

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/32-1-1755> (Volle uitgawe / Full volume)

Lexikos 32

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/32-1-1755> (Volle uitgawe / Full volume)

Lexikos 32

Redakteur

Editor

Dion Nkomo

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 32:2022

AFRILEX SERIES 32:2022



BURO VAN DIE WAT

STELLENBOSCH

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/32-1-1755> (Volle uitgawe / Full volume)

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2022 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2022

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

ISBN 978-1-990998-48-5
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

Asian Digital Library; Arts and Humanities Citation Index®, Current Contents®/ Arts & Humanities, Current Contents®/Social and Behavioral Sciences; ERIH Plus; EuroPub Index; 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

Mr. André H. du Plessis (RSA)
Dr. H.S. Ndinga-Koumba-Binza (RSA en Gaboen/RSA and Gabon)
Prof. Dion Nkomo (RSA en Zimbabwe/RSA and Zimbabwe)
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. T.J. Otlogetswe (Botswana)
Prof. A.N. Otto (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	ix
Foreword	xii
<i>Dion Nkomo</i>	
'n Woord van AFRILEX	xiv
A Few Words from AFRILEX	xvi
<i>Langa Khumalo</i>	
Redaksionele doelstellings	xvii
Editorial Objectives	xviii

Artikels / Articles

An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years	162
<i>Esra M. Abdelzaher</i>	
Stigmatised Dictionaries Housing a Stigmatised Variety of English: The Use of Korean English Online Dictionaries as a Teaching Tool within the EFL Classroom	250
<i>Alex Baratta</i>	
Die Behandlung der Wortarten in Wörterbüchern zur Lexikographie	200
<i>Monika Bielińska</i>	
'n Prototipe-aanlyn Afrikaanse akademiese redigeerwoordeboek	301
<i>Maret Blom de Wet</i>	
The Influence of Grammatical Gender on the Sequence of Near-synonyms in Serbian Dictionaries in Contrast to English Thesauri	272
<i>Dragana Čarapić</i>	
The Effect of Learning Conditions on Collocation Gains: A Case Study of Task-based Dictionary Use Instruction	1
<i>Yuzhen Chen</i>	

The Mental Lexicon in Lexicography: The Dictionarios Valladolid-UVa <i>Pedro A. Fuertes-Olivera</i>	118
Differentiated Treatment of Cultural Items in Lexicographical Products: A Necessary Adaptation to the Digital Environment <i>Qian Li and Sven Tarp</i>	90
Les débuts de la lexicographie multilingue roumaine: ressorts prag- matiques et influences culturelles <i>Mihaela Mocanu</i>	31
A Comparative Analysis of Piotr Borkowski's (1963) and Roman Gajda's (1970) English–Polish Phraseological Dictionaries: Practice vs. Theory <i>Mirosława Podhajecka</i>	326
After the Digital Revolution: Dictionary Preferences of English Majors at a European University <i>Bartosz Ptasznik</i>	220
Teksgebaseerde tesourusgebruik in 'n Afrikaanse taalonderrig- konteks <i>Michele F. van der Merwe</i>	180
Who is Sharpest at Looking Up <i>sharp</i> ? Comparing Two Parallel Groups of Dictionary Users <i>Marjeta Vrbinc, Donna M.T.Cr. Farina and Alenka Vrbinc</i>	368
Corpus-based Headword Selection Procedures for LSP Word Lists and LSP Dictionaries <i>Milica Vuković-Stamatović and Branka Živković</i>	141
Improving the Compilation of English–Chinese Children's Dictionaries: A Children's Cognitive Perspective <i>Dan Zhao and Hai Xu</i>	49

Projekte / Projects

Popma's Monolingual Dictionary of Latin Synonyms <i>Nikita Murzintcev</i>	357
--	-----

Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology <i>Sven Tarp</i>	66
--	----

Resensies / Reviews

Howard Jackson (Ed.). <i>The Bloomsbury Handbook of Lexicography</i> <i>Lorna Hiles Morris</i>	364
---	-----

Mariusz Piotr Kamiński. <i>Defining with Simple Vocabulary in English Dictionaries</i> <i>Thomas Herbst</i>	88
--	----

Clifford Simango. <i>Concise Chindau Dictionary</i> <i>Emmanuel Sithole</i>	392
--	-----

Heming Yong. <i>English Lexicography from British Tradition to World Englishes</i> <i>Yongfang Feng</i>	398
--	-----

Publikasieaankondigings / Publication Announcements 404

Voorskrifte aan Skrywers	405
Instructions to Authors	406

Voorwoord

Lexikos 32 verskyn in die nadraai van die COVID-19-pandemie. Tydens daardie tydperk van globale krisis het alles darem nie verlore gegaan nie, ten minste nie ten opsigte van die innovasies wat die mensdom ontwikkel het om by die realiteit van die pandemie aan te pas nie. Terwyl praktiese leksikografiese werk onafwendbaar ongerieflike vertraging moes verduur, het leksikografiese navorsing en beskouings nie tot 'n algehele stilstand gekom nie. *Lexikos* het steeds 'n groot aantal manuskripte ontvang. Ongelukkig kon almal nie vir publikasie in hierdie volume gefinaliseer word nie. Sommige word tans deur die outeurs gefinaliseer of hersien, terwyl ander nog in die anonieme keuringsproses verkeer. Met ons verfynde deurlopende aanlynmodel sal die voltooide artikels steeds vroeg volgende jaar gepubliseer word, sonder dat daar op die finalisering van *Lexikos* 33 einde 2023 gewag hoef te word.

Soos vir die res van die mensdom, was dit vir leksikograwe ook moontlik om normaliteit in hul aktiwiteite te hervat, soos geïllustreer deur vanjaar se van-aangesig-tot-aangesig-konferensies van die internasionale leksikografiese verenigings soos AFRILEX en EURALEX. Weens die pandemie is die 25ste Internasionale AFRILEX-konferensie wat aanvanklik geskeduleer was om die assosiasie se silwerjubileumviering by die Stellenbosch Universiteit te herdenk, in 2021 virtueel aangebied. AFRILEX-lede en -vriende kon uiteindelik in Junie 2022 vir die 26ste Internasionale Konferensie by die Stellenbosch Universiteit bymekaar kom. Daar word met groot genoeë kennis geneem daarvan dat drie van die artikels in hierdie volume hersiene weergawes is van referate wat in Stellenbosch gelewer is.

Die bydraes wat in hierdie volume gepubliseer is, weerspieël as geheel diversiteit as kenmerk van *Lexikos*. Sommige artikels werp lig op die leksikografiese tradisies van sekere tale en gemeenskappe. Ander fokus op die verskillende benaderings tot leksikografiese take, insluitend die bydraes van korpora en die algehele impak van tegnologie op die praktiese leksikografie. In *Lexikos* 32 kan "die omwenteling" in die leksikografie waargeneem word wat ons laat besin oor hoe die woordeboekmaakproses, woordeboekgebruik en die totale dissipline "ná die digitale revolusie" sal lyk, om die titels van twee van die artikels hier aan te haal. *Lexikos* gaan voort om as platform te dien vir oorsigtelike, beskouende, en transformerende vakkundigheid waarvan die kwaliteit die tydskrif se internasionale status bevestig. Ons strewe is om voort te gaan om globaal betrokke te wees by leksikografiese vraagstukke terwyl ons steeds plaaslik relevant bly om lig te werp op die leksikografiese praktyk op die Afrika-kontinent. Die verteenwoordiging en deelname van AFRILEX by Globalex illustreer aan die een kant die globale verbintenis, aangesien Globalex samewerking tussen die redakteurs van verskillende leksikografiese tydskrifte fasiliteer. Aan die ander kant: AFRILEX se volgehoue betrokkenheid by die Nasionale Leksiko-

grafie-eenhede en uitgewers in Suid-Afrika met sessies van woordeboekuitgewers by hul konferensies, sowel as die toekomstige werkswinkels in lande soos Kenia en Zimbabwe as deel van die vereniging se lidmaatskapsveldtog, bevestig die begeerte na kontinentale geworteldheid, wat hopelik in toekomstige *Lexikos*-volumes weerspieël sal word, soos wat ook die norm in die eerste twee dekades van die tydskrif was.

Die jaar 2022 is ook van historiese belang vir *Lexikos* aangesien hierdie volume aangevul word deur twee spesiale uitgawes. Die eerste spesiale uitgawe, *Lexikos* 32(2), is 'n huldeblyk aan die legendariese prof. Danie Prinsloo, wat die leksikografiese gemeenskap met groot onderskeiding gedien het as een van die stigterslede van AFRILEX, 'n lang dienende lid en President van die AFRILEX-raad, 'n produktiewe *Lexikos*-bydraer en redakteur, om nie eens sy bydraes tot die globale arena te noem nie. Die tweede spesiale uitgawe word tans gefinaliseer en sal kort ná hierdie gewone uitgawe verskyn. Dit is slegs in 2009 dat die gewone uitgawe van *Lexikos* aangevul is deur 'n spesiale uitgawe. Hierdie spesiale uitgawes sal in die toekoms 'n kenmerk van *Lexikos* word om buitengewone leksikografiese bydraes en individue te gedenk en te vier.

Soos gebruiklik vir *Lexikos* is die suksesvolle voltooiing van hierdie volume te danke aan die wonderlike spanwerk aan hierdie tydskrif. As die hoofredakteur van hierdie volume, het ek deurentyd staatgemaak op die steun van die kollegas waaruit die span van roterende redakteurs saamgestel is, aan wie ek veel verskuldig is. Die eerste aantal artikels van *Lexikos* 32 wat in April aanlyn gepubliseer is, is van verlede jaar oorgedra en is in geheel deur prof. Taljard, die redakteur van *Lexikos* 31, hanteer. Saam met mnr. André du Plessis, was prof. Taljard ook verantwoordelik vir die Afrikaanse artikels in hierdie volume, soos wat dr. Hugues Steve Ndinga-Koumba-Binza verantwoordelik was vir die Franse bydraes. Mnr. Du Plessis is nie 'n nuweling in leksikografiese en redaksionele werk nie. Hy is deel van 'n sterk redaksionele span van die Buro van die Woordeboek van die Afrikaanse Taal (WAT) en was reeds voorheen, maar vanjaar tot 'n baie groter mate, betrokke by die werk aan *Lexikos*.

Die keurders van manuskripte wat aan *Lexikos* voorgelê is, te veel om by die naam te noem en oral oor die wêreld heen versprei, het voortgegaan om hul tyd op te offer en hul gesaghebbende kennis te verskaf om sodoende die kwaliteit van die tydskrif te verhoog. Ek wil tog vir dr. Gertrud Faaß (Universiteit van Hildesheim) en dr. Natasha Engelbrecht (Rhodes Universiteit) uitsonder vir die aanbied van hul kundigheid in Duits. 'n Soortgelyke erkenning is aan dr. Daniel Melamis (Rhodes Universiteit) verskuldig vir sy bydrae ten opsigte van kundigheid in Latyn. Geen redakteur kan portuurbeoordelaars genoeg bedank vir hul onskatbare bydraes nie.

Hierdie voorwoord, soos die hele *Lexikos*-volume, sal onvolledig wees sonder 'n woord van waardering aan me. Tanja Harteveld en me. Hermien van der Westhuizen. Oor die jare was hulle die steunpilare in die produksie van ons tydskrif. Hul betrokkenheid by outeurs en die volgehoue professionele steun wat hulle aan die redakteurs verskaf, speel 'n integrale rol daarin om

Lexikos 'n aantreklike tydskrif vir beide outeurs en lesers te maak. Hulle doen al die bogenoemde met onberispelike grasiae.

Laastens, 'n groot dankie aan die outeurs wat *Lexikos* as platform kies om jul buitengewone vakkundigheid te deel, vir jul toegewyde verbondenheid tot anonieme terugvoer en vir jul geduld gedurende die anonieme keuringsproses tot nou. Laat ons voortwerk na die volgende volume!

Dion Nkomo
Redakteur

Foreword

Lexikos 32 comes in the aftermath of the COVID-19 pandemic. During that period of global crisis, not all was lost, not least in terms of innovations that humanity developed to adapt to the reality of the pandemic. While practical lexicographic work inevitably suffered inconvenient delays, lexicographic research and reflections did not come to a total halt. *Lexikos* continued to receive a high volume of manuscripts. Unfortunately, not all of them could be finalized for publication in this volume. Some are being finalized or revised by the authors while others remain under peer review. With our refined publish-as-we-go model, the finished articles will still be published early next year, without having to wait for the finalization of *Lexikos* 33 towards the end of 2023.

Like the rest of humanity, lexicographers were able to resume normalcy in their activities, as illustrated by this year's face-to-face conferences of the international lexicographical associations such as AFRILEX and EURALEX. Due to the pandemic, the 25th International conference of AFRILEX initially scheduled to commemorate the association's silver jubilee at Stellenbosch University in 2020 was held virtually in 2021. AFRILEX members and friends were eventually able to converge at Stellenbosch University for the 26th International Conference in June 2022. It is with great delight to note that three of the articles in this volume are revised versions of the papers that were presented at Stellenbosch.

Overall, the contributions published in this volume typify diversity as the hallmark of *Lexikos*. Some articles reflect on lexicographic traditions of certain languages and communities. Others reflect on different approaches to lexicographic tasks, including the contributions of corpora and the overall impact of technology on practical lexicography. In *Lexikos* 32, we see lexicography being "turned upside down", which makes us ponder how the dictionary making, dictionary use and the entire discipline would look like "after the digital revolution", to quote the titles of two of the articles. *Lexikos* continues to be a platform for reflective, contemplative, and transformative scholarship whose quality affirms the journal's international standing. Our endeavour is to continue to be globally engaged in lexicographic matters while also remaining locally relevant in terms of illuminating lexicographic practice on the African continent. The representation and participation of AFRILEX on Globalex, on the one hand, illustrates the global engagement, as Globalex facilitates cooperation between the editors of different lexicography journals. On the other hand, AFRILEX's continued engagements with the National Lexicography Units and publishers in South Africa through dictionary publishers' sessions at its conferences, as well as its forthcoming workshops in countries such as Kenya and Zimbabwe as part of the association's membership drive, affirm the desire for continental rootedness, which is hoped to reflect in future *Lexikos* volumes, as was the norm in the first two decades of the journal.

The year 2022 is also historic for *Lexikos* in that this volume is accompanied by two special issues. The first special issue, *Lexikos* 32(2), is a festschrift for the legendary Prof. Danie Prinsloo who served the lexicographic community with great distinction as one of the founding elders of AFRILEX, a long-time member and President of the AFRILEX Board, a prolific *Lexikos* contributor and editor, not to mention his contributions on the global arena. The second special issue is being finalized and will appear shortly after this regular issue. It is only in 2009 that the regular issue of *Lexikos* has been supplemented by a special issue. Such special issues will feature in the future of *Lexikos* to mark and celebrate exceptional lexicographic contributions and individuals.

As usual for *Lexikos*, the successful completion of this volume is thanks to the journal's great teamwork. As the main editor for this volume, I consistently relied on the support of colleagues constituting the team of revolving editors to whom I am greatly indebted. The first batch of *Lexikos* 32 articles which were published online in April were carried over from last year and wholly handled by Prof. Taljard, the editor of *Lexikos* 31. Together with Mr André du Plessis, Prof. Taljard also took responsibility with the Afrikaans articles in this volume, as did Dr Hugues Steve Ndinga-Koumba-Binza regarding French contributions. Mr Du Plessis is not new to lexicographical and editorial work. He is part of a strong Bureau of the Woordeboek van die Afrikaanse Taal (WAT) editorial team and has already been involved with *Lexikos* work, more significantly, this year.

