*. Paris: Presses Universitaires de France.
- Sablayrolles, J.-F.** 2000. *La néologie en français contemporain, examen du concept et analyse de productions néologiques récentes*. Paris: Honoré Champion.
- Sablayrolles, J.-F.** 2006. La néologie aujourd'hui. Gruaz, C. 2006. *À la recherche du mot: de la langue au discours*: 141-157. Limoges: Lambert-Lucas.
- Szende, Th.** 1996. Problèmes d'équivalence dans les dictionnaires bilingues. Béjoint, Henri et Philippe Thoiron. 1996. *Les dictionnaires bilingues*: 111-126. Louvain-la-Neuve, Belgique: Editions Duculot.

Reflections on the Making of the *Grand dictionnaire chinois–français contemporain**

Jianhua Huang, *Centre for Linguistics and Applied Linguistics, Guangdong
University of Foreign Studies, Guangzhou, China*

and

Hai Xu, *Centre for Linguistics and Applied Linguistics, Guangdong
University of Foreign Studies, Guangzhou, China*
(Corresponding Author, xuhai1101@gdufs.edu.cn)

Abstract: This article discusses how the *Grand dictionnaire chinois–français contemporain* (GDCFC, 2014), one of the largest Chinese–French dictionaries, was designed and compiled. Due to the limited resources in the language pair of Chinese and French, GDCFC was designed as a bidirectional bilingual dictionary. To meet the needs of Chinese-speaking learners of French, the main user group, GDCFC extends the lemma list, and enriches the information of French equivalents in terms of their frequency, grammar and register. To be geared to the needs of French-speaking learners of Chinese, the secondary user group, GDCFC includes some headwords that fall into regional varieties of Chinese, and also provides useful information on the formulaicity and flexibility of Chinese characters (classifiers in particular). Some other features, such as the indication of POS and equivalents in context, and establishment of a semantic network of the nomenclature, were designed to benefit both groups of users. The making of this dictionary has implications for the compilation of other less-resourced bilingual dictionaries.

Keywords: CHINESE–FRENCH DICTIONARY, BIDIRECTIONAL, DESIGN FEATURES, USER NEEDS

Opsomming: *Gedagtes oor die samestelling van die Grand dictionnaire chinois–français contemporain.* Hierdie artikel bespreek hoe die *Grand dictionnaire chinois–français contemporain* (GDCFC, 2014), een van die grootste Chinees–Franse woordeboeke, ontwerp en saamgestel is. Weens die beperkte hulpbronne in die taalpaar Chinees en Frans, is die GDCFC as 'n tweerigting- tweetalige woordeboek ontwerp. Om te voorsien in die behoeftes van Chineessprekende aanleerders van Frans, die hoofgebruikersgroep, het die GDCFC die lemmalys, sowel as die inligting oor Franse ekwivalente in terme van hul frekwensie, grammatika en register, uitgebrei. Om aan te pas by die behoeftes van Franssprekende aanleerders van Chinees, die sekondêre groep, het die GDCFC sommige trefwoorde wat geklassifiseer kan word as streeksvariante van Chinees ingesluit, en ook nuttige inligting oor die formulisme en plooibaarheid van Chinese karakters (in

* A few of the ideas of this article were presented at the 11th International Conference of the Asian Association for Lexicography (ASIALEX), hosted by the Guangdong University of Foreign Studies in Guangzhou, Guangdong Province, China, 10-12 June 2017.

die besonder klassifiseerders), verskaf. Sommige ander kenmerke soos die aanduiding van woordsoort en ekwivalente in konteks, en die daarstelling van 'n semantiese netwerk van terminologie, is ontwerp om albei gebruikersgroepe daarby te laat baat vind. Die samestelling van hierdie woordeboek het ook implikasies vir die samestelling van ander hulpbronnbeperkte tweetalige woordeboeke.

Sleutelwoorde: CHINEES–FRANSE WOORDEBOEK, TWEERIGTING-, ONTWERPKENMERKE, GEBRUIKERSBEHOEFTE

1. Introduction

The project of the *Grand dictionnaire chinois–français contemporain* (GDCFC), which lasted 16 years¹, was finally completed in October 2014. With the coverage of more than 100,000 entries and over 100,000 illustrative examples, GDCFC is regarded as one of the largest Chinese–French dictionaries ever made. Since its publication, this dictionary has won public acclaim². In this article, we discuss how GDCFC was designed and compiled to meet the needs of both Chinese-speaking learners of French and French-speaking learners of Chinese.

According to the theoretical principles of lexicography, a dictionary is supposed to be of a specific type, and responds to the needs of certain users in certain tasks, for example, the needs of reception, production or translation by L1/L2 speakers (Atkins and Rundell 2008: 40–43; Svensén 2009: 14–18; Adamska-Salaciak 2013: 216). In fact, different types of monolingual Chinese dictionaries as well as bilingual English ones have been published in China. However, as far as the language pair of Chinese and French is concerned, the number of L2 Chinese learners in French-speaking countries and L2 French learners in China is quite small, and the resources (including dictionaries available)³ for this language pair are quite limited. Therefore, from the start of the dictionary project, GDCFC was designed to be a versatile bilingual dictionary with "a fusion of reception-oriented and production-oriented information" (Hannay 2003: 149). Just as Szende (2006: 20) claimed when it was still in preparation, this Chinese–French dictionary "vise à refléter la langue chinoise d'aujourd'hui dans toutes ses dimensions, intègre les besoins des deux publics et apporte différentes options de traduction, allant du sens littéral vers une expression authentiquement idiomatique".⁴

2. Features designed for the main user group

Since GDCFC was published in China, it is natural that priority was given to the needs of Chinese users, specifically those of adult learners of French and professional translators. Potential user situations include production in the foreign language (i.e. French) and translation into the foreign language (i.e. French), in addition to users' cognitive needs for foreign-language (i.e. French) learning (Tarp 2008). What this group of users need most is the availability of an extensive lemma list and rich information of French equivalents.

2.1 Extensive lemma list

To list or not to list semi-fixed and free expressions in a dictionary is a thorny issue. On the one hand, because of their structure and meaning, those expressions do not gain the status of an independent headword. If the editorial team followed the principle of one Chinese–French dictionary and listed in the entries such expressions as "锅的容量" (the volume of a wok), "艺术家的书籍" (books by an artist), and "婚姻有效性的障碍" (a barrier to the validity of marriage), the dictionary macrostructure would be redundant and devoid of rule (Huang 2014). On the other hand, for the purpose of production or translation, users may wish to have access to a rendering of those frequent semi-fixed and free Chinese expressions.

Based on their frequency of usage, GDCFC provides as many useful Chinese expressions as possible in an entry, for example,

【工业】 gōngyè industrie *f* ▷ 化学~ industries chimiques || 轻(重)~ industrie légère (lourde) || ~设备 équipement industriel || ~体系 système industriel || ~城市 villes industrielles || ~区 une zone industrielle || ~产品 articles [produits] industriels || ~化学 chimie industrielle || ~艺术 art industriel || ~技术 technique industrielle || ~污染 pollution industrielle || ~布局 répartition géographique des industries; répartition industrielle || ~界 monde industriel; milieu industriel || ~学校 école des arts et métiers; école polytechnique

The article 【工业】 consists of 14 subentries. Such an arrangement helps to expand, in a compact way, the number of useful multiword expressions in the dictionary.

To meet the production needs of users, GDCFC also enters approximately 3,000 new Chinese words or senses that occurred in recent years. Those neologisms have been strictly selected according to their frequency counts in Chinese language corpora as well as their established usages in some dictionaries of new words and phrases. For instance,

【房奴】 fángnú accros *m.pl* au [dépendants du, esclaves du] credit logement; individu ou ménage *m* surendetté [écrasé par ses mensualités] de credit immobilier
【微信】 wēixìn WeChat (application de messagerie rapide pour téléphone mobile, «micro message»)
【银色人才】 yínsè réncái personne *f* de talent à la retraite qui continue à contribuer à la vie en société

Undoubtedly, it would be untenable to clutter a dictionary with some "fad" words which would soon become obsolete.