The adjudicators of manuscripts submitted to *Lexikos*, too many to mention by names and spread all over the world, continued to sacrifice their time and provide their expert knowledge, thereby enhancing the journal's quality. However, I would like to single out Dr Gertrud Faaß (University of Hildesheim) and Dr Natasha Engelbrecht (Rhodes University) for offering their expertise in German. A similar acknowledgement is due to Dr Daniel Melamis (Rhodes University) for his expertise in Latin. No editor will ever thank peer adjudicators enough for their priceless contributions.

This foreword, just like the entire *Lexikos* volume, would be incomplete without a word of appreciation to Ms Tanja Harteveld and Ms Hermien van der Westhuizen. Over the years, they have been the mainstay in the production line of our journal. Their engagement with authors and the professional support they continue to provide to the editors play an integral role in making *Lexikos* an attractive journal for both authors and readers. They do all this with impeccable grace.

Finally, a big thank-you to the authors for choosing *Lexikos* as a platform for sharing your outstanding scholarship, for your diligent engagement with peer feedback and for your patience during the peer review process until now. Let us work towards the next volume!

Dion Nkomo
Editor

'n Woord van AFRILEX

Die wêreld herstel steeds van die nagevolge van die COVID-19-pandemie wat tydens die hoogtepunt daarvan gelei het tot 'n omvattende sluiting van byna alle menslike aktiwiteite. Vir die African Association for Lexicography (AFRILEX) het dit veroorsaak dat die Assosiasie hul 25ste jaarlikse internasionale konferensie in 2020 moes kanselleer, en dat hulle gevolglik die Assosiasie se eerste volledig virtuele internasionale konferensie in 2021 aangebied het. Met vanjaar se konferensie is teruggekeer na die (nuwe) normaal, met die eerste internasionale post-COVID-19 globale pandemie-konferensie wat weer in persoon by Stellenbosch Universiteit plaasgevind het. Dit was verfrissend om te sien dat die AFRILEX-familie van die uithoeke van die wêreld by die tuiste van die Suid-Afrikaanse leksikografie byeengekom het.

Die 26ste AFRILEX jaarlikse internasionale konferensie by Stellenbosch Universiteit was 'n heuglike geleentheid, waar die AFRILEX-raad en die hele AFRILEX-familie 'n gepaste eerbetoon aan die lang dienende lid en President van die AFRILEX-raad, professor Danie Prinsloo, bewys het. Die AFRILEX-raad het 'n verrassingsbekendstelling gemaak van die spesiale uitgawe, *Lexikos* 32(2), 'n huldeblyk aan prof. Danie as stigterslid van AFRILEX, 'n doyen van die teoretiese en korpusleksikografie, en 'n gerekende vakkundige wat die leksikografiese gemeenskap plaaslik en globaal met soveel toewyding en groot onderskeiding gedien het.

Die 26ste AFRILEX jaarlikse internasionale konferensie het baie interessante referate opgelewer, sommige waarvan op rekenaarmatige benaderings tot die leksikografie gefokus het en wat nuwe stelsels en oplossings binne die raamwerk van die Vierde Industriële Revolusie verskaf het. Dit is opmerklik dat sommige van hierdie interessante referate tot wetenskaplike artikels verwert is wat in hierdie *Lexikos*-volume verskyn. Die tydskrif gaan voort om 'n groot aantal bydraes van plaaslike sowel as internasionale outeurs te lok soos weereens in hierdie volume gesien kan word. Ten spyte van die voor-die-handliggende uitdagings van 'n stadige omkeertyd vir portuurbeoordeelde keuringsverslae, het die *Lexikos*-redaksie nog 'n uitgawe van die tydskrif voltooi, om sodoende hierdie belangrike kanaal van akademiese diens oop te hou. Daarom is besondere dank verskuldig aan die redakteur, professor Dion Nkomo, van Rhodes Universiteit, en die resensieredakteur, me. Tanja Harteveld, asook me. Hermien van der Westhuizen, vir die tegniese ondersteuning.

Die belangrike rol wat die Buro van die WAT oor die jare as uitgewer van *Lexikos* vertolk het, moet weer eens erkenning geniet. Die Buro van die WAT bly steeds 'n sleutel- en strategiese vennoot van AFRILEX.

Namens die AFRILEX-raad, asook die algemene lede van AFRILEX, bedank ek die redaksiespan, die Buro van die WAT en die bydraende outeurs van harte

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/32-1-1755> (Volle uitgawe / Full volume)

vir nommer 32 van *Lexikos*. Dit is nog 'n indrukwekkende bydrae tot die immer-groeiende kennisekonomie.

Langa Khumalo
President: AFRILEX

A Few Words from AFRILEX

The world is recovering from the aftermath of the COVID-19 global pandemic, which at its peak resulted in a massive lockdown of most human activities. For the African Association for Lexicography (AFRILEX) this resulted in the Association's inability to host its 25th annual international conference in 2020, and the subsequent hosting of the Association's first fully virtual international conference in 2021. This year's conference saw the return to the (new) normal, with the first post-COVID-19 global pandemic in-person international conference taking place at Stellenbosch University. It was refreshing to see the AFRILEX family from all corners of the globe converge at the home of South African lexicography.

The 26th AFRILEX annual international conference at Stellenbosch University was a fittingly auspicious occasion, where the AFRILEX Board and the entire AFRILEX family bestowed a befitting honor to its long-time member and President of the AFRILEX Board, Professor Danie Prinsloo. The AFRILEX Board had a surprise launch of the special issue, *Lexikos* 32(2), a festschrift in honor of Prof. Danie as a founding member of AFRILEX, a doyen of theoretical and corpus lexicography, and a distinguished scholar who has served the lexicographic community locally and globally with such dedication and great distinction.

The 26th AFRILEX annual international conference had very engaging papers, some of them focusing on computational approaches to lexicography, providing novel systems and solutions in the context of our present reality of the Fourth Industrial Revolution. It is notable that some of these engaging presentations were converted into scientific articles that are published in this *Lexikos* volume. The journal continues to attract a large number of contributions from both local and international authors as again is shown in this volume. Despite the obvious challenges of slow turnaround of peer reviews, the *Lexikos* editorial team has successfully concluded another issue of the journal, thereby maintaining this important channel of academic service. To this end, I want to especially recognise the effort and tenacity of the editor, Professor Dion Nkomo, of Rhodes University and the review editor, Ms Tanja Harteveld, as well as Ms Hermien van der Westhuizen, for technical support.

I want to also acknowledge the important role that the Bureau of the WAT has played over the years as the publisher of *Lexikos*. The Bureau of the WAT remains a key and strategic partner of AFRILEX.

On behalf of the AFRILEX Board, and indeed the general members of AFRILEX, I sincerely thank the editorial team, the Bureau of the WAT and the contributing authors for volume 32 of *Lexikos*. It is another important contribution to the ever-growing knowledge economy.

Langa Khumalo
President: AFRILEX

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

The Effect of Learning Conditions on Collocation Gains: A Case Study of Task-based Dictionary Use Instruction

Yuzhen Chen, *College of Foreign Languages,
Putian University, Fujian, P.R.C. (287323222@qq.com)*

Abstract: To examine the effect of learning conditions on collocation gains, 88 Chinese EFL students were assigned randomly to one of three different learning conditions, i.e. dictionary use with prior instruction, dictionary use without training, or explicit collocation teaching. They were asked to fill in the missing verb in ten V + N target collocations embedded in sentences. A screen recorder was used to keep track of the students' lookup behaviour in the two conditions involving dictionary use such as every input of the searched word, every move and click of the mouse, every step of collocation search and the time length of dictionary consultation. After completion of the assignment all the students were given corrective feedback to the collocation task and marked their corrections on the test paper. Two weeks later, an unexpected retention test was administered. The study showed that dictionary use following a five-step training session reaped significantly more collocation gains than dictionary use without prior instruction and explicit collocation teaching. Compared with those who used the dictionary without prior training, the students with prior dictionary instruction employed more effective strategies concerning the selection of lookup words and the location and identification of collocation information. They took a more conscientious approach to dictionary use and retrieved more correct target verbs. Apart from lending new support to the continuous appeal for a position of dictionary use instruction in the EFL pedagogy, the study also provides a detailed demonstration of task-based dictionary training applicable to classroom practice.

Keywords: LEARNING CONDITIONS, COLLOCATION LEARNING, DICTIONARY SKILLS, DICTIONARY LOOKUP BEHAVIOUR, DICTIONARY USE INSTRUCTION

Opsomming: Die invloed van onderwysomstandighede op die aanleer van kollokasies: 'n Gevallestudie van taakgebaseerde woordeboekgebruiksonder-rig. Om die invloed van onderwysomstandighede op die aanleer van kollokasies te ondersoek, is 88 Chinese EVT-studente lukraak aan een van drie verskillende onderwysomstandighede onderwerp, nl. woordeboekgebruik met voorafgaande onderrig, woordeboekgebruik sonder enige opleiding, of gebruik met uitvoerige aanwysings vir die aanleer van kollokasies. Hulle is ook gevra om die ontbrekende werkwoord in tien V + N-doelkollokasies wat in sinne gebruik is, in te vul. 'n Skermopnemer is gebruik om die studente se naslaangedrag in die twee onderwysomstandighede wat woordeboekgebruik behels, te monitor, bv. elke keer wanneer die woord wat nageslaan word, ingetik word, elke beweging en klik van die muis, elke stap van die kollokasiesoektog en die tyds-

duur waarin die woordeboek geraadpleeg is. Ná voltooiing van die opdrag is al die studente korrektiewe terugvoer rakende die kollokasieopdrag gegee en hulle het hul korreksies op die toetsblad aangebring. Twee weke later is 'n onverwagse retensietoets uitgevoer. Die studie het aangetoon dat woordeboekgebruik wat volg op 'n onderrigsessie bestaande uit vyf stappe 'n baie groter invloed op die aanleer van kollokasies gehad het as woordeboekgebruik sonder enige voorafgaande opleiding en uitvoerige aanwysings vir die aanleer van kollokasies. Indien die studente wat die woordeboek sonder enige opleiding gebruik het, vergelyk word met dié wat voorafgaande onderrig ontvang het, kan gesien word dat laasgenoemde groep effektiewer strategieë rondom die seleksie van naslaanwoorde en die vind en identifisering van kollokasie-inligting toegepas het. Hulle het 'n meer doelgerigte benadering tot woordeboekgebruik gevolg en meer korrekte doelwerkwoorde gevind. Buiten die verlening van nuwe steun aan die volgehoue pleidooi vir woordeboekgebruiksopleiding in die EVT-pedagogie, verskaf hierdie studie ook 'n gedetailleerde voorbeeld van taakgebaseerde woordeboekopleiding wat op die klaskamerpraktyk toegepas kan word.

Slutelwoorde: ONDERWYSOMSTANDIGHEDE, AANLEER VAN KOLLOKASIES, WOORDEBOEKVAARDIGHEDE, WOORDEBOEKNASLAANGEDRAG, WOORDEBOEKGEBRUIKS-ONDERRIG

1. Introduction

As an important indicator of learners' communicative competence and language proficiency, collocation has been a recurrent topic of interest for EFL researchers and teachers. Abundant evidence shows that collocation acquisition is a slow and difficult process which is complicated by various interlexical and intralexical factors (Peters 2016, Boers et al. 2014, Yamashita and Jiang 2010) and that even advanced learners experience difficulties in producing appropriate collocations (Durrant and Schmitt 2009, Li and Schmitt 2010, Laufer and Waldman 2011, Alzi'abi 2017). To assess the role that the dictionary can play in helping EFL learners meet the big challenge of collocation learning, an increasing number of dictionary-based collocation studies have been carried out. Researchers have examined the contribution of dictionary use to collocation reception, production and retention (Laufer 2011, Li and Xu 2015, Alzi'abi 2016, 2017, Chen 2017, 2020), investigated users' lookup behaviour and skills (Komuro 2009, Wu 2011, Lew 2012, Chen 2017), and compared the efficacy of collocation learning between dictionaries of different types, media and presentation modes (Lew and Radłowska 2010, Dziemianko 2020, 2011, 2012, 2014, 2017, Dai et al 2019). In addition, studies have been conducted to explore the approaches to and effectiveness of dictionary skills training for collocational competence development (Kim 2017, 2018, Basal 2019). However, no efforts have been made yet to gauge the effect of learning conditions on collocation learning which involves dictionary use. There is no evidence that dictionary use is more effective than the prevailing approach of explicit collocation teaching in the classroom and vice versa. It has been established that dictionary use can contribute positively to collocation production and retention as compared

with no dictionary use (Laufer 2011, Li and Xu 2015, Chen 2017, Alzi'abi 2017), but it remains unproven whether dictionary use with prior training will necessarily bring about significantly better results than other learning conditions like dictionary use without instruction and explicit collocation teaching. Besides, more research is needed to compare how trained dictionary users behave differently from the untrained during collocation consultation. Although Kim (2017) validated the effectiveness of dictionary training for improving the participants' productive use of collocation, she took a product-based approach, i.e. comparing the participants' performance of collocation tasks before and after dictionary training without investigating their specific use of lookup strategies during a collocation search. More insights would have been gained if a process-oriented approach had also been adopted. In Kim (2018), think-aloud protocols were used to examine the participants' lookup behaviour, but as cited by the same author, there are disadvantages inherent in this data collection method such as overshadowing the non-reportable perceptual and memory processes and participants' being prone to reactivity which can potentially trigger changes in their cognitive processes. Actually, thanks to technological development, nowadays screen recorders are easily available which can be used to keep track of dictionary users' lookup behaviour, including which words are searched, the number of lookups, every move and click of the mouse, what information is used, and the time length of dictionary use. It is believed that screen video recordings can provide a lot of authentic data to show how learners actually use the dictionary. In view of all this, the present study aims to investigate the effect of learning conditions on collocation gains, comparing the collocation retention scores yielded by dictionary use with prior training, dictionary use without training and explicit collocation teaching. It also attempts to compare the lookup behaviour of learners with and without dictionary use instruction by means of screen recording.

2. Defining collocation

Despite a broad consensus on the significance of collocation for fluent and idiomatic use of language, no universal agreement has been reached when it comes to the definition of collocation. Generally speaking, there are two approaches to collocation: the phraseological approach and the frequency-based approach. The former regards collocation as a type of restricted word combination based on semantic criteria (Cowie 1981, Wray 2002, Nesselhauf 2003, Laufer and Waldman 2011) while the latter refers collocation to the regular co-occurrence of words within a given span (Sinclair 1991, Hoey 2005). In the present study, the author adopts the former approach, defining collocation as habitually occurring lexical phrase that is characterized by relative transparency in meaning and form-restricted co-occurrence of elements like *lift a ban*, *hold a record*, *say a prayer* (Laufer and Waldman 2011: 648).

3. Previous dictionary-based collocation studies

3.1 Users' lookup behaviour

A number of researchers have investigated how learners use dictionaries for collocation learning. Wu (2011) conducted a questionnaire survey on Chinese EFL learners' lookup habits and preferences for collocation search. It was found that the participants usually looked for collocation information during writing and translating. Most of them attributed their reluctance to seek collocational information to a lack of time and even if they did, they would encounter such difficulties as confusion about where and how to find the relevant words and how to deal with what they had found. The study also investigated what type of dictionary was used most often for collocation consultation, what type of collocation was searched for most frequently, how the participants chose the keyword in a collocation for dictionary lookup, and how to optimize the presentation of collocation information in the dictionary. Instead of trying to paint a general picture of learners' collocation lookup habits, Lew (2012) focused on one specific lookup strategy, i.e. how learners decided which component word of a multi-word combination to look up in the dictionary. Forty Polish learners were required to select one word from each of 36 multi-word expressions that they would most readily look up. The results showed that word frequency and part of speech were strong predictors, yet the position of the word within the multi-word expression did not make much difference.