Nowadays, a general-purpose dictionary cannot ignore scientific and technical terms, especially those that are used in everyday life. GDCFC is not

an exception, and lists approximately 10,000 technical terms, for example,

- 【彩超】 cǎicāo [(abrév.) pour 彩色多普勒超声] échographie *f* Doppler couleur
- 【磁浮列车】 cífú-lièchē train *m* à lévitation magnétique
- 【云计算】 yúnjìsuàn [[Inform.] Informatique *f* en nuage(s); infonuagique *f*; nuage *m* informatique; calcul *m* dématérialisé [dans les nuages]; informatique *f* virtuelle [dématérialisée, dans le nuage]; le cloud computing (angl.)

GDCFC distinguishes itself from other dictionaries of the same type in that the selection of those entries has drawn on the expertise of a group of specialists from forty disciplines. It thus ensures the representativeness of those terms and the accuracy of their translations.

GDCFC also offers some colloquial expressions and slang, for example,

- 【倒贴】 dàotiē ① subventionner *v.t* au lieu d'en gagner ② <fam. > (femme amoureuse) fournir de l'argent à son amant
- 【粉丝?】 fěnsī <fam.> fan *n*; admirat-eur(rice) *n*; groupie *f* ▷ 某歌星的~ les fans d'une chanteuse étoile

Some of the expressions are not even available in the authoritative *Contemporary Chinese Dictionary* (CCD, 2012) and the *Normative Dictionary of Contemporary Chinese* (NDCC, 2014). They are included, for they could satisfy the needs of users with different kinds of sociolects.

GDCFC embraces in its lemma list some frequent abbreviations as well. For example,

- 【东盟】 Dōngméng [(abrév.) pour 东南亚国家联盟] Association de l'Asie du Sud-Est (ANASE) ▷ ~国家 pays de l'ANASE
- 【妇代会】 fùdàihuì [(abrév.) pour 妇女代表大会] Assemblée *f* des déléguées de femmes
- 【个唱】 gèchàng [(abrév.) pour 个人演唱会] concert *m* donné par un seul chanteur [une seule cantatrice]; récital *m*

In brief, the reference needs of dictionary users can be largely satisfied by the complementary list of frequently-used semi-fixed/free Chinese expressions, neologisms, scientific/technical terms, colloquial expressions/slang, and common abbreviations.

2.2 Common French equivalents

In each article, GDCFC provides only common French equivalents to Chinese words or expressions, and carefully excludes rare or obsolete terms. In general, the vocabulary used in this dictionary does not exceed that in *Le Petit Larousse* (2013) or *Dictionnaire du français contemporain* (1980), except a small number of

scientific and technical terms (cf. Cowie 1999; Xu 2012).

Consider the entry 【傲然】. "Hautainement" is given as the first French equivalent in the *Dictionnaire chinois–français* (1990) and the *Nouveau Dictionnaire pratique chinois–français* (1996). But today this term is no longer mentioned in the contemporary French dictionaries, neither in *Le Petit Larousse* nor in *Le Robert* (1996). In its 1932–35 edition, the *Dictionnaire de l'Académie française* clearly indicates that the word "est peu usité" (is rarely used). Obviously, an equivalent such as "hautainement" would mislead Chinese users rather than helping them to master French. GDCFC instead offers some more common equivalents to 【傲然】: *fièrement adv*; *de manière hautaine [imposante]*; *avec fierté*.

2.3 Useful grammatical information of French equivalents

Almost all existing Chinese–French general dictionaries, including *Dictionnaire Ricci chinois–français* (2014), do not offer grammatical information of French equivalents. Only the *Dictionnaire chinois–français* (1990) indicates the gender information (masculine or feminine) in the nouns of French translations, but it gives no clues for verbs and other word classes. In a dictionary targeting French native speakers, this information might be redundant. However, in a dictionary primarily for Chinese-speaking learners of French, the grammatical information is very useful and even indispensable. Therefore, GDCFC made an effort to explicate the grammatical functions of French translations. For instance,

- 【会面】 *huì//miàn* *se rencontrer v.pr*; *se voir v.pr*; *rencontrer v.t*; *avoir une entrevue (avec qn)*; *rencontre f..*
- 【远航】 *yuǎnháng* *naviguer v.i au long cours [sur un long parcours]*; *partir v.i en expédition*; *navigation f lointaine [hauturière, au long cours]..*
- 【远客】 *yuǎnkè* (Q. 个、位) *hôte m (hôtesse f) [voyageu-r(se) n] qui vient de loin*
- 【原初】 *yuánchū* *début m*; *au début loc.adv*; *à l'origine loc.adv*; *d'abord loc.adv*
- 【原生】 *yuánshēng* *primordial(e) a*; *protogène a*; *primiti-f(ve) a*; *proto- préf*
▷ ~矿物 *minerais primitifs, minéral originel* ◆ ~植物 *protophyte m*

The grammatical information of French equivalents, such as gender of nouns (*m/f*), transitivity of verbs (*v.i/v.t/v.pr*), and subcategories of adjectives (*a*) and adverbs (*loc.adv*), will help learners to produce grammatically correct French. Those syntactic codes might not appear user-friendly, but adult learners of French in China, the majority of users that this dictionary targets, are actually quite familiar with those codes as a result of their French grammar instruction at college.

2.4 Appropriate register of French equivalents

More often than not, there are several possible translations⁵ of a word or phrase. While using a dictionary for production or translation, learners often ignore the register of an L2 word, and hence fail to use it appropriately. GDCFC fills this gap by briefly indicating the register of some French equivalents. For example,

- 【妓院】jìyuàn (Q. 家、座) maison *f* de prostitution; maison *f* close; maison *f* publique; bordel *m* (fam.) • ≈ 娼家
- 【淫欲】yínyù lubricité *f*; désir *m* impur [impudique, sexuel]; concupiscence *f* (litt.) ▷ 饱暖思~ Bien nourri et habillé, on commence à désirer la lubricité.
- 【怨恨】yuànhèn ❶ en vouloir à *loc.v*; garder rancune à *loc.v*; haïr *v.t*; abhorrer *v.t* (litt.) ❷ ressentiment *m*, rancune *f*, rancœur *f*, haine *f*, animosité *f* ▷ 充满~ être plein de haine
- 【带头】dài//tóu prendre la tête de; donner l'exemple; prendre l'initiative de; mener la danse (péj.) ▷ 他~跳进水里。Il fut le premier à se jeter à l'eau. || ~作用 rôle de promoteur; exemple d'un pionnier; influence exercée par les initiateurs ◆~羊 sonnailler *m* de troupeau || ~者 locomotive *f* (fig.) • ≈ 领头

In the above articles, the French equivalents "bordel" is marked with (fam.), "concupiscence" and "abhorrer" with (litt.), "mener la danse" with (pej.), and "locomotive" with (fig.). Those metalanguage labels help to prevent Chinese learners from using the French words inappropriately.

3. Features designed for the secondary user group

Due to a small number of L2 Chinese learners in French-speaking countries and limited resources of the Chinese–French language pair, publishers are reluctant to produce a Chinese–French dictionary exclusively for this group of users. Hence, GDCFC targets L2 Chinese learners as well, though as a secondary intended user group. Their potential user situations embrace reception of the foreign language (i.e. Chinese) and translation from the foreign language (i.e. Chinese), in addition to their cognitive need for foreign-language (i.e. Chinese) learning (Tarp 2008). This group of users may wish to obtain information on the usage of Chinese characters or words.