To explore learners' ability to retrieve information from a collocation dictionary, Komuro (2009) gave 26 Japanese university students a task to fill in the sentence blank for target collocations. They were found to be more successful in completing adjective–noun collocations (71.5%) than they were for verb–noun collocations (61.2%) and preposition–noun collocations (57.3%). Most of the participants had difficulty in making decisions about which collocate to choose from the near-synonymous collocates listed in the entry.

Research findings show that learners have inadequate dictionary skills. In a study to examine Arab EFL learners' use of electronic dictionaries to judge the appropriateness of collocations, Alzi'abi (2012, cited in Alzi'abi 2017) pointed out that even with dictionary assistance, the participants did poorly on the tests as they were incapable of taking full advantage of the collocational information in dictionaries. Li and Xu (2015) invited 32 Chinese EFL learners to perform a meaning determination task of verbal phrases with an online dictionary. The participants were unable to differentiate meanings of words and were pre-occupied with familiar meaning. They also tended to ignore the use of hyperlinks for cross reference. Similar findings were reported in Chen (2017) which tracked users' consultation behaviour with CALL software. The participants hardly used or even noticed the hyperlink function of the electronic dictionary. They were unable to distinguish between senses of a polysemous word, inclined to choose the sense listed at the beginning of an entry, and apt to lose

patience when faced with overcrowded entry information. Those findings were confirmed again by Chen (2020) through a retrospective questionnaire survey administered to EFL learners who used smartphone dictionaries for a collocation production task. Based on their findings, the researchers (Chen 2017, 2020, Laufer 2011, Li and Xu 2015) advocated inclusion of dictionary use instruction in the EFL pedagogy.

3.2 Dictionaries' contribution to collocation learning

Ample evidence indicates that dictionaries can provide useful assistance in collocation reception, production and retention. Laufer (2011) presented 95 EFL learners with a task to fill in the missing verb in target verb–noun collocations. It turned out that for the intermediate group, the use of the dictionary resulted in an increase of 150% of correct collocations as compared with the pre-test, and the pre-intermediate group improved by 96%. Chen (2017) asked 52 English majors at a Chinese university to complete a collocation fill-in task first dictionary-free and then with a CALL dictionary. Statistics showed that the number of correct verbs for the 12 target collocations increased from 0.25 in the pretest to 5.90 in the test, and in the posttest, 2.13 correct verbs were retained. As further corroborated by Chen (2020) in which 62 English majors completed a ten-item collocation fill-in task with smartphone dictionaries, the use of a dictionary also contributed significantly to the learners' productive knowledge of collocations, increasing from 0.23 correct verbs in the pretest to 5.82 in the test. Despite a significant loss of the target collocations after a week (from 5.82 to 2.00 words), the participants still gained 1.77 words as compared with the pretest.

Li and Xu (2015) examined the use of a dictionary for collocation decoding rather than encoding. The participants were required to complete a meaning determination task, using an online dictionary to choose from the given options the right meaning for the target verbal phrases. A significant difference was identified between the task performance before and after dictionary consultation (26.52% vs 49.69%). Alzi'abi (2012, cited in Alzi'abi 2017) also looked at dictionary use in relation to collocation reception. More than 100 English majors at a Syrian university were instructed to judge the appropriateness of 20 verb–noun combinations. They substantially improved their performance with the aid of an electronic dictionary. In Alzi'abi (2016), the participants were asked to provide three noun collocates to replace the *etc.* used in verb definitions to stand for the nouns that can collocate with the defined verb. Then they were guided to judge, according to the verb definitions, the appropriateness of a set of four noun collocates used with each defined verb. It was found that the participants could only provide 40% correct noun collocates, but they did much better in the judgment test, suggesting a positive role of a dictionary in helping users identify erroneous collocates. The limited help the dictionary offered for the production of collocations was attributed to the lack of clarity in the defini-

tions as well as the use of *etc.* that proved more difficult to decipher (Alzi'abi 2016: 322). It should be noted though, that such a testing form seemed to be too challenging for EFL learners who may have limited vocabulary knowledge in general and collocational knowledge in particular.

In Alzi'abi (2017), 88 MA English majors were instructed to provide, upon dictionary consultation, three additional adverb collocates for each target verb. There was a significant improvement in the participants' performance when they used the dictionary. Nevertheless, only about 10% of the responses were appropriate and the scores were all well below an average level. Actually, as with Alzi'abi (2016), this kind of testing form seemed to be more directly concerned with the participants' collocational knowledge than with the effect of dictionary use.

Compared with its contribution to collocation reception and production, the dictionary seems to play a lesser role in collocation retention. Laufer (2011) found that the intermediate group retained about 1.5 new collocations on average and the pre-intermediate group could only remember about 0.5. Chen (2017) also demonstrated that the participants' collocation retention rate was only 36.1%, close to what was found in Chen (2020), i.e. 34.5%. The low retention scores were actually not surprising, because for one thing, retrieving information successfully with a view to completing a task at hand by no means guarantees that the retrieved information will be stored in long-term memory, and for another, users are not always successful at finding the desired information in the first place (Boers and Lindstromberg 2012: 92). Presumably, the dictionary may be likely to play a greater part in collocation retention if a learning process is involved.

3.3 Skill training in collocation consultation

To date, there is only sporadic empirical research on how to train learners to master collocation lookup skills and how to increase learner autonomy through dictionary use instruction. Among the few studies are Kim (2017, 2018). In Kim (2017), 59 Korean EFL learners were trained to consult an online dictionary for collocation production in the order of node word selection, word sense distinction, collocate type location and feasible collocate identification. A comparison between a pretest and a posttest on collocation production demonstrated that teaching dictionary skills substantially improved learners' ability to produce natural collocations. As indicated by an end-of-semester survey, the participants perceived the instruction as necessary and helpful in gaining collocational competence. More importantly, they began to develop the habit of consulting collocation dictionaries after receiving the instruction. Innovative as it is, the study did not monitor the process of dictionary use. No information was provided about what strategies were adopted for collocation search or how the learners changed their dictionary consultation behaviour. The findings would

be more convincing if the learners' lookup behaviour were observed closely and analyzed in detail.

In her another noteworthy study, Kim (2018) compared four participants' dictionary consultation behaviour while correcting collocation errors before and after dictionary use instruction by using think-aloud protocols. At the first meeting, the participants wrote an essay and then used an online dictionary to correct the wrong collocations sorted out and marked by the researcher. They verbally reported on the cognitive process they went through during the error correction. At the second meeting four weeks later, the participants were given the same four-step dictionary skills training as in Kim (2017) on the selection of a node word, the distinction of senses, the location of collocate types and the identification of possible collocates. After the instruction, the participants performed the same correction task as was given at the first meeting. The study showed that all the four participants made remarkable progress in the collocation correction task, increasing the success rate from 55% at the first meeting to 90% at the second meeting. What was more encouraging was that, the participants significantly changed their dictionary consultation behaviour. The research thus highlighted the pedagogical need to impart collocation lookup skills in L2 writing classes to foster collocational competence and learner autonomy. Kim's research added further evidence about the effectiveness of dictionary training for collocation production, yet it should be pointed out that the number of participants ($n = 4$) was far too small to warrant generalizations. Besides, the participants were highly motivated language learners and were not quite typical of Korean college students in general. Furthermore, as acknowledged by the author, completing the collocation task while thinking aloud may slow down task execution, neglect the non-reportable process or potentially distort the participants' cognitive processes. More authentic data would be collected if a non-intrusive method like screen recording were used.

Basal (2019) carried out an interesting experiment in which the experimental group behaved like novice lexicographers and were guided to create a collaborative online dictionary while learning adjective–noun collocations. By using concordances, the Oxford Online Dictionary, the World Wide Web and Google Docs, 53 EFL learners at a Turkey university were instructed to find from authentic texts the target collocations and appropriate example sentences. It was found that compared with the traditional group who learned the target collocations through exercises and teaching, the experimental group performed significantly better on both the immediate and the delayed posttests. The study confirmed the effectiveness of using online tools for learning collocations. It should be noted that the small number of participants in the experimental group ($N = 28$) limited the generalizability of the results. Nevertheless, the introduction of new technologies and online resources to collocation teaching can present new windows of opportunities for this challenging task.

Alzi'abi (2017) identified a non-significant relationship between training in dictionary use and collocation production performance. Among the 88 partici-

pants, only about 55% had received some kind of dictionary training. Data analysis indicated that basic prior training in dictionary usage did not make any positive impact on the participants' overall performance on verb–adverb collocation production. According to the researcher, the participants probably did not receive proper instructions on the best way to use dictionaries of any type, or they might have been taught how to utilize bilingual dictionaries or possibly paper dictionaries rather than electronic ones. Still, there might be another possible explanation, i.e. the form of the test may not be an effective indicator of dictionary use competence. The participants were asked to give three additional adverbs other than those in the dictionary. This does not directly concern their ability to use the dictionary. Instead, it is more related to their collocational knowledge.

4. The study

Inspired by Kim (2017, 2018), the present study also involved the instruction on collocation lookup skills, but the author took both product-based and process-based approaches, looking at both the performance of a collocation task as a result of dictionary use and the whole process of using the dictionary. Besides, Kim (2017, 2018) focused on the use of a collocation dictionary, which may not be applicable to EFL learners in China who seldom turn to collocation dictionaries but prefer to use general dictionaries for collocation searches. Therefore, the present author devised her own teaching approach based on the use of learners' dictionaries. Another noteworthy point is that, different from most previous studies reviewed in Section 3.2, the present study involved the learning process, i.e. all the participants were given the corrective feedback to the task and asked to rectify their original wrong answers on the test paper.

4.1 Research questions

The author seeks to address the following research questions:

- (1) Do different learning conditions bring about significantly different results of collocation gains? Which learning condition yields the highest retention scores? Here three conditions are involved, i.e. dictionary use with prior training, dictionary use without training and explicit collocation teaching.
- (2) What are the differences in the lookup behaviour between the participants who use the dictionary with prior instruction and those without?

4.2 Participants

The participants of the study were from three parallel classes at a Chinese university. They are the first-year English majors with at least six years of EFL

learning experience and similar linguistic and cultural backgrounds. As measured by one-way ANOVA, there is no significant difference in their score of National English Examination for University Entrance [$F(2, 87) = 1.308, p = 0.276 > 0.05$], indicating that the three classes are at a comparable level of English proficiency. An informal survey done by the author showed that none of the students had received any dictionary training back in high school, and that they had never used any online English monolingual learners' dictionary. Ninety students took part in all the three stages of the study, but two of them who knew three target collocations were sifted out, thus leaving 88 students for the final analysis.

4.3 Design of the collocation task

Target collocations

The study targets at V+N collocations which are difficult for EFL learners to acquire (Durrant and Schmitt 2010, Nesselhauf 2003, Boers et al. 2014). The target collocations were selected according to several criteria. The components of the collocations are high-frequency words which are familiar to the participants but the combinations are assumed to be new to them. The mutual information of each collocation is above 3 as checked through the BNC, which indicates that they are habitually co-occurring pairs. All the target collocations are transparent in meaning without distinctive register features and they are all covered by the dictionary used in the study. Originally, 15 V+N collocations had been chosen. After a brief survey administered to a parallel class of 30 students who had similar English proficiency to the participants of the study, five were ruled out, as they were familiar to some of the students. Thus, only ten collocations remained.

Task design

A collocation task was designed which was composed of ten sentences, each of which contained a V + N target collocation with the verb missing. The participants were asked to provide a verb to complete the target collocation according to its Chinese translation. Gap filling is a common form of learning exercise used in EFL classroom teaching in China. It is also one of the five sections included in the Test for English Majors (TEM, Band 4), a nation-wide test to measure English majors' language proficiency.

All the sentences were carefully modeled on the concordance lines in BNC or the example sentences from learners' dictionaries other than the dictionary used for the present study. They consisted of frequent words that are familiar to the participants and have a stand-alone context with the Chinese explanation of the target collocation to avoid ambiguity, e.g. *The tree ___ a small, bitter fruit.* (结果实) (see the appendix).

4.4 Dictionary used for the study

The dictionary used for the study was the fifth edition of *Longman Dictionary of Contemporary English* (LDOCE5) which can be accessed freely on <https://www.ldoceonline.com>. LDOCE5 presents collocation information in at least four ways. As displayed in Figure 1, collocation patterns are presented in bold in each sense of the headword above an example (or examples) like *pace of, at a steady/slow etc. pace*. Some collocations are integrated in bold into examples like *quicken one's pace, a walking pace* and *take a pace*. Some are treated as a multi-word subsense headword that can be retrieved through hyperlinks, such as *keep pace, set the pace, and force the pace*. In addition, a COLLOCATIONS box shows different types of collocations related to the entry headword, such as adjective–noun collocations, noun–noun collocations, verb–noun collocations and other phrases. Generally, LDOCE5 offers rich and easily accessible collocation information.

pace

From Longman Dictionary of Contemporary English

pace¹ /peɪs/ ●●○ **W3** noun 🗣️ 🔊

- 1 **SPEED OF EVENTS/CHANGES** [singular] the speed at which something happens or is done

pace of

🗣️ The pace of change in our lives is becoming faster and faster.

at a steady/slow etc pace

🗣️ Public spending continues to rise at a steady pace.

- 2 **WALK/RUN** [singular] the speed at which someone walks, runs, or moves

pace of

🗣️ You need to step up the pace of your exercises.

at a slow/leisurely/brisk etc pace

🗣️ Lucy set off at a leisurely pace back to the hotel.

🗣️ He quickened his pace, longing to be home.

🗣️ Traffic slowed to a walking pace.

- 3 **STEP** [countable] a single step when you are running or walking, or the distance you move in one step

pace backwards/towards/forwards etc

🗣️ He took a pace towards the door.

🗣️ Rebecca walked a few paces behind her mum.

4 → keep pace (with something/somebody)

5 → go through your paces

6 → put somebody/something through their paces

7 → set the pace

8 → force the pace

9 → be able to stand the pace

The screenshot shows the 'COLLOCATIONS' section of the LDOCE5 dictionary entry for 'Pace'. It is divided into two main categories: 'ADJECTIVES/NOUN + PACE' and 'VERBS'. Under 'ADJECTIVES/NOUN + PACE', there are four entries: 'rapid/fast' with the example 'The rapid pace of change creates uncertainty.', 'slow' with 'The pace of life in the countryside is slower.', 'a steady pace' with 'The economy was growing at a slow but steady pace.', and 'at your own pace (=at the pace that suits you)' with 'This allows each child to learn at his or her own pace.'. Under 'VERBS', there are five entries: 'the pace quickens/accelerates' with 'The pace of change is quickening.', 'the pace slows/slackens' with 'After a surge in exports, the pace slackened considerably the following year.', 'gather pace (=happen more quickly)' with 'Support for the campaign is gathering pace.', 'keep up the pace (=continue to do something or happen as quickly as before)' with 'China's society is transforming but can it keep up the pace?', and 'keep up with the pace (=do something as fast as something else is happening or being done)' with 'It's essential that we constantly update our skills and keep up with the pace of change.'