3.1 Words in regional varieties of Chinese

In addition to *Putonghua* (Mandarin Chinese), L2 Chinese learners are likely to encounter some words and phrases that are mainly used in Taiwan, Hong Kong, Macau or the overseas Chinese communities. To help users understand

their meanings, GDCFC offers some of the most common nomenclature, and marks them with ⟨dial.⟩. For example,

- 【三脚猫】 sānjiǎomāo ⟨dial.⟩ chat *m* à trois pattes (trad. litt.); personne *f* qui n'a que des connaissances superficielles, personne *f* dénuée de compétence
【烧刀子】 shāodāozi ⟨dial.⟩ eau-de-vie *f* forte
【失失慌慌】 shīshīhuānghuāng ⟨dial.⟩ • ≈ 慌张
【手袜】 shǒuwà ⟨dial.⟩ • = 手套

One might argue that the label ⟨dial.⟩ is not informative enough, and that it would be preferable to specify the regions by their name. However, as far as L2 Chinese learners are concerned, they would show more interest in the meanings of those words than their regional origin.

3.2 Classifiers in Chinese

Classifiers (or quantifiers), which can be roughly divided into "sortal classifiers" and "measure words" (Ahrens and Huang 2016), often challenge L2 Chinese learners. Why does one say "一匹马" (a horse) and "一头牛" (a cow/bull) instead of *"一头马" and *"一匹牛"? It is difficult to deduce the usage of classifiers from a grammatical rule. Thus, L2 Chinese learners often wish to find information on classifiers from a dictionary.

Under the noun entries, GDCFC uses the sign (Q) to specify the classifiers that they take. For example,

- 【秘诀】 mījué (Q. 条) méthode *f* secrète; recette *f* secrète; secret *m*; truc *m*; clé *f* ▷ 成功的~ secret du succès; recette de la réussite
【模型】 móxíng (Q. 套、具、副) ① modèle *m*, maquette *f* ▷ 飞机~ modèle d'avion || 车身~ modèle de carrosserie || 建筑~ modèle d'un édifice || 缩比~ modèle réduit, maquette réduite || 全尺寸~ maquette grandeur réelle || 百分之一的~ modèle réduit au 1/100 ② gabarit *m*; moule *m*; matrice *f*; forme *f* ▷ 石膏~ moule en plâtre • ≈ 模子[múzi]
【蘑菇¹】 mógu (Q. 朵、片、棵) champignon *m* comestible; champignon *m* ▷ 采~ cueillir des champignons ◆~房 champignonnière *f*; meule *f* (de terreau et de fumier)

In some entries, several interchangeable classifiers are given.

GDCFC also indicates those words or terms that can be used as classifiers with the sign ⟨q.⟩, and at the same time exemplifies the contexts of usage in which they apply. For example,

- 泡³ pāo ⟨q.⟩ [s'appliquant uniquement à l'urine et aux excréments] ▷ 撒泡尿 pisser (un coup); faire pipi || 拉泡屎 aller à la selle || 一泡狗屎 un étron de chien

匹² (*疋⁰) pǐ <q.> ① [s'appliquant au cheval, à l'âne, etc.] ▷ 一匹马 un cheval
|| 两匹骡子 deux mulets || 100匹马的马群 un troupeau de chevaux de
cent têtes ② [s'appliquant à la pièce ou au rouleau d'étoffe] *pi*, 1 *pi* = 50
ou 100 *chi* (尺*) ▷ 一匹布 une pièce d'étoffe; un rouleau de tissu
【梭子²】suōzi ① cartouchière *f* ② <q.> [s'appliquant à la cartouche] char-
geur *m*; rafale *f* ▷ 打了一~子弹 vider un chargeur, tirer une rafale

Apparently, such treatment will help L2 Chinese learners to have a clearer picture of specific classifiers preceding a noun.

3.3 Formulaicity of some Chinese characters

Whether one Chinese character can be freely combined to form a word is contingent upon morphological restrictions on it. Our experience in teaching Chinese as a foreign language shows that learners in the West have difficulty in distinguishing the stand-alone contemporary Chinese characters from those that are not. GDCFC indicates the Chinese characters, which are used only as morpheme, syllable or affix, respectively with the labels of <morph.>, <syll.> or <aff.>. For example,

丫 (*桠⁰ 榧⁰) yā ① <morph.> fourche *f*; bifurcation *f* ▷ 树丫 fourche d'un
arbre || 枝丫 branche fourchue; rameau; branchette ② <morph.> <p. ext.>
chose *f* [objet *m*] en forme de fourche ③ <morph.> <dial.> fillette *f* ▷
小丫 petite fille; fillette ④ nom de famille
俩 (俩) liǎng <syll.> 伎[jì]俩 •另见liǎ
老³ lǎo <aff.> ① [devant certains noms désignant des personnes] ▷ 老百姓* ||
老大娘* ② [devant certains noms d'animaux ou de plantes] ▷ 老虎* ||
老鹰* || 老玉米* ③ [devant un nom de personne monosyllabique] ▷ 老王
Lao Wang ④ [devant un numéral, de deuxième à dixième, désignant
l'ordre des rangs d'âge] ▷ 她是老二。Elle est née la deuxième. || 王老三
Wang le troisième ⑤ <dial.> cadet(te) *a*; benjamin(e) *n* ▷ 老儿子 fils cadet;
le benjamin de la famille || 老闺女 fille cadette; la benjamine de la famille

In this way, French learners of Chinese can gain useful information on the formulaicity of some Chinese characters, and reduce their errors as a result of arbitrary use of those characters.

3.4 Flexibility of some Chinese characters

In contrast to formulaicity, some Chinese words, which are composed of two or more characters, can be used in a "flexible" way. One may, where appropriate, detach a character (i.e. to insert a meaning unit), repeat certain characters (to strengthen the tone or to mitigate meaning), or even reverse the characters (to form another word or synonym).⁶

3.4.1 Detachedness of some Chinese characters

GDCFC marks such a headword with "/" in the middle of its phonetic notation, and offers examples of usage:

- 【搭伴】dā//bàn (～儿) tenir compagnie (à qn); aller [partir] ensemble (à l'occasion); voyager en compagnie de (qn) ▷ 他也去北京, 你可以和他搭个伴儿。 Il va aussi à Beijing, tu pourras l'avoir comme compagnon de voyage. || 不与某人～ fausser compagnie à qn
- 【带好儿】dài//hǎo saluer qn de la part de...▷ 你回家时, 给叔叔带好儿。 Dis bonjour à mon oncle [Salue mon oncle] de ma part quand tu rentres à la maison.

3.4.2 Repeated use of some Chinese characters

GDCFC highlights the Chinese characters (morphemes) which can be used repeatedly. There are three types:

- 1) Type AB-AAB: One morpheme is repeated, e.g. 操心-操操心.
- 2) Type AB-AABB: Both morphemes are repeated, e.g. 安稳-安安稳稳.
- 3) Type AB-ABAB: The whole word is repeated, e.g. 反省-反省反省.

【操心】cāo//xīn se préoccuper (de) *v.pr.*; se soucier (de) *v.pr.*; s'inquiéter (de) *v.pr.*; se faire du souci; se donner du mal ▷ 你不必为那件事～。 Ne vous embarrassez pas de cette affaire-là. || 他为孩子操碎了心。 Il s'est fait beaucoup de souci pour ses enfants. || 我得为儿子的前途多操操心。 Je dois me préoccuper de l'avenir de mon fils.

【安稳】ānwěn ① stable *a*; paisible *a*; solide *a*; ferme *a*; sûr(e) *a* ▷ 安安稳稳的生活 vie paisible || 睡得安安稳稳 dormir tranquillement || 船行～。 Le bateau s'avance doucement. ② [en parlant surtout des enfants] sage *a*; posé(e) *a*; sérieux-x(se) *a*

【反省】fǎnxǐng faire un retour sur soi-même; faire réflexion sur soi-même; faire un examen de conscience ▷ 你是怎么犯错误的, 该好好～～。 Comment as-tu pu commettre une telle erreur? Tu dois faire un examen de conscience [effectuer un sérieux retour sur toi-même].

3.4.3 Reversed use of some Chinese characters

GDCFC also marks the keyword entries in which both morphemes (characters) can be reversed (i.e. AB-BA).

This unique linguistic phenomenon in the Chinese language increases the learning burden on L2 Chinese learners. Although a general dictionary cannot deal with this problem in length, we believe that a simple indication will be

useful to French users. In this dictionary,

- 1) the sign "=" indicates that the two morphemes are interchangeable, the word and the inverted keeping the same meaning (AB = BA);
- 2) the sign "≈" indicates that the two morphemes can be reversed in some cases, the inverted word keeping a close meaning; and
- 3) the sign "≠" indicates that the reverse word has a completely different meaning.

For example,

【脊背】 jǐbèi (=背脊*) dos (d'un être vivant) *m* ●=背¹①

【建构】 jiàngòu (≈构建*) [appliqué surtout à qch. d'abstrait] construire *v.t*; constituer *v.t*; instaurer *v.t* ▷~新的理论体系 établir un nouveau système théorique

【爱心】 àixīn (≠心爱*) (Q 份、番) cœur *m* compatissant; amour *m*; affection *f*

Obviously, such information given in GDCFC will help L2 Chinese learners to gain a better understanding of the flexibility of some Chinese characters, and use them correctly.

4. Features designed for both groups of users

GDCFC's design lies in its practical consideration of the needs of both groups of Chinese and French speakers: the needs of Chinese users producing French, and the needs of reception of Chinese by French speakers. We believe that this approach which is oriented to the needs of both groups of users is not necessarily opposed or contradictory, but is, to some extent, complementary. Even in the case of flexible use of Chinese characters as mentioned above, although the arrangement mainly aims to facilitate French speakers' comprehension, it can still help Chinese speakers to be sensitive to this type of linguistic phenomena.

The following design will benefit both groups of users at large.

4.1 Indication of POS in context

GDCFC attempts, whenever possible, to reflect the grammatical categories of Chinese headwords. The issue of indication of parts of speech (POS) of Chinese characters is, however, complex, and has provoked controversy among specialists. Sometimes, the POS indication in CCD and NDCC — the two authoritative monolingual Chinese dictionaries — does not properly correspond to the actual usage. Consider the word "免费" (free, gratis, free of charge). The two Chinese dictionaries categorize it into 副 (*v.*). Yet in natural context, "免费" is used more often than not as adjective or adverb. In addition, the two dictionaries sometimes contradict each other in POS indication. For example, while CCD

labels "恼火" (irritated; be fuming at) with 形 (*adj.*), NDCC marks it as 动 (*v.*).

Since it is not easy to determine POS labels from existing Chinese dictionaries and the editorial team of GDCFC could hardly spare time to establish their own based on the words' usage patterns found in large Chinese corpora, it is advisable not to fall into the trap of POS indication. Instead, GDCFC illustrates the various usages of a headword with examples, and uses, if possible, translation equivalents of the same POS to reflect its grammatical categories. For instance,

- 【免费】 miǎn//fèi dispenser (qn) des frais; gratuit(e) *a*; gratuitement *adv*, gratis *adv*; à titre gratuit *loc.adv* ▷ ~入场 entrée gratuite; entrée libre || ~教育 enseignement gratuit || ~学校 école gratuite || ~门诊 consultations gratuites || ~随带行李 bagages en franchise || ~为病人看病 traiter les malades gratis; soigner les malades gratuitement
- 【恼火】 nǎohuǒ se fâcher *v.pr*; s'emporter *v.pr*; s'irriter *v.pr*; se vexer *v.pr* ▷ 我直说了, 你可别~。 Je vais te parler franchement, seulement ne te fâche pas. || 这些话使他很~。 Ces propos l'avaient exaspéré. || 他慢条斯理, 叫我真~。 Sa lenteur me fait bouillir. || 他动不动就~。 Il s'emporte facilement, même contre un rien. • ≈ 恼怒 ①
- 【概述】 gàishù ① résumer *v.t*; exposer *v.t* sommairement ▷ 他向我们~了他的意图。 Il nous a exposé sommairement ses intentions. ② exposé *m* sommaire; aperçu *m* ▷ 对当前形势作一个~ donner un aperçu [brosser un tableau] de la situation actuelle

The corresponding French equivalents of "免费" as exemplified in context of usage show that this word can function as adverb, adjective and verb. In the same vein, those contextual equivalents illustrate the usage of "恼火" as either verb or adjective, and that of "概述" as either verb or noun.

4.2 Equivalents in context

Since cognitive (semantic, systemic or prototypical) equivalents are not easily available (Adamska-Sałaciak 2010, 2011), we argue that as an alternative, equivalents presented in a dynamic way can cater for users' reference needs. In the translation of entry words, especially in the rendering of idioms, GDCFC offers many explanatory and translational equivalents. A headword is first translated literally, then explained, and finally, a fixed French expression, which roughly corresponds to its Chinese counterpart but uses different lexical means and/or imagery, is given if available. All this would be accompanied by examples of usage. For instance,

- 【云开日出】 yúnkāi-rìchū Les nuages se dissipent, le soleil réapparaît: revirement de fortune. | Le soleil dissipe les nuages. | Après la pluie vient le beau temps. (prov.) ▷ 他历经许多困苦, 今天总算~, 看到希望了。 Aujourd'hui, après maintes difficultés, les nuages se sont dissipés, le soleil réapparaît, et finalement il voit de l'espoir dans l'avenir.

【游戏人生】yóuxì- rénshēng jouer avec la vie; prendre la vie comme un jeu; mener une vie désinvolte; agir sans aucune contrainte ▷ 现在有些年轻人无所事事, ~。De nos jours, certains jeunes mènent une vie oisive et pleine de désinvolture.

In our view, different types of equivalents at various levels can assist Chinese speakers in finding an appropriate French counterpart of a Chinese idiom, and assist French speakers in understanding the meanings of an idiomatic expression in Chinese.

4.3 Semantic network of the nomenclature

The nomenclature in a dictionary binds, as a whole, the semantic network of words. In GDCFC, synonyms or near synonyms of some headwords are thus offered. Unlike other dictionaries in which a static list of synonyms are given, GDCFC presents the synonyms in a typical context. For example,

【得当】dédàng convenable *a*; adéquat(e) *a*; opportun(e) *a*; à point *loc.adv*, à propos *loc.adv* ▷ 采取~ (适当*) 的措施 prendre des mesures convenables [opportunes] || 措词~ (恰当*) s'exprimer en termes convenables [appropriés] || 回答~ (恰当*) répondre de façon adéquate
【寄存】jìcún mettre *v.t* à la consigne; mettre *v.t* en dépôt; consigner *v.t*; déposer *v.t* ▷ ~ (存放*) 行李 consigner ses bagages ...

Typical examples of usage will help users differentiate synonyms more accurately.