Figure 1: Screenshots of *Pace* in *LDOCE5*

4.5 A five-step dictionary training session

To impart the basic skills and strategies needed for collocation searches, a dictionary training session was devised. Chen (2017, 2020) provided useful reference for the training design as the studies shared some similarities in research design to the present one in terms of participants (similar linguistic background and English proficiency level), the dictionary used for research (the same type of learners' dictionary in the electronic form), and the collocation task involved (the same type of collocation). Based on Chen's findings, the training focused on five steps of dictionary use for collocation production, i.e. **Decide — Search — Locate — Interpret — Use.**

Step 1: Decide

To look for collocation information in a learner's dictionary, the first step is to determine which keyword of a collocation to look up. Chen (2017) identified through log files that some learners tended to figure out the English equivalent to the Chinese verb and then look it up in the dictionary, an ineffective or even futile strategy also found by Chen (2020) through a retrospective questionnaire survey. Therefore, to begin with, the importance of the keyword determination strategy was stressed and the participants were trained how to choose an appropriate keyword for consultation. They were also reminded not to employ inefficient or even wrong strategies. For example, to search for a verbal collocate for *__ a ban* (撤销禁令), instead of using the English equivalent to 撤销 as an entry word, the participants should select the noun *ban* as the keyword to look up, as conventionally most learners' dictionaries list V+N collocations under the noun entry. Besides, there may be more than one English equivalent to a Chinese verb, and some may be semantically acceptable yet syntagmatically inappropriate to collocate with a certain noun. For instance, *cancel, annul, revoke* are all equivalent to 撤销, yet *cancel/annul/revoke a ban* are obviously unacceptable.

Step 2: Search

After the keyword is decided on, the second step is to enter its correct form in the search bar. The participants' attention was drawn to the fact that different from print dictionaries, electronic dictionaries may have different search routes. They were shown that in *LDOCE5*, inflected forms like singular and plural nouns may have different collocates, that the use of the article before a noun may or may not be useful for collocation searches, and that some collocations can be retrieved by inputting the whole collocation while others cannot. To increase the chances of success and to save time as well, a more desirable practice was to use the base form of the entry word, i.e. to drop all its affixes or the article. For instance, to complete *__ one's teeth* (长牙齿), a better suggestion was to use the singular form *tooth* instead of the plural form *teeth*. In the same manner, to fill in the blank in *___ a bet* (下赌注), *bet* would be a better option than *a bet*.

Step 3: Locate

The third step is to find where the target collocation is. Some headwords can be used in more than one word class, for example, as both a verb and a noun. The participants were instructed to locate the right entry quickly according to word classes. For instance, in *LDOCE5*, *bet* as a verb is explained first, followed by its noun entry. To search a verbal collocate for *___ a bet* (下赌注), the participants

were advised to skip the verb entry, scroll down the page quickly and get directly to the noun entry where the target collocate can be located.

Chen (2017, 2020) reported that learners tended to look at the first or the beginning part of an entry. They seldom used the hyperlink function of the dictionary for a further search and were apt to lose patience when reading lengthy entries. Those problems were taken into consideration during dictionary training in which the following major points were introduced to the participants. Firstly, read collocation patterns and sentence examples within the entry, especially words in bold, where target collocations may be presented. Secondly, do not ignore the hyperlink function of the dictionary, for some collocations are listed as subsense headwords with their meaning and usage hidden. For example, as shown in Figure 2, some collocations of *joke* are presented together with a hyperlink label. A click on the label will lead one instantly to the meaning and examples of the searched collocation. Thirdly, pay attention to the COLLOCATIONS box which contains various collocations about the headword. And fourthly, try to be patient during dictionary search, especially when reading a long entry.

joke¹ /dʒəʊk \$ dʒoʊk/ ●●● **S2** **W3** noun [countable] 🔊 🔊

1 something that you say or do to make people laugh, especially a funny story or trick
joke about

🔊 a joke about absent-minded professors
🔊 I couldn't go out with someone **for a joke**, could you?

2 → **be a joke**

3 → **go/get/be beyond a joke**

4 → **something is no joke**

5 → **somebody can take a joke**

6 → **make a joke (out) of something**

7 → **somebody's idea of a joke**

8 → **the joke's on somebody**

Figure 2: Screenshot of *joke*

Step 4: Interpret

After the desired target collocation is located, the next step is to interpret. Previous studies (Li and Xu 2015, Chen 2017, 2020) proved that dictionary users have difficulty in understanding the meaning of collocations, especially in distinguishing different collocations related to a same headword. Therefore, the participants were given the following advice. Firstly, learn to distinguish word senses within a polysemous entry. Take *drop a case* (撤销诉讼) for example. *Case* has several different senses, some of which have a signpost at the very beginning of each sense for users to capture its meaning, like EXAMPLE, SITUATION, REASON/ARGUMENT, LAW/CRIME, BOX/CONTAINER (see Figure 3). The participants were instructed to read all the signposts, compare different senses and choose the one which fits the given context in the sentence.

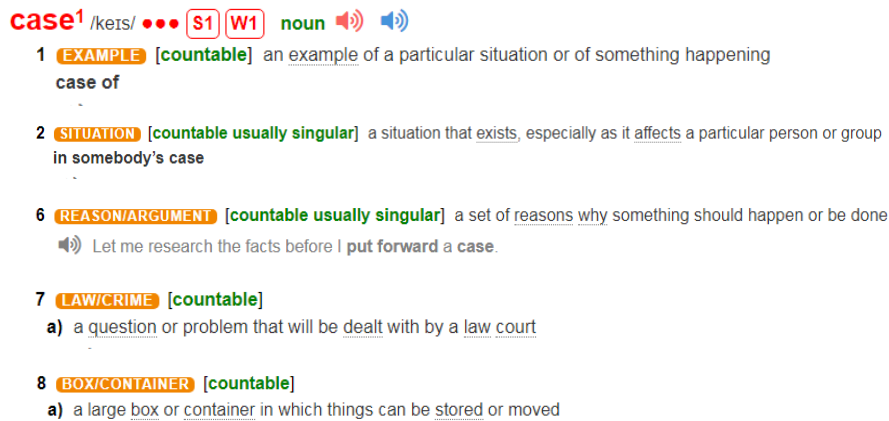


Figure 3: Screenshots of *case*

Secondly, pay close attention to the explanation and particularly the example that supports the meaning of a certain collocation. For instance, Figure 4 shows that more than five collocations are included in the COLLOCATIONS box of *nose*, all of which are different in meaning. The participants should read and compare these collocations carefully before deciding on which one to choose.

VERBS

blow your nose (=clear your nose by blowing strongly into a piece of soft paper or cloth)
She blew her nose on a large white handkerchief.

wipe your nose (=wipe liquid away from your nose)
The boy wiped his nose on his sleeve.

pick your nose (=remove substances from inside your nose with your finger)
Stop picking your nose, Freddy.

wrinkle your nose (=move the muscles near your nose when you do not like something)
Susan looked at the meal and wrinkled her nose.

hold your nose (=so that you cannot smell a bad smell)
The smell was so revolting that I had to hold my nose.

Figure 4: Screenshot of COLLOCATIONS box of *nose*

Thirdly, learn to distinguish between different collocations boxes. *LDOCE5* offers more than one COLLOCATIONS box for some headwords. For example, two boxes of COLLOCATIONS are presented, each for one sense of *perfume* (see Figure 5). The participants were guided to choose the box intended for the meaning that suits the given context.

COLLOCATIONS – Meaning 1: a liquid with a strong pleasant smell that women put on their skin or clothing to make themselves smell nice

VERBS

wear perfume
What's that perfume you are wearing?

dab perfume on something (=quickly put perfume on)
She dabbed some perfume on her throat.

COLLOCATIONS – Meaning 2: a sweet or pleasant smell

ADJECTIVES

sweet
She breathed in the sweet perfume of the roses.

Figure 5: Screenshots of *perfume*

Step 5: Use

When the desired collocates is identified, the last step is to use it in the context. The participants were trained to use the right form of the verb and make inflectional changes when necessary. For example, the sentence *The government has ____ a 10-year ban on arms sales to Africa* involves the present perfect tense, which means the verb here should be a past participle. Therefore, the right form is *lifted* instead of *lift*.

4.6 Procedure

The study was divided into three stages. In the first week, a preparatory session was given to all the three classes to acquaint them with the linguistic phenomenon of collocation including its definition, features and classification, along with a brief introduction of *LDOCE5* covering its website, search routes, features and functions.

The second week started with a pretest on the participants' prior knowledge about the target collocations. The pretest paper was the same as the one used in the condition treatment. The participants were asked to complete it without a dictionary or any other form of assistance. Then the three classes were assigned randomly to one of the following treatment conditions:

- (1) Dictionary use with prior instruction (Condition A): the participants received a fifteen-minute training session focusing on the five steps of dictionary use for collocation search, i.e. Decide — Search — Locate — Interpret — Use (see Section 4.5), and after that, they used *LDOCE5* to complete the collocation task.
- (2) Dictionary use without prior instruction (Condition B): the participants were asked to perform the collocation task with *LDOCE5* on their own.
- (3) Explicit collocation teaching (Condition C): the teacher listed each of the target collocations on the blackboard with its verbal collocates capitalized for visual salience, explained its meaning, illustrated the L1 and L2 congruency and gave example sentences which are different from the task sentence to support its use in a specific context. After that, the participants were asked to complete the task.

Students in Conditions A and B performed the task in a language lab equipped with computers, each of which was downloaded with a convenient access route to *LDOCE5*. *Screen Recording Expert V7.5* (<http://www.tlxsoft.com/index1.htm>) was preinstalled in the computers to record users' lookup behaviour, including what headwords they input, which part of dictionary information they focused on, whether or not they clicked on hyperlinks, and how long they spent on searching. Every shift of screen window, every move of the mouse and each step of dictionary search were captured in real time.

After completing the collocation task, all the participants received the corrective feedback to the task. They were asked to mark their corrections with colour pens so that their corrections can stand out against their original answers on the test paper before they handed it in. The marking of task performance was based on the original non-colour answers on the test paper.

An unexpected retention test was administered two weeks later. The participants were required to complete the same task again without any form of assistance. The retention test was the same as that in the previous stage except for the reshuffled order of sentences to avoid a carry-over effect.

4.7 Data analysis

Four kinds of data were collected: pretest score, task score before correction, retention test score, and each participant's screen recording (from Conditions A and B). The maximum test score is 10 points with one point for each correct response. Minor mistakes such as misspelling or a wrong use of verb tenses still merited one point. Screen recordings were scrutinized for the following data: 1) the time length of dictionary search, 2) the number of words searched, and more importantly 3) each step of dictionary lookup. Quantitative data were put into analysis using SPSS 20 and qualitative data were compared to identify the differences in dictionary consultation behaviour between the two dictionary classes.

5. Results and discussions

5.1 Effect of learning conditions on collocation retention

Table 1 shows that in the pretest, the students can only provide an average of 0.31 correct verbs for the ten target collocations. It can be seen from the raw data that only three out of 88 students were familiar with two target collocations and several others knew one. This is an expected finding as the target collocations were supposed to be unfamiliar to the students. The results of univariate GLM in Table 2 indicated that the students in the three conditions were not significantly different from each other in their prior knowledge about the target collocations [$F(2,85) = 0.946, p > 0.05$].

Table 1: Descriptive Statistics (Dependent Variable: pretest score) (Max = 10)

Condition	Mean	Std. Deviation	N
A	0.43	0.634	28
B	0.23	0.430	30
C	0.27	0.640	30
Total	0.31	0.575	88

Table 2: Tests of Between-Subjects Effects (Dependent Variable: pretest score)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	0.625 ^a	2	0.313	0.946	0.392	0.022
Intercept	8.422	1	8.422	25.484	0.000	0.231
Condition	0.625	2	0.313	0.946	0.392	0.022
Error	28.090	85	0.330			
Total	37.000	88				
Corrected Total	28.716	87				

a. R Squared = .022 (Adjusted R Squared = -.001)

The retention scores in the three conditions were also analyzed through univariate GLM. As displayed in Table 3, Condition A brought about the best retention score ($M = 6.43$), followed by Condition B ($M = 4.73$) and then Condition C ($M = 4.37$). A significant difference was identified in the mean scores of the three conditions with a medium effect size [$F(2, 85) = 4.417, p = 0.015 < 0.05, \eta^2 = 0.094$] (see Table 4). The results of Post-hoc tests (LSD) in Table 5 demonstrated that Condition A yielded significantly higher retention scores than Condition B ($p = 0.024 < 0.05$) and Condition C ($p = 0.006 < 0.05$) while Conditions B and C did not differ from each other substantially ($p = 0.614 > 0.05$). The plot in Figure 6 also showed that Condition A was more conducive to collocation retention than Conditions B and C.

Table 3: Descriptive Statistics (Dependent Variable: posttest score) (Max = 10)

Condition	Mean	Std. Deviation	N
A	6.43	2.873	28
B	4.73	2.559	30
C	4.37	2.965	30
Total	5.15	2.911	88

Table 4: Tests of Between-Subjects Effects (Dependent Variable: posttest score)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	69.389 ^a	2	34.695	4.417	0.015	0.094
Intercept	2355.287	1	2355.287	299.839	0.000	0.779
Condition	69.389	2	34.695	4.417	0.015	0.094
Error	667.690	85	7.855			
Total	3069.000	88				
Corrected Total	737.080	87				

a. R Squared = .094 (Adjusted R Squared = .073)

Table 5: Multiple Comparisons (Dependent Variable: posttest score)

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
A	B	1.70*	0.736	0.024	0.23	3.16
	C	2.06*	0.736	0.006	0.60	3.53
B	A	-1.70*	0.736	0.024	-3.16	-0.23
	C	0.37	0.724	0.614	-1.07	1.81
C	A	-2.06*	0.736	0.006	-3.53	-.60
	B	-0.37	0.724	0.614	-1.81	1.07

Based on observed means.

The error term is Mean Square (Error) = 7.855. *. The mean difference is significant at the .05 level.

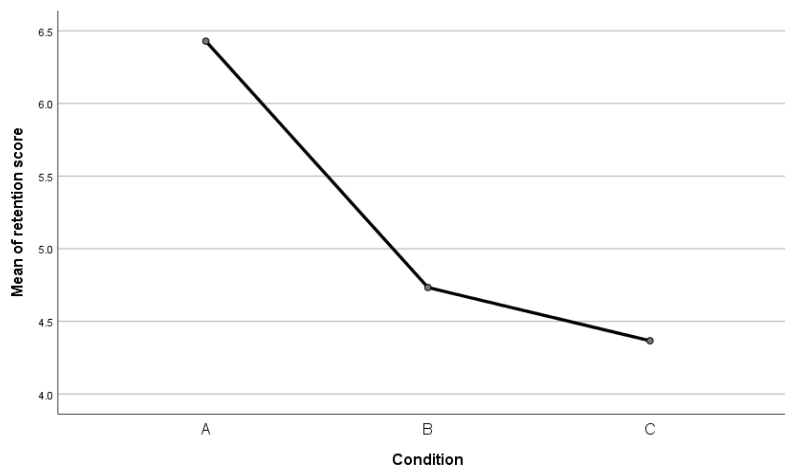


Figure 6: A plot of means of retention scores

The study reveals that learning conditions exert a significant effect on collocation gains. The students who had been trained on dictionary use achieved the best collocation retention. As will be explained in Section 5.2, compared with the students who used the dictionary without prior training, students who received training were equipped with better lookup skills and more effective strategies. They searched more words, decided on the right lookup word, read information more carefully and were more aware of the hyperlink function, thus retrieving more correct verbs in the first place. Obviously, dictionary training proved to be effective as it helped to improve students' skills and facilitate the long-term retention of retrieved collocations. Compared with traditional collocation teaching in which students remained as passive receivers of

collocational knowledge, the use of a dictionary involved more cognitive efforts on the part of users, which might leave a deeper memory trace, thus better retention. Nevertheless, it should be noted that dictionary use without prior training did not bring about significantly better retention than the explicit teaching condition (see Table 5). As will be presented in Section 5.2, those students had limited dictionary skills and encountered many problems and difficulties during dictionary consultation. The inadequate competence in dictionary use apparently reduced the usefulness of the dictionary for collocation learning, which further points to the significance of dictionary use instruction.

Different from previous research which mostly examined the contribution of dictionaries to collocation comprehension, production and retention in incidental conditions, the present study involved the learning process, i.e. the students were given corrective feedback to the task and were asked to rectify their own answers with colour pens for the sake of visual salience. It turned out that the retention scores in the study were higher than those in most previous research. As displayed in Table 3, the retention rate of the two dictionary classes reached 64.3% and 47.3% respectively. Compared with Laufer's study (2011) in which the intermediate participants retained only about 1.5 new collocation against a total of 12 target collocations, Dziemianko's (2012) in which the electronic dictionary users recalled 28.7% of productive collocational knowledge, and Chen's (2017, 2020) which yielded a retention rate of 36.1% and 34.5% respectively, the present study produced more promising results. It can be hypothesized that dictionary use combined with learning activities can contribute more to collocation gains than dictionary use merely intended for task completion.

5.2 Differences in dictionary use between students with and without dictionary training

Each student's screen recording was scrutinized carefully and comparisons were made between the two conditions of dictionary use from the following aspects:

5.2.1 The success rate of task completion

The data for this part were collected from the test paper. An independent samples T-test was conducted to compare task scores from Conditions A and B. As Table 6 shows, the students in Condition A obtained an average of 6.21 correct verbs from the dictionary while those in Condition B made only 3.63 right responses to the ten target collocations. A significant difference was identified ($p < 0.001$). In other words, dictionary use with prior training yielded significantly better task performance than autonomous dictionary use.

Table 6: Independent Samples Test of Task Score (Max = 10)

	Condition	N	Mean	Std. Deviation	Std. Error Mean	t	Sig.(2-tailed)
Task score	A	28	6.21	1.475	0.279	6.219	0.000
	B	30	3.63	1.671	0.305		

5.2.2 The number of words looked up

The author identified the number of words searched by each student from their screen recordings. Table 7 indicates that the students in Condition A consulted 9.43 headwords on average whereas those in Condition B looked up an average of 7.63 words. The results of the independent sample T-test revealed a significant difference ($p = 0.002 < 0.005$), pointing to the fact that substantially more words were consulted in Condition A than in Condition B.

Table 7: Independent Samples Test of the Number of Lookup Words

	Condition	N	Mean	Std. Deviation	Std. Error Mean	t	Sig.(2-tailed)
Number of lookup words	A	28	9.43	1.399	0.264	3.196	0.002
	B	30	7.63	2.646	0.483		

5.2.3 The time length of dictionary consultation

Table 8 shows that the average time length of dictionary consultation was 13.82 minutes in Condition A and 12.57 minutes in Condition B. The results of an independent sample T-test revealed no significant difference between the two conditions ($p = 0.180 > 0.05$).

Table 8: Independent Samples Test of the Time Length of Lookup

	Condition	N	Mean	Std. Deviation	Std. Error Mean	t	Sig.(2-tailed)
Time length of lookup	A	28	13.82	3.300	0.624	1.357	0.180
	B	30	12.57	3.739	0.683		

5.2.4 Lookup skills and strategies

A. Deciding which keyword of a collocation to look up

The screen recordings demonstrated that some students in Condition B

adopted an "equivalent" strategy when making decisions on which keyword of a collocation to look up. For example, to complete the collocation *grant one's wish* (满足愿望), they chose to search *meet* or *satisfy*, which are English equivalents to 满足. Similarly, a few students entered *cancel* when they tried to fill up the collocation *drop the case* (撤销诉讼), taking *cancel* as an English equivalent to 撤销. Another problem is that some students searched part of a collocation or a whole collocation which they assumed to be correct, using *the time*, *the case*, *the queue*, **force strike*, **mind the time*, **recognize the time*, **pray a prayer* as lookup words instead of searching the noun in the collocation which is a more effective strategy. The case was much better in Condition A. The screen recordings showed that a majority of the students selected the noun in the collocation as the lookup word. Only two students searched *the time*, *the case*, probably due to their absent-mindedness during dictionary training.

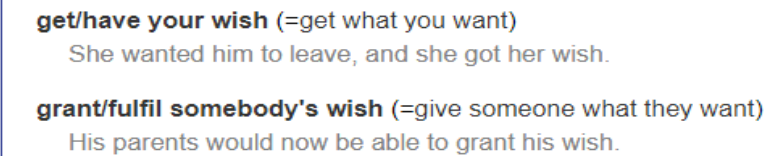
B. Locating the needed collocation information

As can be observed from screen recordings, the students in Condition B encountered many problems when they tried to locate the desired collocation information. In *LDOCE5*, headwords of different word classes are treated according to use frequency. For instance, *strike*, *fight* and *wish* are explained as a verb first, followed by a noun entry. Many students were unaware of the arrangement order of word classes and started by reading the whole long verb entry, only to find that the noun entry is far behind. And when they finally got to the noun entry, they had probably lost their patience, thus making a hasty choice. Confused by the design of dictionary information displays, some students spent unnecessarily long time reading irrelevant information such as **More Results** listed on the right of the screen which are intended to lead users to other information. Curiously, several students read the information in the Register box or even the Grammar box in the entry. Instead of paying close attention to the examples within the entry which may contain collocation information, many students chose to read **Examples From Corpus** presented at the end of the entry, which are just a pile of sentences from corpus without specific senses distinguished. Furthermore, a majority of the students did not click on the hyperlink label, either because they did not see the label at all or because they did not know that it would lead them to the information they may need. Another noticeable problem is that the longer entry the dictionary offers, the less chances of success in retrieving the correct information.

Fortunately, these above-mentioned problems are much less found in Condition A. Most of the students chose to skip the verb entry and came directly to the noun entry, which saved them a lot of time and energy. They were aware of the hyperlink function of the dictionary and, with a couple of exceptions, did click on the label. Nevertheless, these students also had trouble with long entries. The screen recordings showed that quite a few students missed the correct information when faced with a lengthy entry.

C. Interpreting the collocation information

Interestingly, the students in both conditions of dictionary use experienced similar difficulties in interpreting and distinguishing meanings. Many were confused by the seemingly synonymous collocations and failed to choose the right one for the given context. The following pairs of collocations caused high error rates in both conditions, i.e. *get/have ones' wish* vs. *grant/fulfill one's wish*, *have a fight* vs. *pick a fight*, and *dismiss a case* vs. *drop a case*. To fill in the sentence *So far we have been unable to ___ her wish* (满足愿望), some students read the COLLOCATIONS box (see Figure 7) and chose *get one's wish* (实现愿望) instead of *grant/fulfil one's wish* (满足愿望), probably because they failed to notice the difference in the subject when the two verbs *get* and *grant* are used.

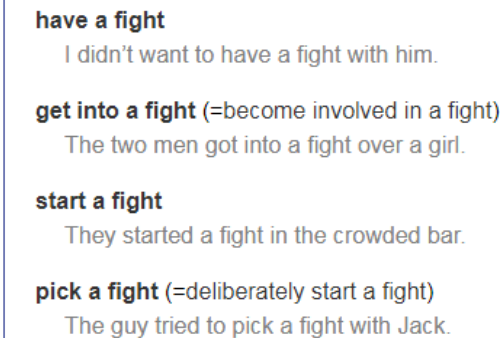


get/have your wish (=get what you want)
She wanted him to leave, and she got her wish.

grant/fulfil somebody's wish (=give someone what they want)
His parents would now be able to grant his wish.

Figure 7: Screenshot of the COLLOCATION box of *wish*

In the sentence *He tried to ___ a fight with me* (挑衅打架), the Chinese translation clearly states that the fight is deliberately started, but many students still used *have a fight* (打架). It can be seen from Figure 8 that the four collocations, *have a fight*, *get into a fight*, *start a fight* and *pick a fight* are synonymous in some degree. It is easy for users to get confused if they do not read carefully.



have a fight
I didn't want to have a fight with him.

get into a fight (=become involved in a fight)
The two men got into a fight over a girl.

start a fight
They started a fight in the crowded bar.

pick a fight (=deliberately start a fight)
The guy tried to pick a fight with Jack.

Figure 8: Screenshot of the COLLOCATIONS box of *fight*

Likewise, many students were unable to distinguish *dismiss a case* (驳回、撤销诉讼) from *drop a case* (撤销、撤回诉讼). Figure 9 demonstrates that the difference

between those two collocations lies in one key word, i.e. *officially*. Both the explanation and the sentence example indicate that *dismiss a case* should be an action taken by a court or other authority instead of by an individual. The sentence *The lawyer advised him to ___ the case since he stands little chance to win* (撤销诉讼) shows that the action is performed by an individual, so the right answer should be *drop a case* rather than *dismiss a case* which was used by many students. It would be more helpful for users to choose the right collocation if the dictionary provided a more explicit example to support the use of *drop a case*.

dismiss/throw out a case (=officially stop it from continuing)

The case was thrown out by New York state's highest court.

drop a case (=not continue with it)

The case was dropped because of a lack of evidence.

Figure 9: Screenshot of the COLLOCATIONS box of *case*

In a nutshell, the difficulties in distinguishing between different collocations resulted either from the carelessness or the limited lexical/grammatical knowledge on the part of users or from the insufficient examples given by the dictionary.

D. Other issues

The screen recordings revealed some other issues which are mostly concerned with the attitude towards dictionary use. The students in Condition B were more inefficient in dictionary use. One typical example is that a student spent more than five minutes reading the entry of *time*, moving the mouse up and down repeatedly over the entry, but eventually failed to locate the target collocation. Many students were also not patient enough during dictionary consultation. For example, the COLLOCATIONS box of *strike* contains many collocations related to the headword, such as *go on a strike*, *begin a strike*, *call a strike*, *stage a strike*, *end/call off a strike*, and the target collocation *break a strike* is given at the end of the box. Many students did read the box, but they did not scroll down to the end of the box, thus missing the target collocation. It was also found that a couple of students consulted the dictionary for only three headwords out of the ten target ones, either out of laziness or self-confidence that they knew the target collocations so it was unnecessary to use the dictionary.

Comparatively speaking, the students in Condition A took a more conscientious approach to dictionary use during the task. Many were found to read the same entry information more than once until they located the desired in-

formation. They also searched more headwords. Four students even took notes and wrote down the dictionary information on the test paper. Nevertheless, quite a few students were still not attentive enough. Sometimes they read repeatedly over the entry but still failed to identify the correct information. The students' task scores showed that none of the students retrieved all the right verbs for the ten target collocations. Given that all the target collocations are covered by the dictionary and that they had been trained how to search a collocation, a success rate of 62.1% is still not satisfactory enough. Being careless may be one of the reasons. Obviously, it takes time to develop good habits of dictionary use.

6. Conclusion

By engaging students in collocation learning under different conditions, the study identifies a significant effect of learning conditions on collocation gains. Dictionary use with prior training yielded better collocation retention than dictionary use without training and explicit collocation teaching. The task-based dictionary use instruction proves to be viable and effective. Compared with those who used the dictionary autonomously, the students with prior training were more skillful and employed more effective strategies concerning the selection of lookup words and the location and identification of collocation information. Being more careful, patient and involved in dictionary consultation, they had a higher success rate of task completion and achieved better collocation retention. In addition to proving the effectiveness of dictionary training for the improvement of users' skills and collocation retention, the study also demonstrates how to devise and perform a five-step procedure of task-based dictionary use instruction.

It is no easy job to help language learners grow into proficient and competent dictionary users. A well-designed task-based training session in dictionary use can help learners to develop practical skills, adopt effective strategies and modify their lookup behaviour, but a single training session is far from enough. As evidenced by the study, the students who were taught how to use *LDOCE5* still had difficulties in dealing with long entries and distinguishing between different collocations. Some were still too careless to notice easily retrievable information. To grow into ideal dictionary users, EFL learners should receive systematic and comprehensive dictionary use instruction. To this end, teacher training has become necessary and important. To educate EFL learners on dictionary use, EFL teachers should be confident about their own dictionary use skills and familiar with teaching content and methodology. An in-service training course in dictionary use offered by Bae (2015) and short-term workshops of teacher training proposed by Chi (2020) are worthwhile endeavors to equip EFL teachers with "lexicographical knowledge and teaching methodology to ensure they are well-equipped to design dictionary use learn-

ing activities suitable for their students" (Chi 2020: 92). Still, more efforts should be made to explore and experiment in this field.

Based on the study, some suggestions are made to improve the design of the online version of *LDOCE5*. When it comes to information displays of polysemous headwords with more than one word class, it is advisable to make use of the hyperlink function, making the entries of each word class available upon a mouse click and concealing entries of other word class, so that dictionary users can directly get to the target entry. Since long entries pose problems for many dictionary users, such a treatment can reduce distraction from other unrelated information and improve the success rate of information retrieval. The design of search routes should be more user-friendly. In *LDOCE5*, some collocations can be searched by typing in the whole collocation while others cannot. It would be time and effort-saving if all the collocations presented in the COLLOCATIONS box were made retrievable via inputting the whole collocation in the search bar. Sometimes users may need to confirm the existence or the use of certain collocations. A same search route should also be applicable to all cases of search, as in *LDOCE5* the use of *a/an* and *the* plus a noun may lead one to the target information only in some cases. In addition, to help users better understand and distinguish between collocations, all collocations in the box should be given an explanation and supported with more typical examples. Furthermore, as revealed by previous research (Li and Xu 2015, Chen 2017, 2020) and confirmed by the present study, many users tend to ignore the use of hyperlink function of e-dictionaries, so it would be more desirable if the hyperlink labels in *LDOCE 5* were highlighted for visual salience. The colour and font of collocations in example sentences should also be more eye-catching, so that users will not miss them easily. As suggested by Diemand-Yauman et al (2011: 111), superficial changes to learning materials could yield significant improvements in educational outcomes. This may also apply to dictionary use. Dziemianko (2015) proved that functional labels in colour significantly increased the speed and effectiveness of online dictionary search and enhanced short-term retention. It is believed that judicious use of colour and font or other highlighting methods in the dictionary interface design will benefit users of e-dictionaries.

It should be acknowledged that the study has a number of limitations. Due to time constraints in the classroom, the task only involved ten target collocations. The results would be more convincing if a bigger sample size of data were collected. Besides, the study did not consider the influence on collocation extraction of such factors as the type of target collocations, the sentences used, the differences in entry length or the positioning of collocations within entries. In addition, the study employed a screen recorder to keep track of how each student consulted the dictionary. This method can provide authentic data about every input of searched words, every move or click of the mouse and every step of dictionary consultation, yet more insights would be obtained if it were combined with other methods like retrospective interviews. Furthermore, the comparisons were made between the students with and without dictionary

training. If the same group of students were compared before and after they received dictionary use instruction, more would be revealed about how they improved their skills and modified their consultation behaviour.

As a worthwhile research area, dictionary use instruction should receive more attention from both lexicographic researchers and EFL teachers. Future research can focus on the integration of dictionary training with language learning, the scientific design of learning tasks to practice dictionary consultation, or the optimization of the content and methodology of dictionary use instruction. It may also be interesting to explore the long-term effects of dictionary training on user's overall language development. And it should be noted that with e-dictionaries gaining popularity, the dictionary use instruction should be based on the consultation habits and reference needs of e-dictionary users.

Acknowledgements

This study was funded by Fujian Province Educational and Pedagogical Reform Program (FBJG20210176). I'm grateful to the anonymous reviewers for their insightful comments and constructive suggestions. I would also like to thank all the students for their participation in the study.