GDCFC also strives to cross-reference words or related terms. In this way, it has established a link between different words that belong to a system, and helps users to create a semantic network in their mental lexicon. For instance,

【霸王鞭¹】bàwángbiān ① ... ② danse *f* folklorique du bâton coloré
●=花棍舞、打连厢
【扛大梁】káng dàliáng (fig.) assumer une grande responsabilité; jouer le rôle principal ●=挑大梁
【跟斗】gēndou (dial.) ●≈跟头
【客舍】kèshè (litt.) ●≈客店
【干租】gānzū louer *v.t* sans équipage; location *f* sans équipage ●←湿租
【倒果为因】dàoquǒwéiyīn inverser effet et cause; prendre l'effet pour la cause ●←倒因为果

In the above entries, "●=" is used to indicate synonyms; "●≈" represents near synonyms; and "●←" links antonyms. Thus, GDCFC shows users, back and forth, useful links in the dictionary text. They learn not only a word or a single term, but also a series of words or related terms. To our knowledge, no Chinese-French dictionaries, and even no general Chinese-English ones, have treated

items in such a way. This dictionary might be the first attempt of this kind.

5. Conclusion

This dictionary has some distinctive features that other Chinese bilingual dictionaries (including Chinese–English/English–Chinese ones) are devoid of. Apart from an extensive and up-to-date lemma list, GDCFC pays special attention to the frequency, grammar and register of equivalents. It gives a lengthy treatment of formulaicity and flexibility of L2 morphology. It illustrates the POS as well as meaning of a headword in context of usage. It integrates the entry words into a semantic network by offering synonyms in context and cross-referencing related words.

The making of GDCFC has implications for the compilation of other less-resourced bilingual dictionaries. Dictionaries of this type often have to follow the design of bidirectional dictionaries⁷ (Hausmann and Werner 1991; Hannay 2003; Adamska-Salaciak 2013), for "it is cheaper to produce one dictionary which will be sold in both markets than invest in two different ones" (Adamska-Salaciak 2013: 215). This does not mean that a bidirectional dictionary is confined to a skeleton structure, with only equivalents given. Far from it. A bidirectional dictionary can be "learnerized" in the way as a monodirectional bilingual dictionary does (Granger and Lefer 2016). In other words, such a dictionary can be designed to meet the needs of both groups of users in question. Priority is given to the morphological, semantic, syntactic and pragmatic points of an L2 lexical item with which learners may run into difficulty. It has to be admitted that some information in a bidirectional dictionary might look redundant to the group of L1 users, but that is a compromise before a monodirectional dictionary targeting a specific learner population is finally produced. A well-designed bidirectional dictionary can still boast some of its notable features.

Endnotes

1. It took such a long time to complete it, because a number of professors of French, who had been involved in the making of this dictionary, were dropped from the editorial team due to other commitments, and this project almost came to a halt when the editor-in-chief was diagnosed with cancer in 2008.
2. For instance, in 2018, this dictionary won the "Chinese Government Award for Publishing" — the highest-level award in the Chinese publishing industry.
3. Apart from several pocket bilingual dictionaries, there are only 4 medium-sized Chinese–French dictionaries published before 2014: *Dictionnaire français de la langue chinoise* (1976/1986/1994/1999), *Dictionnaire chinois–français* (1990), *Nouveau dictionnaire pratique chinois–français* (1996), and *Grand dictionnaire Ricci de la langue chinoise* (2001), and most of them have not been updated.
4. This Chinese–French dictionary "aims to reflect today's Chinese language in all its dimensions, to integrate the needs of both Chinese and French speakers, and to show different translation options, ranging from a literal to an authentic idiomatic expression".

5. See Section 4.2 for further discussion of the equivalency issue.
6. One major category of those Chinese words in flexible use is "separable word" or *liheci* (Packard 2016: 74-76).
7. The terminology is inconsistent in the literature. Landau (2001) and Atkins and Rundell (2008) use the term "bidirectional" dictionaries to refer to what Hausmann and Werner (1993), Han-nay (2003) and Adamska-Salaciak (2013) call "biscopal" (Lx-Ly and Ly-Lx) dictionaries.

Acknowledgements

This study was supported by the National Social Science Fund of China (Grant No. 15BYY062), and by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities in P.R. China (Project No. 17JJD740004). We would like to thank Professor Arleta Adamska-Salaciak and Professor Danie Prinsloo for their valuable comments and suggestions.

References

A. Dictionaries

- Académie française (Ed.).** 1932–35. *Dictionnaire de L'Académie française*. Eighth edition. Paris: Hachette.
- Association Ricci pour le grand dictionnaire français de la langue chinoise and The Commercial Press Dictionary Research Centre (Eds.).** 2014. *Dictionnaire Ricci chinois-français*. Beijing: The Commercial Press.
- Dubois, J. (Ed.).** 1980. *Dictionnaire du français contemporain illustré*. Paris: Larousse.
- Huang, J.H. (Ed.).** 2014. *Grand dictionnaire chinois-français contemporain*. Beijing: Foreign Language Teaching and Research Press. (GDCFC)
- Institut Ricci (Ed.).** 1976/1986/1994/1999. *Dictionnaire français de la langue chinoise*. Taipei: Institut Ricci-Kuangchi Press.
- Jiang, L.S., J.C. Tan and R. Cheng (Eds.).** 2012. *Contemporary Chinese Dictionary*. Sixth edition. Beijing: The Commercial Press. (CCD)
- Larousse (Ed.).** 2013. *Le Petit Larousse illustré en couleurs*. Paris: Larousse.
- Li, X.J. (Ed.).** 2014. *Normative Dictionary of Contemporary Chinese*. Third edition. Beijing: Foreign Language Teaching and Research Press, and Language and Culture Press. (NDCC)
- Liu, J. and X.R. Liu (Eds.).** 1996. *Nouveau dictionnaire pratique chinois-français*. Beijing: Social Sciences Academic Press.
- Morvan, D. (Ed.).** 1996. *Le Robert illustré d'aujourd'hui en couleur*. Paris: Le Robert.
- Xu, J.Z. and L.G. Guo (Eds.).** 1990. *Dictionnaire chinois-français*. Beijing: The Commercial Press.

B. Other literature

- Adamska-Salaciak, A.** 2010. Examining Equivalence. *International Journal of Lexicography* 23(4): 387-409.
- Adamska-Salaciak, A.** 2011. Between *designer drugs* and *afterburners*: A Lexicographic-Semantic Study of Equivalence. *Lexikos* 21: 1-22.

- Adamska-Sałaciak, A.** 2013. Issues in Compiling Bilingual Dictionaries. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 213-231. London: Bloomsbury Academic.
- Ahrens, K. and C.-R. Huang.** 2016. Classifiers. Huang, C.-R. and D.X. Shi (Eds.). 2016: 169-198.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- Granger, S. and M.-A. 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.
- Hannay, M.** 2003. Types of Bilingual Dictionaries. Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 145-153. Amsterdam/Philadelphia: John Benjamins.
- Hausmann, F.J. and R.O. Werner.** 1991. Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: Eine Übersicht. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1991. *Dictionaries. An International Encyclopedia of Lexicography*: 2729-2769. Berlin/New York: Mouton de Gruyter.
- Huang, C.-R. and D.X. Shi (Eds.).** 2016. *A Reference Grammar of Chinese*. Cambridge: Cambridge University Press.
- Huang, J.H.** 2014. Jubi Weixia — Jian Ping Lishi Hanfa Cidian (A Review of Ricci de caractères chinois). *Cishu Yanjiu (Lexicographical Studies)* 202(4): 72-75.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York/Cambridge: Cambridge University Press.
- Packard, J.** 2016. Lexical Word Formation. Huang, C.-R. and D.X. Shi (Eds.). 2016: 67-80.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Szende, T. (Ed.).** 2006. *Le français dans les dictionnaires bilingues*. Paris: Honoré Champion.
- Tarp, S.** 2008. *Lexicography in the Borderline between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Xu, H.** 2012. A Critique of the Controlled Defining Vocabulary in *Longman Dictionary of Contemporary English*. *Lexikos* 22: 367-381.