References

- Alzi'abi, S.E. 2012. On Arab EFL Learners' Use of Collocational Information in Electronic Dictionaries. *Damascus University Journal* 28(3+4): 63-99.
- Alzi'abi, S.E. 2016. Help or Hindrance: "Etc" in Learner's Dictionaries? A Study of Verb–Noun Collocations. *European Journal of Applied Linguistics* 4(2): 301-327.
- Alzi'abi, S.E. 2017. Guessing Verb–Adverb Collocations: Arab EFL Learners' Use of Electronic Dictionaries. *Lexikos* 27: 50-77.
- Bae, S. 2015. A Course in Dictionary Use for Korean EFL Teachers. *Lexicography. Journal of ASIALEX* 2(1): 45-69.
- Basal, A. 2019. Learning Collocations: Effects of Online Tools on Teaching English Adjective–Noun Collocations. *British Journal of Education Technology* 50(1): 342-356.
- Boers, F. and S. Lindstromberg. 2012. Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics* 32: 83-110.
- Boers, F., S. Lindstromberg and J. Eyckmans. 2014. Some Explanations for the Slow Acquisition of L2 Collocations. *VIAL—Vigo International Journal of Applied Linguistics* 11(1): 41-62.
- Chen, Y. 2017. Dictionary Use for Collocation Production and Retention: A CALL-based Study. *International Journal of Lexicography* 30(2): 225-251.
- Chen, Y. 2020. The Efficacy of Dictionary Use for Collocation Learning — A Case Study of Smartphone Dictionary. *Lexicographical Studies* 2: 20-31.
- Chi, M.L.A. 2020. Reconstructing the Lexicographical Triangle Through Teaching Dictionary Literacy to Teachers of English. *Lexicography. Journal of ASIALEX* 7(1–2): 79-95.

- Cowie, A.P.** 1981. The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics* II(3): 223-235.
- Dai, Y., Z. Wu and H. Xu.** 2019. The Effect of Types of Dictionary Presentation on the Retention of Metaphorical Collocations: Involvement Load Hypothesis vs. Cognitive Load Theory. *International Journal of Lexicography* 32(4): 411-431.
- Diemand-Yauman, C., D.M. Oppenheimer and E.B. Vaughan.** 2011. Fortune Favors the Bold (and the Italicized): Effects of Disfluency on Educational Outcomes. *Cognition* 118(1): 111-115.
- Durrant, P. and N. Schmitt.** 2009. To What Extent Do Native and Non-native Writers Make Use of Collocations? *International Review of Applied Linguistics* 47(2): 157-177.
- Dziemianko, A.** 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.
- Dziemianko, A.** 2011. Does Dictionary Form Really Matter? Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX2011 Proceedings. Lexicography: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto, Japan, August 22–24, 2011*: 92-101. Kyoto: Asian Association for Lexicography.
- Dziemianko, A.** 2012. Why One and Two Do Not Make Three: Dictionary Form Revisited. *Lexikos* 22: 193-216.
- Dziemianko, A.** 2014. On the Presentation and Placement of Collocations in Monolingual English Learners' Dictionaries: Insights into Encoding and Retention. *International Journal of Lexicography* 27(3): 259-279.
- Dziemianko, A.** 2015. Colours in Online Dictionaries: A Case of Functional Labels. *International Journal of Lexicography* 28(1): 27-61.
- Dziemianko, A.** 2017. Dictionary Form in Decoding, Encoding and Retention: Further Insights. *ReCALL* 29(3): 335-356.
- Hoey, M.** 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Kim, S.** 2017. "Minimum Input, Maximum Output, Indeed!" Teaching Collocations Through Collocation Dictionary Skills Development. *Lexikos* 27: 265-286.
- Kim, S.** 2018. EFL Learners' Dictionary Consultation Behaviour During the Revision Process to Correct Collocation Errors. *International Journal of Lexicography* 31(3): 312-326.
- Komuro, Y.** 2009. Japanese Learners' Collocation Dictionary Retrieval Performance. Barfield, A. and H. Gyllstad (Eds.). 2009. *Researching Collocations in Another Language: Multiple Interpretations*: 86-98. Basingstoke: Palgrave Macmillan.
- Lauffer, B.** 2011. The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. *International Journal of Lexicography* 24(1): 29-49.
- Lauffer, B. and T. Waldman.** 2011. Verb–Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61(2): 647-672.
- Lew, R.** 2012. The Role of Syntactic Class, Frequency, and Word Order in Looking up English Multi-Word Expressions. *Lexikos* 22: 243-260.
- Lew, R. and M. Radłowska.** 2010. Navigating Dictionary Space: The Findability of English Collocations in a General Learner's Dictionary (LDOCE4) and Special-purpose Dictionary of Collocations (OCD). Ciuk, A. and K. Molek-Kozakowska (Eds.). 2010. *Exploring Space: Spatial Notions in Cultural, Literary and Language Studies*. Volume 2: *Space in Language Studies*: 34-47. Newcastle upon Tyne: Cambridge Scholars Publishing.

- Li, J. and N. Schmitt.** 2010. The Development of Collocation Use in Academic Texts by Advanced L2 Learners: A Multiple Case Study Approach. Wood, D. (Ed.). 2010. *Perspectives on Formulaic Language: Acquisition and Communication*: 22-46. London/New York: Continuum.
- Li, Lingling and H. Xu.** 2015. Using an Online Dictionary for Identifying the Meanings of Verb Phrases by Chinese EFL Learners. *Lexikos* 25: 191-209.
- Nesselhauf, N.** 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24(2): 223-242.
- Peters, E.** 2016. The Learning Burden of Collocations: The Role of Interlexical and Intralexical Factors. *Language Teaching Research* 20(1): 113-138.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Wray, A.** 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wu, J.H.** 2011. *A Cognitive Approach to the Lexicographic Representation of Collocations in English-Chinese Learners' Dictionaries*. Beijing: Kexue Publishing House.
- Yamashita, J. and N. Jiang.** 2010. L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly* 44(4): 647-668.

Appendix

A Collocation Task

Directions: Each of the following sentences contains a V + N collocation. Please provide an appropriate verb to complete the target collocation according to the Chinese translation given in the brackets.

1. When people are lining up in a shop, it is impolite to try to _____ **the queue**. (插队, 加塞儿)
2. Her youngest daughter has just learned to _____ **the time**. (识钟表, 看钟表)
3. The government has threatened to bring in the army to _____ the 10-month-old **strike**. (迫使罢工结束)
4. She _____ **a prayer** for their safe return. (祈祷)
5. He tried to _____ **a fight** with me. (挑衅打架)
6. So far we have been unable to _____ her **wish**. (满足愿望)
7. The tree _____ a small, bitter **fruit**. (结果实)
8. What he wanted me to do was to _____ him a big **cheque**. (开支票)
9. The lawyer advised him to _____ **the case** since he stands little chance to win. (撤销诉讼)
10. Look at the colour of you! You really _____ **the sun**, didn't you? (被晒黑)

Les débuts de la lexicographie multilingue roumaine: ressorts pragmatiques et influences culturelles

Mihaela Mocanu, *Institut de Recherche Interdisciplinaire,
L'Université "Alexandru Ioan Cuza" de Iasi, Roumanie*
(mocanu.mihaela@uaic.ro)

Résumé: Les débuts de la lexicographie multilingue roumaine sont étroitement liés à la parution sur le territoire des Principautés des premiers textes ecclésiastiques traduits de la langue slavonne. Les premiers dictionnaires répondent à des besoins d'ordre pratique, étant conçus comme des annexes à la littérature religieuse, slavonne ou traduite, qui circulait. L'étude réalise une radiographie des premiers projets lexicographiques roumains, avec l'identification des prototypes que ceux-ci enregistrent et la description des ressorts d'ordre pragmatique qui sont à la base des initiatives lexicographiques. Malheureusement, beaucoup de ces travaux n'ont pas été conservés, et d'autres n'ont été préservés que de manière fragmentaire, étant identifiés dans des bibliothèques privées. Dans ces conditions, pour reconstruire la lexicographie roumaine de cette période, nous avons fait appel à deux catégories de sources: les notations et les témoignages des auteurs des premiers projets lexicographiques, respectivement les études faisant référence à la lexicographie du roumain ancien. Comme ils n'avaient pas à la base la conception rigoureuse, scientifique, d'une école lexicographique, les premiers dictionnaires roumains sont l'expression de l'initiative des érudits, reflétant tant leur vision sur le rôle de tels instruments lexicographiques, comme le degré de développement de la langue roumaine à l'époque.

Mots-clés: LEXICOGRAPHIE ROUMAINE, LES XVIe–XVIIe SIECLES, DICTIONNAIRES MULTILINGUES, TERMINOLOGIE, INFLUENCES CULTURELLES, RESSORTS PRAGMATIQUES

Abstract: The Beginnings of Romanian Multilingual Lexicography: Pragmatic Sources and Cultural Influences. The beginnings of Romanian multilingual lexicography are deeply rooted in the first religious texts translated from Slavonic on the territory of the Romanian Principalities. The first dictionaries were aimed at fulfilling practical needs, being conceived as annexes to Slavonic or translated religious literature circulating at the time on the Romanian territory. This study provides an analysis of the first Romanian lexicographic projects, identifying their prototypes and describing the pragmatic background that these lexicographic initiatives were based on. Many of these works have sadly been lost, while others have been just partially preserved in private libraries. In these conditions, for the reconstruction of the Romanian lexicography belonging to that specific period, we resorted to two categories of sources: the notations and testimonies of the authors of the first lexicographic projects, and studies on the old Romanian lexicography. As they were not based on the rigorous scientific conception of a proper

lexicographic school, the first Romanian dictionaries reflect the initiative of some scholars, revealing both their perspective on the role of such lexicographic tools and the stage of development of the Romanian language at the time.

Keywords: ROMANIAN LEXICOGRAPHY, 16TH–17TH CENTURIES, MULTILINGUAL DICTIONARIES, TERMINOLOGY, CULTURAL INFLUENCES, PRAGMATIC SOURCES

1. Introduction

L'église a joué sans doute un rôle essentiel dans l'apparition et le développement de la lexicographie européenne: en Europe occidentale médiévale, où la langue des érudits et du clergé était le latin, le besoin d'évangélisation a déterminé les missionnaires à créer des listes de correspondances (la langue latine — la langue vulgaire). Ainsi, les chercheurs sont d'avis que:

«in many of the languages of Europe, the origins of lexicography can be traced back to interlinear glosses in medieval manuscripts. Monks noted the vernacular equivalents of unfamiliar Latin Words on the manuscript as an aid to understanding the text» (Brown 2006).

[dans beaucoup de langues d'Europe, les origines de la lexicographie peuvent être tracées jusqu'aux gloses interlinéaires des manuscrits médiévaux. Les moines notaient les équivalences vernaculaires des mots latins moins familiers sur le manuscrit pour aider à comprendre le texte].

Ces glossaires contenaient d'habitude des termes usuels, nécessaires à parcourir les textes bibliques et liturgiques et servaient aux gens des églises et aux traducteurs, assurant la fidélité de ceux-ci par rapport à la lettre du texte religieux, considéré saint.

Considérée comme étant l'époque d'or de la lexicographie multilingue, la Renaissance a marqué la relance de l'étude des langues hébraïques, syriennes et grecques. L'admiration vers les valeurs de l'Antiquité a remis à l'attention la langue grecque qui devenait ainsi égale à la langue latine. D'autre part, le développement des relations internationales et l'intensification des mobilités a rendu nécessaire l'élaboration des manuels et des dictionnaires qui facilitent la compréhension des langues étrangères. Le latin, qui continuait à être le moyen privilégié de communication des érudits et des savants, constituait la langue source pour la plupart des travaux lexicographiques réalisés dans cette période. L'émancipation de la lexicographie de sous la tutelle du latin se produira vers la fin du Moyen Âge, surtout dans la période de la Renaissance, quand les lexicographes et les typographes qui se dédiaient à la traduction commencent à créer des dictionnaires bilingues et multilingues, dans lesquels le français et l'anglais prenaient la place du latin des travaux antérieurs (Delisle et Woodsworth 1995).

À partir du XVIIe siècle, l'augmentation de l'intérêt pour les langues étrangères, par des raisons politiques, commerciales, didactiques ou d'autre nature, a conduit à l'augmentation du nombre de travaux lexicographiques.

L'anglais et le français deviennent des langues de base pour un nombre varié de combinaisons, étant concurrencées par l'espagnol et le néerlandais. Surnommé le siècle des encyclopédistes, le XVIII^e siècle n'est pas étranger à la pratique de l'inventaire des langues, de sorte que le nombre des dictionnaires polyglottes et de ceux bilingues est en continuelle augmentation. Les langues européennes — l'allemand, l'anglais, l'italien, l'espagnol, le néerlandais, continuent à se rapporter au français dans les dictionnaires bilingues. Le XIX^e siècle marque un phénomène intéressant de point de vue lexicographique: comme suite à l'importance accordée par les linguistes à la langue sanscrite, on assiste à une multiplication des dictionnaires dans les langues assyrienne et perse. Même si des langues mortes, le latin et le vieux grec apparaissent eux aussi dans des travaux de lexicographie, en combinaison avec les principales langues européennes. Parallèlement, les dictionnaires dédiés aux langues «vivantes» continuent à se développer de plus en plus. L'intensification des échanges internationaux dans le XX^e siècle a conduit au développement des activités lexicographiques. Après la première guerre mondiale, le français, dominant dans les relations diplomatiques occidentales, commence à partager le monopole avec l'anglais.

Si au début le support matériel pour créer des dictionnaires était représenté par la table en argile, le papyrus, le parchemin, avec l'apparition de l'imprimerie, en 1450, la lexicographie européenne entre sur des nouvelles coordonnées. L'édition et la diffusion des ouvrages de lexicographie augmente d'une année à l'autre, étant dictées par les besoins d'ordre pratique du monde européenne. À l'époque moderne, le développement de la lexicographie informatisée a ouvert de nouvelles possibilités: elle a fait possible non seulement la digitalisation des travaux lexicographiques et la création de nouveaux dictionnaires à l'aide des instruments informatiques, mais a aussi permis la réalisation des banques de données lexicales (qui ont la capacité de stocker un très grand nombre de mots), à base desquelles sont réalisés des dictionnaires multilingues par des moyens exclusivement informatiques. D'autre part, l'intensification du phénomène de la migration a imposé aux pays européens l'identification des solutions concrètes pour garantir la communication interlinguale. Tous ces facteurs, sur le fond de la globalisation accentuée, de la mondialisation des relations politiques, économiques, scientifiques et culturelles, ont conduit au développement d'une manière accélérée de la lexicographie polyglotte, au niveau européen.

Ce travail comprend une présentation critique des premiers dictionnaires polyglottes parus dans l'espace roumain, à partir de la fin du XVII^e siècle et jusqu'au début du XIX^e siècle, quand le premier dictionnaire roumain à caractère explicatif, étymologique et polyglotte est édité. Dans les conditions où beaucoup de ces premiers travaux lexicographiques n'ont pas été conservés, et d'autres n'ont été préservés que partiellement, la démarche de reconstitution de la période mentionnée s'appuie sur une série de sources indirectes. Ainsi, pour la présentation des premiers projets lexicographiques nous avons fait appel, d'une part, aux notations des initiateurs de ces projets et aux témoignages des contemporains et, d'autre part, aux études de spécialité dédiées à la lexicographie roumaine ancienne.