Roderick McConchie and Jukka Tyrkkö (Editors). *Historical Dictionaries in Their Paratextual Context*. 2018, xii + 318 pp. ISBN: 978-3-11-057286-5 (Hardback). Lexicographica. Series Maior 153. Berlin/Boston: Walter de Gruyter. Price € 99.95.

This volume of essays examines dictionaries and dictionary-making practices in their broadly-interpreted paratextual context.¹ The term 'paratext', borrowed from literary theory, has been infrequently used in lexicography (e.g. Van Male 2004). As the Editors put it (McConchie and Tyrkkö 2018: vii), "Dictionaries exist in and are abound by a context ... These works have usually been seen as finished, immutable product, without asking how this product was produced, or what its subsequent fate was." They go on to explain that "Reflection on the nature and role of dictionaries raises many questions. Who wrote and compiled dictionaries and why? Who patronized their publication and their authors, financed them, and to whom were they dedicated? How were they set up for printing, advertised, sold, and distributed? What were the conventions of dictionary layout? How did this change over the years? Who bought and read them? What collections did they find their way into, and for what reasons? What is the individual history of individual copies of dictionaries?" The wealth of questions for which reliable, or even tentative, answers are sought is not confined to the traditional field of dictionary criticism (cf. Akasu 2013, Swanepoel 2017). Instead, with its integrative and insightful 'anthropological' approach to historical dictionaries and their compilers, the monograph takes us into a broader area of cultural history, opening new and exciting avenues of research.

The historicity of dictionaries merits a comment. The *Oxford English Dictionary* (OED), alluded to by Michael Adams and discussed by Sarah Ogilvie, remains the only indisputably historical dictionary under scrutiny here. Numerous endeavours compiled in the past, which remain of great interest to historians of lexicography, may also contribute to a better understanding of historical lexicography. To make these sources a legitimate object of study, however, it would be recommended to call them more adequately semi- and quasi-historical dictionaries (Podhajecka 2018: 143, 145).

The preface, introducing the methodological issues behind the topic, is followed by twelve articles, each of which highlights a particular paratextual context in specifically English lexicography. Seija Tiisala's story of Jacob Serenius's trilingual dictionary, *Dictionarium Anglo-Svethico-Latinum*, is the only exception to the rule. The book closes with the contributors' biographies and a comprehensive index.

The selection of essays opens with Michael Adams's "Reading Trench Reading Richardson," which looks at the possible motivations and reasons for Richard Chenevix Trench's ample annotations of Charles Richardson's *New Dictionary of the English Language* (1836–37). Adams claims that this is "important evidence of Trench's practice as a lexicographer and dictionary critic," which became a basis not only for Trench's seminal paper, *On Some Deficiencies*

in *Our English Dictionaries* (1857), but also for the OED. It would be facile to refer to the reasoning that emerges from Adams's examination of the subsequent layers of Trench's annotations as a detective story, but a detective story it undoubtedly is, one drawn out brilliantly. I should perhaps offer a minor suggestion: Adams admits to having consulted the fifth edition of Trench's *Select Glossary of English Words Used Formerly in Senses Different from Their Present* (1879) for comparison, but one may find online all the previous editions.²

Fredric T. Dolezal and Ward J. Risvold's article is dedicated to John Wilkins's *Essay Towards a Real Character and a Philosophical Language* (1668), an ambitious attempt to produce a universal language. It has been considered an important, albeit forgotten, project in the history of linguistics and its component, *An Alphabetical Dictionary ...*, in the history of lexicography (Dolezal 1985: 1). Dolezal and Risvold examine a group of authors (e.g. William Lloyd), booksellers (e.g. Samuel Gellibrand), and printers (e.g. Thomas Roycroft) in seventeenth-century England in order to identify the most likely candidate to have printed the *Essay*, that is, Anne Maxwell. In so doing, they demonstrate an impressive knowledge of the London book trade, people associated with it, and events that may have exerted an influence, such as the 1666 Great Fire of London. This contribution offers unique glimpses into the early publishing market, providing evidence hitherto unknown.

M. Victoria Domínguez-Rodríguez and Alicia Rodríguez-Álvarez's contribution, "As Well for the Entertainment of the Curious, as the Information of the Ignorant," looks at variously-titled encyclopedic supplements prefixed or appended to eighteenth-century dictionaries of English. Having taken into account a huge selection of general dictionaries listed in Alston (1966) and included in EEBO, they excluded those without paratexts as largely irrelevant. The remaining list of sixteen dictionaries served as the basis for the study. It may be important to note that Domínguez-Rodríguez and Rodríguez-Álvarez not only explored the educational, informative, or entertaining functions of paratextual elements, but they also provided a working typology of them. This is a solid piece of research based on an accurately applied methodology, which has led the authors to arrive at sound conclusions.

The following essay, "Printed English Dictionaries in the National Library of Russia to the Mid-Seventeenth Century," was written by Olga E. Frolova and Roderick McConchie. They seek to ascertain how sixteenth- and seventeenth-century dictionaries in the holdings of St. Petersburg's library were acquired and from where. Frolova and McConchie present us with brief accounts of the first seventeen dictionaries, from Thomas Cooper's *Biblioteca Eliotæ ...* (1552) to John Rider's *Riders Dictionarie, Corrected and Augmented ...* (1640), as well as biographies of the dictionary-makers, carefully tracing the provenance of the books through manuscript inscriptions. They succeed in establishing that the works come from a variety of collections, of which that owned by the Polish bibliophiles Andrzej (not Andrzej) Stanisław Załuski and his brother Józef Andrzej Załuski was the largest.

Giovanni Iamartino's article, "A Hundred Visions and Revisions: Malone's Annotations to Johnson's *Dictionary*," deals with a copy of Samuel Johnson's *Dictionary of the English language* (1755) annotated throughout by the Shakespearean scholar Edmond Malone. Iamartino shows a selection of over five hundred annotations researched thus far in four main categories: new entries and new definitions, added quotations, new or verified etymologies, and miscellaneous notes. The annotations, we are told, need not be perceived as the reflection of Malone's eccentric pastime. On the contrary, they served a practical purpose, which is why many found their way into Henry Todd's revision of Johnson's *Dictionary*. This is an informative account of the value of annotations in updating and improving lexicographical data.

Roderick McConchie's "The Use of 'Mechanical Reasoning': John Quincy and His *Lexicon Physico-Medicum* (1719)" applies still another perspective that helps to bring to light another interesting finding. He focuses, within the area of under-researched English medical lexicography (McConchie 2019: 1), on the extent to which the lexicographer's interest might influence his dictionary, treating Quincy and his passion for Newtonian mechanical principles as a test case. In *Lexicon Physico-Medicum* (1719), McConchie encounters a number of headwords, such as *cohesion*, *energy*, *particle*, *vision*, and *water*, that appear unrelated to medical sciences. He then compares the dictionary with later editions and with two editions of Blancard's *The Physical Dictionary* to verify whether Quincy's innovation exerted an influence on other medical lexicographers.

In "Paratexts and the First Edition of the *Oxford English Dictionary*: 'Content Marketing' in the Nineteenth Century?," Sarah Ogilvie sets out to discover the links between the creation, content, and context of the OED by looking, on the one hand, at the prefaces and dedications ('peritexts') and, on the other, at archival sources ('epitexts'). She skillfully examines the subsequent prefaces to the fascicles and volumes of the OED and dedications, "a window into the historical and cultural setting," revealing that the editors used prefaces to establish the authority of the dictionary in the eyes of its users, whereas the dedications are indicative of the editors' aspiration to attain prestige and power. Ogilvie weaves threads of evidence into a coherent tapestry, tackling the editors' treatment of World Englishes (see Ogilvie 2013), the contributors to the OED, comparisons between the lexical coverage and that of competitor dictionaries, and aspects of the editorial process. Her research indicates that paratexts, while being a mine of information on the 'hidden' history of the OED, aimed to promote the dictionary's comprehensiveness, scholarly rigour, and prestige.

Rebecca Shapiro's contribution, "The 'Wants' of Women: Lexicography and Pedagogy in Seventeenth- and Eighteenth-Century Dictionaries," explores the links between women and works on lexicography and matters educational. Beginning with James Murray's citations on women (his employing women on "various lexicographical projects" is somewhat ambiguous), Shapiro first analyses early modern English dictionaries in relation to women. This is a viable research stance inasmuch as some early lexicographers, including Robert

Cawdrey (1604), William Bullokar (1616), Thomas Blount (1656), and John Kersey (1702), targeted women as their primary readership; they were expected to improve their own knowledge of languages as learners and pass it onto their children as teachers. The seventeenth and eighteenth centuries were a period of burgeoning literacy, which not only granted women better education, but also saw the first female lexicographers, such as Hester Lynch Piozzi and Maria Edgeworth, whose biographies Shapiro also outlines.

In her essay "Claudius Hollyband: A Lexicographer Speaks His Mind," Gabriele Stein, an acclaimed expert on sixteenth-century dictionaries (e.g. Stein 2014), looks at the judgmental descriptions included by Hollyband, a French Huguenot settled in England, in his *Dictionarie French and English* (1593). Based on a meticulously close reading of the dictionary text, which is her particular area of expertise, Stein provides instances of Hollyband's authorial involvement, entries of an autobiographical nature, and subjective opinions ranging from appreciation to disapproval. They are found in entries related mainly to food and drink, sex (including homosexuality), and the Catholic Church. Stein's conclusions are clear: the lexicographer was by no means a neutral recorder of the vocabulary, openly speaking his mind and, in particular, bluntly expressing his criticism. The use of first and second person pronouns are, moreover, manifestations of his "quasi-oral" teaching method.

Seija Tiisala's article reflects on Jacob Serenius's *Dictionarium Anglo-Svethico-Latinum* (1734). Scrutiny of the paratext of this dictionary and its subsequent editions sheds light on the circles in which the clergyman-cum-politician-cum-dictionary-maker moved, and how they may have influenced his thinking. We thus see him among members of the Royal Society, Freemasons, and subscribers to his dictionary. It is primarily the social spectrum of the subscribers that attracts Tiisala's attention. Her analysis shows that Serenius mixed with merchants, industrialists (e.g. Balthasar Leyel, a director of the English East India Company), bankers, diplomats, consuls, artists, clergymen, scholars, and collectors, and he no doubt benefitted from this influential and inspiring network of contacts.

Yukka Tyrkkö's interest in medical lexicography is well known (e.g. Tyrkkö 2013). His essay under the intriguing title "'Weak Shrube or Underwood': The Unlikely Medical Glossator John Woodall and His Glossary" concerns John Woodall, a military surgeon and innovator, adventurer, and businessman. Tyrkkö concentrates on Woodall's biographical sketch before he carries out an analysis of his six surgical manuals and the glossary appended to *The surgions mate* (1617), in the latter case tracing the medical terms to their sources. The comparison of the paratextual features in the 1617 and the 1639 editions of the book suggests convincingly that Woodall, not a man of letters, became a medical glossator and writer by a sheer accident.

Ruxandra Vişan's article "A 'Florid' Preface about 'a Language That is Very Short, Concise and Sententious'" is the last contribution in this volume. Her narrative is structured around the preface to the second edition of Nathan Bailey's

Dictionarium Britannicum (1736) and its acknowledged and hitherto unacknowledged sources, including Dominique Bouhours's *Les Entretiens d'Artiste et d'Eugène* (1671). By comparing the paratexts and the entries in the dictionaries under analysis, Vişan illustrates how the old lexicographical material was successfully reintegrated, by decontextualisation, recontextualization, and "cultural transplantation," into new material. This is an illuminating essay indicating the anticipated, but never before so explicitly articulated (cf. Considine 2018: 508), creative uses and re-uses of linguistic and lexicographical data in eighteenth-century English dictionary-making.

To conclude, it is hoped that this collection of essays raises questions and encourages "historians of lexicography to examine the paratextual matter of dictionaries from new angles" (McConchie and Tyrkkö 2018: xi). In my opinion, this hope is well-founded. The volume is original in its methodological approach and provides a useful model against which other studies of dictionaries, techniques of dictionary compilation, and biographies of dictionary-makers and their social networks may be conducted in the future in order to investigate previously neglected areas. The content, together with the stylistic coherence and terminological consistency, for which the Editors deserve special praise, makes it a must-have for anyone interested in the complex interplay between lexicography and socio-cultural, political, and economic factors. The only matter remaining to be dealt with is that of a higher resolution which would render the images more readily readable.

Endnotes

1. 'Paratext' may be regarded as an umbrella term, as it refers here both to the internal history of dictionaries that focuses on the textual or visual elements in addition to the main dictionary text, and the external history that focuses on the socio-cultural background in which the dictionaries were published, marketed, and sold.
2. These include the first British edition (London: John W. Parker and Son, 1859, 232 pp.), the first American edition (New York: Redfield, 1859, 218 pp.), the second British edition (London: John W. Parker and Son, 1859, 218 pp.), the third British edition (London: Macmillan, 1865, 229 pp.), the fourth British edition (London/Cambridge: Macmillan, 1873, 275 pp.), the fifth British edition (London: Macmillan, 1879, 309 pp.), and a later British edition (London: G. Routledge and Sons, 1906, 230 pp.). They are available from the Internet Archive at archive.org, Google Books at books.google.com, or HathiTrust at hathitrust.org.

References

- Akasu, K. 2013. Methods in Dictionary Criticism. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 48-61. London: Bloomsbury.
- Considine, J. 2018. Review of: Shapiro, R. (Ed.). 2017. *Fixing Babel: An Historical Anthology of Applied Lexicography*. *International Journal of Lexicography* 31(4): 507-511.

- Dolezal, F.** 1985. *Forgotten but Important Lexicographers: John Wilkins and William Lloyd. A Modern Approach to Lexicography before Johnson*. Tübingen: Max Niemeyer.
- McConchie, R.** 2019. *Discovery in Haste. English Medical Dictionaries and Lexicographers 1547 to 1796*. Berlin: De Gruyter.
- Ogilvie, S.** 2013. *Words of the World. A Global History of the Oxford English Dictionary*. Cambridge: Cambridge University Press.
- Podhajecka, M.** 2018. History in Lexicography and Lexicography in History: A Reappraisal. *Dictionaries: Journal of the Dictionary Society of North America* 39(1): 131-165.
- Stein, G.** 2014. *Sir Thomas Elyot as Lexicographer*. Oxford: Oxford University Press.
- Swanepoel, P.** 2017. Methods in Dictionary Criticism. Bielińska, M. and S.J. Schierholz (Eds.). 2017. *Wörterbuchkritik. Dictionary Criticism*: 81-112. Berlin: Walter de Gruyter.
- Tyrkkö, Y.** 2013. 'My Intent is Onelie to Further Those That Be Willing to Learne': The Lexicon of Mid-Sixteenth-Century Surgical Books in Context. McConchie, R.W. et al. (Eds.). 2013. *Selected Proceedings of the 2012 Symposium on the New Approaches in English Historical Lexis (HEL-LEX3)*: 177-188. Somerville, MA: Cascadilla Proceedings Project.
- Van Male, T.** 2004. From Incipit to Iconophor. Coleman, J. and A. McDermott (Eds.). 2004. *Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology at the University of Leicester, 2002*: 95-108. Tübingen: Max Niemeyer.