2. Bref historique de la lexicographie roumaine

Une perspective diachronique sur la lexicographie roumaine indique trois stades d'évolution: (i) l'époque ancienne (du XVI^e siècle jusqu'en 1824); (ii) l'époque moderne (1825–1880) et (iii) l'époque postmoderne (de 1881–jusqu'à présent). La première étape se caractérise par la prépondérance des dictionnaires bilingues et plurilingues:

«Caracterul exclusiv bilingv și plurilingv pe care îl are, până în secolul al XIX-lea, lexicografia românească este determinat (...) de rolul prin excelență practic îndeplinit de aceasta» [Le caractère exclusivement bilingue et plurilingue que la lexicographie roumaine a, jusqu'au XIX^e siècle, est déterminé (...) par le rôle par excellence pratique accompli par celle-ci] (Seche 1966: 181).

La deuxième étape est sous le signe des versions intermédiaires du *Dictionnaire de Buda*, considéré comme le premier dictionnaire explicatif et étymologique de la langue roumaine. La troisième étape se caractérise par la parution des dictionnaires terminologiques multilingues, sur le fond du développement progressif des sciences et de la technique, comme réalisation pratique des résultats que la discipline de la terminologie, nouvellement créée (les années '40–'50 du XX^e siècle) obtient.

La parution des premiers ouvrages lexicographiques au XVI^e siècle, desquels on a conservé deux fragments de vocabulaire slave-roumain, est étroitement liée à la circulation des nouveaux textes religieux traduits de la langue slavonne. Les premiers projets lexicographiques répondent à des besoins d'ordre pratique, étant conçus comme des annexes à la littérature religieuse, slavonne ou traduite, qui circulait en Valachie. De la première moitié du XVII^e siècle on a conservé, partiellement, *Vocabular biblic sârbesc-românesc* [Vocabulaire biblique serbe-roumain], dont la rédaction est située par les lexicographes autour de l'année 1630, et, de manière intégrale, *Lexiconul slavo-românesc al lui Macarie Cozianul* [Le dictionnaire slave-roumain de Macarie Cozianul] daté de 1649. La lexicographie roumaine du XVII^e siècle est, en totalité, d'origine valaque. L'augmentation du nombre des dictionnaires est déterminée par la multiplication des traductions du slavon, la source de ceux-ci étant *Lexiconul slavo-rusesc și tâlcuirea numelor* [Le dictionnaire slave-russe et l'explication des noms] imprimé en 1627 par le hiéromoine russe Pamvo Berînda. À partir de ce matériel, les auteurs roumains ajoutent ou éliminent des unités de l'inventaire lexical original, en fonction des besoins pratiques. Parmi les glossaires créés de cette manière on peut distinguer, de point de vue de l'originalité, deux travaux: *Lexiconul păstrat în Codicele Sturdzan* [Le dictionnaire conservé dans le Manuscrit Sturdzan] (environ 1660–1670) et *Lexiconul lui Mihai* [Le dictionnaire de Mihai] (1672), qui comprennent des inventaires lexicaux extrêmement riches (plus de 8000 mots-titre), avec beaucoup de nouvelles entrées par rapport au travail de Pamvo Berînda (Seche 1966: 7-33). „Les dictionnaires roumains inspirés par le dictionnaire de Pamvo Berynda constituent la matière principale de la lexicographie roumaine du XVII^e siècle; il faut y ajouter cinq autres dictionnaires bilingues, deux avec le latin, deux avec l'italien et un avec le grec” (Gînsac et Ungureanu

2018: 849).

La démarche de reconstitution de la lexicographie roumaine ancienne s'avère être extrêmement difficile dans les conditions où la plupart des ouvrages ont disparu ou n'ont été conservés que de manière fragmentaire. Ainsi, certains projets lexicographiques nous sont connus par les annotations des auteurs ou à partir des témoignages des contemporains, et pour d'autres ouvrages on peut se faire une image à l'aide des fragments découverts dans des bibliothèques privées. Ainsi, en ce qui concerne certains travaux, nous pouvons nous en faire une image à l'aide des fragments découverts dans les bibliothèques privées, alors que d'autres initiatives lexicographiques nous sont connues grâce aux notations des auteurs ou des témoignages des contemporains. Une contribution importante à l'histoire de la lexicographie polyglotte roumaine dans son époque de début appartient à Mircea Seche, qui identifie trois grandes étapes dans l'évolution de la lexicographie roumaine multilingue (1687–1825; 1826–1880; 1880–1969), délimitées par rapport à trois moments de référence pour le développement du domaine lexicographique roumain: la parution du premier dictionnaire multilingue — *Dicționar latin-român-maghiar* [Dictionnaire latin-roumain-hongrois] (1687–1701); l'édition du premier dictionnaire explicatif, étymologique, polyglotte — *Lesicon romanesco-latinescu-ungurescu-nemțescu* [Dictionnaire roumain-latin-hongrois-allemand] (1825); le changement du contexte technique et scientifique international, qui a conduit à la parution des dictionnaires terminologiques plurilingues ayant le roumain dans leur composition (Seche 1966: 7-33).

L'analyse de la lexicographie multilingue roumaine relève l'existence de deux prototypes: d'une part, la tradition polyglotte, spécifique pour les XVIIIe et XIXe siècles, et, d'autre part le développement dans l'époque moderne des dictionnaires terminologiques et des vocabulaires techniques correspondant à certains domaines de spécialité (Pricop et al. 2017: 159-165). Le premier dictionnaire polyglotte enregistré dans l'espace roumain est un dictionnaire dans les langues latine, roumaine et hongroise, qui date de 1687–1701. Après la parution de celui-ci il suit une période de plus de moitié de siècle de «silence». Pendant les dernières décennies du XVIIIe siècle on enregistre trois directions de développement de la lexicographie multilingue: des listes multilingues de plantes (la première en 1783), des vocabulaires (manuscrits) dans les langues roumaine, grecque, française, allemande et russe, respectivement des dictionnaires polyglottes réalisés par les représentants de l'École d'Ardeal. L'intérêt pour l'élaboration d'un dictionnaire général de la langue roumaine accroît, mais, vu que pendant cette période la facilitation de la communication entre les différentes ethnies du territoire des Pays Roumains devient impérative, la réalisation de cet objectif reste en plan second, faisant circuler, en échange, les dictionnaires en plusieurs langues.

3. Les facteurs historiques et culturels qui ont favorisé l'apparition de la lexicographie roumaine

Partant de la prémisse que le dictionnaire est un miroir où l'utilisateur reconnaît

tant des données de l'histoire de la langue, comme des données concernant l'histoire de la culture (Guérard 2015), les premiers projets lexicographiques roumains portent l'empreinte du contexte extralinguistique qui les ont générés. Ainsi, une digression sur les débuts de la lexicographie roumaine ouvre une vaste perspective sur les influences culturelles et sur les intérêts d'ordre économique, politique et religieux qui ont gouverné l'apparition des premiers projets lexicographiques sur le territoire actuel de la Roumanie.

Les chercheurs pensent que les racines de la lexicographies roumaine apparaissent au XVI^e siècle, à l'occasion de la traduction de certaines notions de la langue slavonne dans la langue maternelle, dans le cadre de la lecture des textes slaves, quand on faisait appel aux gloses roumaines, tout comme au moment de la réalisation des traductions d'après de tels textes, quand certains mots et expressions des deux langues étaient traités de sorte que, mis les uns à côté des autres, les mots roumains expriment la notion correspondant aux mots slaves (Strungaru 1966). L'étape suivante, caractérisée par la composition des dictionnaires slave-roumains, marque le détachement et l'emploi indépendant de ces listes de mots, qui ne sont plus perçues comme des annexes à différents textes religieux:

«Apariția lucrărilor lexicografice române este aproape concomitentă cu apariția primelor texte bisericești traduse din limba slavonă. În această perioadă apar glosarele și vocabularele bilingve, slavo-române, latino-române ... Nevoile practice, de ordin administrativ, bisericesc și școlar, apoi afirmarea limbii materne au făcut ca într-o perioadă dată să apară la noi astfel de lucrări» (Kelemen 1962: 89-92).

[La parution des travaux de lexicographie roumains a lieu presque en même temps que la parution des premiers textes religieux traduits de la langue slavonne. À la même époque paraissent aussi les dictionnaires et les vocabulaires bilingues, slave-roumains, latin-roumains ... Des besoins pratiques, d'ordre administratif, religieux et scolaire, ensuite l'affirmation de la langue maternelle ont fait paraître chez nous, dans une période donnée, de tels travaux.]

Les efforts de création de certains travaux lexicographiques doivent être mis en relation avec une série de facteurs historiques et culturels qui ont favorisé l'apparition des projets lexicographiques sur le territoire de notre pays. Ainsi, les premières préoccupations lexicographiques de nos érudits ne constituent pas l'effet de certaines influences lexicographiques étrangères, mais ce sont le résultat de certains besoins d'ordre interne concernant l'éclaircissement de certains mots des textes slaves qui circulaient à l'époque. Dans le contexte où la langue slavonne était complètement différente de la langue des compatriotes, les érudits roumains ont fait appel à des gloses roumaines, marquées en marge du texte slave, à côté des mots et des expressions qui n'étaient pas clairs, dans le but d'expliquer leur sens. L'emploi des gloses roumaines dans le cadre des textes slaves, au XVI^e siècle, constitue les premières manifestations lexicographiques de l'espace roumain. La parution des premières gloses roumaines est aussi stimulée par la parution des premières traductions sur le territoire de notre pays.

L'étape suivante de la lexicographie roumaine reste sous le signe de la parution des glossaires slave-roumains, employés de manière indépendante de la circulation d'un certain texte. Etant le résultat de la transcription des gloses roumaines des manuscrits et des textes slaves imprimés sur des listes à part, les glossaires se présentaient sous forme d'inventaires des mots slaves, transcrits avec leur explication en roumain, et rangés sur des colonnes, ayant ainsi l'aspect de certains travaux lexicographiques proprement-dits. L'ordre de l'enregistrement des mots et des expressions reste cependant celui de l'occurrence dans le texte, et non pas l'ordre alphabétique, adopté par la suite par la pratique lexicographique.

La parution de plusieurs dictionnaires slave-roumains qui ont circulé en manuscrits, à la moitié du XVII^e siècle, indique une nouvelle étape dans le développement de la lexicographie roumaine. Parmi ceux-ci, *Lexiconul lui Pamvo Berînda* [Le dictionnaire de Pamvo Berînda] se distingue par l'inventaire et la clarification du thésaurus des mots slaves religieux, étant une synthèse de la lexicographie slave antérieure. Résultat de l'effort de transposer en langue roumaine différents textes slaves qui circulaient sur le territoire de notre pays, les dictionnaires bilingues slave-roumains rédigés au XVII^e siècle reflètent en même temps l'ambition de la diffusion d'une culture dans la langue maternelle dans l'espace roumain.

La fin du XVII^e siècle marque le début de la lexicographie multilingue roumaine, avec la parution du premier dictionnaire trilingue enregistré dans l'espace roumain: *Lexiconul latin-român-maghiar* [Le dictionnaire latin-roumain-hongrois], 1687-1701. Le travail paraît sur le fond d'un changement d'ordre culturel: la langue slavonne perd du terrain devant la langue latine, et la langue hongroise commence à avoir une influence sur le terrain de la lexicographie. Les démarches lexicographiques ont acquis ampleur au fur et à mesure que le contexte social et historique obligeait les provinces roumaines à une synchronisation avec l'Occident. Ainsi, le passage des listes courtes de mots à des glossaires et ensuite à des ouvrages lexicographiques fondamentaux, a constitué un processus de durée, qui a culminé avec l'effort de réaliser un dictionnaire général de la langue roumaine, le plus complexe et représentatif de point de vue diachronique, diatopique, diastratique et diaphasique. La réalisation de ce dictionnaire a été parachevée à peine il y a 10 ans, se concrétisant dans le dictionnaire trésor de la langue roumaine, élaboré par l'Académie Roumaine tout au long d'un siècle (1907-2010). Cette précision est pertinente pour motiver la division de la lexicographie roumaine en trois étapes, en fonction de deux moments de référence: la parution du premier dictionnaire général de la langue roumaine (*Lexiconul de la Buda* [Le dictionnaire de Buda], 1825) et le début de la lexicographie monolingue (les années '70 du XIX^e siècle) (Seche 1966, 1969). Les dictionnaires bilingues et plurilingues sont prépondérants, dans une première étape, fait justifié par leur caractère fonctionnel à l'époque: „Caracterul exclusiv bilingv și plurilingv pe care îl are, până în secolul al XIX-lea, lexicografia românească este determinat (...) de rolul prin excelență practic îndeplinit de aceasta” [Le caractère exclusivement bilingue et plurilingue que

la lexicographie roumaine a, jusqu'au XIXe siècle, est déterminé (...) par le rôle par excellence pratique qu'elle accomplit] (Seche 1966: 181).

4. Les premiers projets lexicographiques multilingues

Avec une histoire qui remonte jusqu'au XVIIe siècle, la lexicographie multilingue roumaine se caractérise par l'hétérogénéité, tant en ce qui concerne la sélection et l'établissement des inventaires des entrées, comme par la manière d'organisation du matériel lexicographique. Une perspective diachronique sur la pratique lexicographique roumaine relève des changements concernant les domaines de spécialité pour lesquels les dictionnaires sont élaborés et concernant les manières d'organisation de l'information lexicale.

Vers la fin du XVIIe siècle — le début du XVIIIe siècle, la lexicographie roumaine enregistre un tournant, la place de la langue slavonne étant reprise dans les dictionnaires bilingues et multilingues par le latin, respectivement par une autre langue romane. De même, on enregistre maintenant une augmentation du nombre des dictionnaires, de sorte que, en moins de deux décennies ont attesté six projets lexicographiques: un vocabulaire latin-roumain et un autre roumain-latin, un vocabulaire italien-roumain, un dictionnaire trilingue, latin-roumain-hongrois, un essai de vocabulaire grec-roumain et un glossaire de néologismes. D'autre part, le centre de gravité de la lexicographie roumaine se déplace maintenant de la Valachie vers la Transylvanie qui, pendant presque deux siècles dorénavant, détiendra la première place en ce qui concerne les initiatives lexicographiques de l'espace roumain. En Transylvanie, le contact direct avec le monde occidental se produit soit par la fréquentation de l'école de Blaj par les jeunes scribes, soit par la continuation des études aux collèges de Rome et Vienne, où ceux-ci se familiarisent avec le latin.

«L'Occidentalisation romane se produit ici, tout d'abord, par l'intermédiaire de l'étude du latin comme langue de culture, ce qui satisfait l'orgueil national de l'origine, de l'ancienneté et de la continuité roumaine, mais aussi comme argument dans la lutte d'émancipation socio-politique et spirituelle et d'affirmation comme la quatrième nation, avec des droits civiques égaux dans l'État» (Coman 1999: 7-8).

Le premier dictionnaire trilingue de la lexicographie roumaine est *Marsili Lexicon Latinum Walachi Ungaric*, identifié dans la bibliothèque du comte italien Luigi Ferdinando Marsigli (Seche 1966). Marsigli avait été ambassadeur à la cour de C. Brîncoveanu, mais il connaissait très bien la Transylvanie, ayant habité ici une période de temps. Selon les spécialistes, l'auteur du vocabulaire que Marsigli a acheté serait un transylvain d'origine allemande et il y a la possibilité que celui-ci ait rédigé le glossaire à la demande du diplomate italien. Le dictionnaire trilingue de Marsigli, connu sous le nom de *Lexiconul marsilian* [Le dictionnaire de Marsigli], marque le début de l'influence hongroise sur la lexicographie roumaine. L'ouvrage reprend une grande partie de la liste d'entrées du

Dictionnaire latin–hongrois, élaboré par Albert Molnár, probablement d’après une édition de 1700 (Seche 1966). L’origine étrangère de l’auteur est reflétée par la transcription à caractères latins des correspondances roumaines et par l’influence de la manière d’écrire en hongrois (ex. *szluga*, *karnacz* = cârnați [saucisse]; *kovács* = covaci «fierar» [forgeron]). Le dictionnaire comprend 2500 entrées, les mots étant choisis au niveau du vocabulaire général. Le dictionnaire réussit à éviter le caractère de glossaire que les dictionnaires bilingues avaient (Tagliavini 1930: 7), mais il présente de nombreuses carences: des omissions des correspondances en roumain et en hongrois, des espaces vides que l’auteur, avec ses connaissances de langue, n’a pas pu compléter.