Mirosława Podhajecka
University of Opole
Poland
(mpodhajecka@uni.opole.pl)

Publikasieaankondigings / Publication Announcements

Afrikaans National Lexicography Unit (Woordeboek van die Afrikaanse Taal) and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Afrikaans*. 2018, 148 pp. ISBN 9781928326076. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-afrikaans-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

Mariëtta Alberts. *Terminologie en terminografie. 'n Handleiding*. 2019, xviii + 429 pp. ISBN 978-0-620-83804-7 (Sagteband). Pretoria: Die SA Akademie vir Wetenskap en Kuns. Prys: R200.

W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Vyftiende Deel: SKOOL–SRI LANKAANS*. 2019, xxii + 724 pp. ISBN-13 978-0-9946528-6-7 (leerband), ISBN-13 978-0-9946528-5-0 (plastiekband). Stellenbosch: Buro van die WAT. Prys: R800 (leerband) / R450 (plastiekband).

María José Domínguez Vázquez, Mónica Mirazo Balsa, Carlos Valcárcel Riveiro (Editors). *Studies on Multilingual Lexicography*. 2019, c. 240 pp. ISBN 978-3-11-060467-2, ISSN 0175-9264. Lexicographica. Series Maior 157. Berlin/Boston: Walter de Gruyter. <https://www.degruyter.com/view/product/538679>. Price: €99.95.

Volker Harm, Anja Lobenstein-Reichmann, Gerhard Diehl (Herausgeber/Editors). *Wortwelten. Lexikographie, Historische Semantik und Kulturwissenschaft*. 2019, vi + 301 pp. ISBN 978-3-11-063212-5, ISSN 0175-9264. Lexicographica. Series Maior 155. Berlin/Boston: Walter de Gruyter. <https://www.degruyter.com/view/product/517068>. Price: €99.95.

isiNdebele National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–isiNdebele*. 2018, 148 pp. ISBN 9780994660145. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-isindebele-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

isiXhosa National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–isiXhosa*. 2018, 148 pp. ISBN 9780994660152. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-isixhosa-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

isiZulu National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–isiZulu*. 2018, 148 pp. ISBN 9780994660121. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-isizulu-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

Khrystyna Lettner. *Zur Theorie des lexikographischen Beispiels. Die Beispielangaben in der ein- und zweisprachigen pädagogischen Lexikographie des Deutschen*. 2019, ix + 429 pp. ISBN 978-3-11-062786-2, ISSN 0175-9264. Lexicographica. Series Maior 158. Berlin/Boston: Walter de Gruyter. <https://www.degruyter.com/view/product/540820>. Price: €109.95.

Roderick McConchie. *Discovery in Haste. English Medical Dictionaries and Lexicographers 1547 to 1796*. 2019, ix + 226 pp. ISBN 978-3-11-063578-2, ISSN 0175-9264. Lexicographica. Series Maior 156. Berlin/Boston: Walter de Gruyter. <https://www.degruyter.com/view/product/517680>. Price: €99.95.

Roderick McConchie and Jukka Tyrkkö (Editors). *Historical Dictionaries in Their Paratextual Context*. 2018, xii + 318 pp. ISBN: 978-3-11-057286-5 (Hardback). Lexicographica. Series Maior 153. Berlin/Boston: Walter de Gruyter. Price: €99.95. (Review in this issue.)

Marné Pienaar, Eleanor Cornelius. *Interpreting Terminology: Terminologie van het tolken / Tolkterminologie / Mareo a botoloki / Amatemu okutolika*. 2018, xv + 88 pp. ISBN 978-1-928314-40-0. Stellenbosch: SUN MeDIA. <https://doi.org/10.18820/9781928314417>. Prys: R220.

Uwe Quasthoff, Sabine Fiedler, Erla Hallsteinsdóttir (Editors/Redakteurs). *Frequency Dictionary Afrikaans/Frekwensiewoordeboek van Afrikaans*. 2019, 113 pp. ISBN 978-3-96023-249-0 (Paperback/Sagteband). Frequency Dictionaries — Häufigkeitswörterbücher Band 13. Leipzig: Leipziger Universitätsverlag. Price: €22.

Mary Reynolds (Editor-in-Chief). *Oxford South African School Dictionary*. Fourth edition. 2019, 772 pp. ISBN 978 0 19 073180 9 (Paperback), ISBN 978 0 19 073236 3 (Hardback). Cape Town: Oxford University Press Southern Africa. Price: R144.95 (Paperback) / R235.95 (Hardback).

Sesotho National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Sesotho*. 2018, 148 pp. ISBN 9781928326014. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-sesotho-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

Sesotho sa Leboa National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Sesotho sa Leboa*. 2018, 148 pp. ISBN 9781928326007. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-sesotho-official-sa-leboa-foundation-phase-caps-picture-dictionary/>. Price: R160.

Setswana National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Setswana*. 2018, 148 pp. ISBN 9781928326021. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-setswana-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

Siswati National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Siswati*. 2018, 148 pp. ISBN 9780994660169. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-siswati-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

Tshivenda National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Tshivenda*. 2018, 148 pp. ISBN 9780994660190. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-tshivenda-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

Herbert Ernst Wiegand, Rufus H. Gouws, Matthias Kammerer, Michael Mann und/and Werner Wolski (Herausgeber und Bearbeiter/Editors and Compilers). *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtex te und Definitionen sowie Äquivalenten in neun Sprachen / With English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages*. 3. Bd. / Vol. 3. 2019, xxv + 1020 pp. ISBN 978-3-11-061874-7. Berlin/New York: Walter de Gruyter. Preis:/Price: €399.

Herbert Ernst Wiegand, Rufus H. Gouws, Matthias Kammerer, Michael Mann und/and Werner Wolski (Herausgeber und Bearbeiter/Editors and Compilers). *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtex te und Definitionen sowie Äquivalenten in neun Sprachen / With English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages*. 4. Bd. / Vol. 4. 2019, xv + 1010 pp. ISBN 978-3-11-061874-7. Berlin/New York: Walter de Gruyter. Preis:/Price: €399.

Herbert Ernst Wiegand, Rufus H. Gouws, Matthias Kammerer, Michael Mann und/and Werner Wolski (Herausgeber und Bearbeiter/Editors and Compilers). *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen / With English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages*. 5. Bd. / Vol. 5. 2019, xv + 928 pp. ISBN 978-3-11-065202-4. Berlin/New York: Walter de Gruyter. Preis:/Price: €399.

Xitsonga National Lexicography Unit and the South African National Lexicography Units (Authors and Editors). *The Official Foundation Phase CAPS Linked Picture Dictionary English–Xitsonga*. 2018, 148 pp. ISBN 9780994660176. Johannesburg: South African National Lexicography Units (SANLU). <https://sanlu.africa/product/english-xitsonga-official-foundation-phase-caps-picture-dictionary/>. Price: R160.

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 engeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evaluerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksiko-opname, Leksikobibliografie, Leksikonuus, Lexikofokus, Leksiko-eerbewys, Leksikohuldeblyk, Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. inkomstegenererende navorsingstydskrif goedgekeur. Dit verskyn ook op die *Institute of Science Index* (ISI).

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

Manuskripte word vir publikasie oorweeg met dien verstande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir

eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies. Indien van toepassing, moet besonderhede van die oorsprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

6. Gratis oordrukke en eksemplare

Lexikos is sedert volume 28 slegs elektronies beskikbaar op <http://lexikos.journals.ac.za>. Geen oordrukke of eksemplare is dus beskikbaar nie.

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 STELENBOSCH, Republiek van Suid-Afrika.

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sageware 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 engeen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n vertaling van die oorspronklike titel bevat.

2. Grafika

Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

4. Aantekeninge/voetnote/eindnote

Aantekeninge moet deurlopend in die vorm van boskripte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website: <http://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 Woordboek 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

Lexikos is only available electronically on <http://lexikos.journals.ac.za> from volume 28 onward. No offprints or copies are available.

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**.