La période 1704–1780 marque un recul dans le développement de la lexicographie roumaine, au moins de point de vue numérique: on atteste pour cette période deux glossaires slavon-roumains, deux glossaires latin–roumains et deux dictionnaires, de petites dimensions, italo-roumains, ces derniers étant rédigés par des auteurs étrangers. Dans la deuxième moitié du XVIIIe siècle apparaissent les premières initiatives d’élaborer des dictionnaires explicatifs unilingues roumains, restés cependant dans la phase de projet. Ils sont dus aux représentants de l’École d’Ardeal, mouvement d’émancipation politique-sociale des Roumains de Transylvanie dont les représentants donneront, d’ailleurs, au début du siècle suivant, les premiers dictionnaires scientifiques en langue roumaine.

À partir de 1780, le nombre des ouvrages lexicographiques augmente de nouveau. En 1783 paraît le premier glossaire trilingue avec des noms de plantes, rédigé par le professeur sicule Benkő József. Ce glossaire de terminologie spéciale, le deuxième par ordre chronologique après le glossaire bilingue de termes géographiques de Constantin Cantacuzino, marque le début de la lexicographie imprimée. Par rapport aux dictionnaires manuscrits, qui avaient une circulation réduite, et, par conséquent, une influence minimale, les dictionnaires imprimés jouissent d’une circulation accrue, étant valorisés dans la rédaction des travaux lexicographiques ultérieurs. Le glossaire botanique de Benkő József, qui comprend 620 noms de plantes, paraît dans le sommaire d’une revue (*Magyar Könyvház*, Pozsony, vol. II, 1783, p. 407) (Seche 1966). La transcription des mots roumains a été réalisée avec l’écriture hongroise, et le matériel lexical a été groupé d’après des classes de plantes et non pas d’après le critère alphabétique. Jusque vers la fin du XIXe siècle, les glossaires des noms de plantes restent prépondérants, représentant ainsi une constante de la lexicographie multilingue roumaine de tous les siècles.

Dans la dernière décennie du XVIIIe siècle, dans l’espace roumain paraissent trois dictionnaires polyglottes et trois dictionnaires bilingues. Le premier dictionnaire polyglotte de la série est *Dictionarii trium lingvarum germano-latina et daco-romana* [Dictionnaire trilingue allemand-latin et roumain], un manuscrit de grandes dimensions, en trois volumes. Le travail appartient au Transylvain Aurelius Antoninus Praedetis Nasody et date de 1792–1793. Il paraît que le manuscrit a été rédigé durant le séjour de l’auteur en Moldavie, fait prouvé par

le grand nombre de régionalismes moldaves dans la liste de mots. Les mots-titre en allemand, classés par ordre alphabétique et rédigés à caractères gothiques en italique, sont suivis par plusieurs correspondants en latin, en relation de synonymie et écrits avec des caractères latins. La liste de mots est rangée sur deux colonnes, et le travail ne présente pas un appareil scientifique.

Un autre dictionnaire polyglotte, *Dicționarul în patru limbi (român–german–francez–latin)* [Le dictionnaire en quatre langues (roumain–allemand–français–latin)], appartient à l’habitant de Banat Paul Iorgovici, qui a rédigé *Observațiile de limbă rumânească* [Les observations sur la langue roumaine] imprimées en 1799. Malheureusement, le travail d’Iorgovici n’a pas été conservé, et les informations concernant la manière de composition du travail nous sont parvenues de manière indirecte, par l’intermédiaire du travail *Observațiile de limbă rumânească*. L’auteur du dictionnaire ouvre la série des lexicographes roumains épurés, se proposant l’impression d’un glossaire spécial, qui comprenne les équivalences latines des mots non-latins de la langue roumaine, dans le but de remplacer les derniers termes avec ceux d’origine latine. Iorgovici visait le remodelage, en sens étymologiste, des mots roumains d’origine latine dont la forme s’éloignait trop de celle des correspondances latines, tout comme la création des mots nouveaux, par la dérivation à suffixes latins des racines avec la même origine. Connaisseur de la langue française, l’auteur a soutenu l’emprunt lexical du français comme moyen d’enrichissement de la langue roumaine. Le dictionnaire de Paul Iorgovici, daté autour de l’année 1800, est le premier dictionnaire en quatre langues de la lexicographie roumaine.

C’est de la fin du XVIIIe siècle que date aussi le petit *Dicționar francez–grec–român* [Dictionnaire français–grec–roumain], découvert par Nicolae Iorga parmi les manuscrits de Gh. A. Sturdza. Le travail, qui comprend 203 mots-titre, semble avoir été réalisé en Moldavie, d’après les particularités phonétiques et lexicales qu’il présente. Au début du XIXe siècle, le centre de la lexicographie roumaine se déplace de la Valachie et de la Moldavie vers la Transylvanie. Il suit une période extrêmement riche en ce qui concerne les préoccupations lexicographiques, étant attestés au moins dix dictionnaires ou projets lexicographiques. Même si sous l’influence formelle de la lexicographie hongroise, les travaux reflètent, au niveau de la conception et des objectifs, les tendances latines spécifiques à l’époque: les auteurs des dictionnaires essaient de démontrer la latinité du lexique de la langue roumaine, en sacrifiant beaucoup de fois la réalité objective. La sélection des mots et des formes lexicales, tout comme les indications étymologiques sont mises au service de cet objectif tendancieux.

En 1801, l’érudit transylvain Samuil Micu-Klein, l’une des figures proéminentes de l’époque et représentant de l’École d’Ardeal, achève la rédaction d’un dictionnaire bilingue: *Dictionarium valachico–latinum* [Dictionnaire valaque–latin]. Deux ans après l’achèvement, l’auteur se propose de transformer le travail dans un dictionnaire polyglotte, en quatre langues, par l’ajout des langues allemande et hongroise. En 1805, le dictionnaire polyglotte est achevé, ce travail étant considéré comme la première forme du *Dictionnaire de Buda*, qui allait

paraître en 1825. Le travail de Micu constitue le premier dictionnaire polyglotte qui, à côté des correspondants dans d'autres langues, comprend aussi des définitions des mots-titre roumains. Même si la dimension explicative du dictionnaire est plutôt incidentelle, elle doit être retenue pour la nouveauté qu'elle apporte dans l'histoire de la lexicographie roumaine. Huit ans après la disparition de Samuil Micu, les «héritiers» du savant transylvain réimpriment «Înștiințarea» [La notification] publiée par l'imprimerie de l'Université de Buda, avec la mention que le dictionnaire de Micu a souffert une série de finissages de la part de Vasile Codoși. Il commençait ainsi la longue transformation de l'ouvrage, qui allait mener, 11 ans après, à la parution du *Dictionnaire de Buda*. Même si la version de Micu est inférieure au *Dictionnaire de Buda*, il y a encore des spécialistes qui lui attribuent de manière intégrale la paternité du travail de 1825.

Vers 1802, un autre transylvain, Ștefan Crișan, ancien professeur aux collèges réformés de Cluj et Târgu Mureș, avait prêté pour être imprimé un *Dictionnaire roumain–latin–hongrois*. Découvrant le travail similaire de Samuil Micu, Crișan retire son dictionnaire de l'imprimerie. Le manuscrit du dictionnaire de Crișan est apporté par Gh. Asachi à Iași, où il se trouve de nos jours encore. Le travail compte plus de 10000 mots-titre et manifeste les mêmes tendances puristes évidentes de l'époque. Par rapport au travail de Micu, *Le dictionnaire* de Crișan comprend aussi une série d'observations étymologiques, les premières de ce genre dans l'histoire de la lexicographie roumaine.

Un autre dictionnaire polyglotte de cette période appartient à Gh. Șincai. Il s'agit du *Vocabularium pertinens ad tria regna naturae*, rédigé vers 1808–1810. L'auteur vise à intégrer la terminologie générale des sciences de la nature, les 427 termes du vocabulaire désignant des noms de plantes, d'animaux et de minérales, dans les langues latine–roumaine–hongroise–allemande et roumaine–latine–hongroise–allemande. Pour réaliser l'inventaire des termes, l'auteur a valorisé les sources allemandes. Les termes roumains utilisés ont, pour la plupart d'entre eux, un caractère populaire, Șincai montrant une réticence envers la terminologie livresque, néologique. Même s'il n'a pas été envoyé à l'imprimerie, l'ouvrage a été valorisé pour achever le *Dictionnaire de Buda*.

Le dernier travail lexicographique multilingue paru avant le *Dictionnaire de Buda* est *Dicționar rumanesc, lateinesc și unguresc* [Le Dictionnaire roumain, latin et hongrois], en deux volumes, réalisés par l'initiative d'Ioan Bobb et imprimé à Cluj entre les années 1822 et 1823. C'est le premier ouvrage lexicographique de grandes proportions imprimé, qui comprend environ 11000 mots. La liste de mots comprend de nombreux noms propres, des mots latins et des termes régionaux d'origine hongroise. Les unités phraséologiques, fréquentes dans le travail, figurent comme des articles indépendants, après le mot-titre important, sans être incluses dans le corpus qui traite le mot. Pour certains mots, peu usuels, l'auteur indique de brèves définitions ou des correspondances synonymiques (ex. *armentariu*: „pastoriu e vite” [berger d'animaux]; *ager*: „iute, grabnic, străduitoriu” [rapide, agile]). Le dictionnaire est entière-

ment imprimé avec des caractères latins, et la forme des mots roumains est généralement latinisée et étymologiste. Le travail ouvre la vague du purisme extrême, par l'isolation, dans une annexe de la fin du IIe volume, des mots populaires roumains d'origine non-latine. L'auteur a utilisé, dans l'élaboration du dictionnaire, trois sources lexicographiques hongroises: la source principale est constituée par le *Dictionnaire* de Fr. Páriz Pápai, dans l'édition de 1708, et les deux autres sources secondaires sont le *Dictionnaire* de A. Molnár et celui de I. Márton (Seche 1966). L'ouvrage d'Ioan Bobb se prouve être inférieur, de point de vue de la réalisation, au *Dictionnaire de Buda*.

5. Le début de la lexicographie roumaine moderne

Le dictionnaire de Buda, qui a marqué le début de la lexicographie roumaine moderne, a paru en 1825, à l'Imprimerie de l'Université de Buda, sous le titre original, *Lesicon românesc–latinesc–unguresc–nemțesc, care de mai mulți autori, în cursul a trideci și mai multor ani s-au lucrat seu: Lexicon valachico–latino–hungarico–germanicum quod a pluribus auctoribus decursu triginta et amplius annorum elaboratum* [Dictionnaire roumain–latin–hongrois–allemand, auquel plusieurs auteurs, au long de plus de trente ans, ont travaillé]. Dictionnaire polyglotte en quatre langues, le *Dictionnaire de Buda* représente le premier dictionnaire explicatif et étymologique de la langue roumaine qui est sorti de l'imprimerie. L'élaboration du *Dictionnaire de Buda* a été un processus de longue durée, qui s'est réalisée dans plusieurs étapes. La première étape vise l'activité lexicographique de l'érudite Samuil Micu-Klein, reviseur et correcteur des livres roumains à l'Imprimerie de Buda. Le 10 mars 1801, Micu achève la rédaction d'un dictionnaire bilingue: *Dictionarium valachico-latinum* [Dictionnaire valaque–latin]. Deux ans après, l'auteur se propose de transformer le travail dans un dictionnaire polyglotte, en quatre langues, ajoutant les langues allemande et hongroise. Pour cela, Micu fait appel à l'aide de deux spécialistes: le professeur I. Molnar-Halitzki pour l'allemand, respectivement Virág Benedek, pour le hongrois. En 1805, le dictionnaire polyglotte est achevé, cet ouvrage étant considéré comme la première variante du *Dictionnaire de Buda*, qui allait être imprimé en 1825. À la demande du directeur des écoles de Timișoara, Samuil Micu-Klein ajoute à l'ouvrage les correspondants en lettres latines des mots roumains (rédigés dans la première variante avec des caractères cyrilliques). Le travail est donné à l'Imprimerie de l'Université de Buda qui, au début de l'année 1806, annonce l'impression de deux travaux lexicographiques constitués par Micu-Klein: *Dictionarium latino–valachico–germanico–hungaricum* [Dictionnaire latin–valaque–allemand–hongrois] et *Dictionarium valachico–latino–germanico–hungaricum* [Dictionnaire valaque–latin–allemand–hongrois]. Il s'agit, en réalité, du même ouvrage, seulement l'ordre des langues composantes est modifié. Après la mort de Micu, l'imprimerie de l'Université de Buda confie le travail d'achèvement du dictionnaire à Vasile Coloși, le curé de la paroisse de Săcărîmb. Celui-ci avait donné lui-même à l'imprimerie, vers 1805, un *Dictionnaire roumain–latin–hongrois–*

allemand en vue de l'impression. Coloși unit les deux dictionnaires, en changeant l'ordre des deux dernières langues du travail de Micu. Fragments de la rédaction de Coloși, découverts dans l'église épiscopale d'Oradea, montrent que l'auteur a profondément modifié la version de Micu. La version de 1810 est amplifiée, et certains des articles rédigés par Coloși restent les mêmes jusqu'à l'édition imprimée du *Dictionnaire de Buda*. Après la mort de Coloși, la suite du travail est reprise, en 1815, par le chanoine d'Oradea Ioan Corneli. Avec le soutien financier de Samuil Vulcan, Corneli part à Buda, où il révisé et amplifie le dictionnaire à partir de la lettre *I*, ajoutant de nombreux termes technico-scientifiques. À partir des fragments, les lexicographes ont reconstitué la contribution de Corneli au dictionnaire: les articles rédigés par celui-ci sont beaucoup plus étendus que dans le *Dictionnaire de Buda*, les définitions étant parfois interminables. La nouveauté que Corneli apporte vise les indications étymologiques, qui détiennent un pourcentage plus important dans sa version que dans la forme finale du *Dictionnaire*. Corneli est obligé de rentrer dans le pays et en 1820 la révision de l'ouvrage est confiée à Petru Maior. Même s'il a peu travaillé au dictionnaire, car il s'est éteint en 1821, la contribution de Maior est extrêmement consistante, surtout en ce qui concerne l'achèvement des étymologies. La révision et l'achèvement du vaste travail ont été réalisés par le prêtre Ioan Teodorovici et par son frère, Alexandru Teodori, docteur en médecine et philosophie (Pricop et al. 2017: 104-106).

Dictionnaire polyglotte, en quatre langues, *Le Dictionnaire de Buda* représente le premier dictionnaire explicatif et étymologique de la langue roumaine qui a été imprimé. Presque tous les mots du dictionnaire présentent de brèves indications sémantiques (périphrases ou synonymes), cet aspect offrant au travail un caractère explicatif. Par rapport à la version de Micu-Klein, la forme finale du *Dictionnaire de Buda* comprend aussi des informations grammaticales: chaque mot-titre est encadré dans la catégorie morphologique correspondante, on atteste le pluriel des noms et des adjectifs, l'infinitif long et le participe des verbes. Les mots polysémantiques sont groupés par sens, à différence de la version de Micu-Klein qui concentrait les nuances sémantiques dans une énumération unique. Les sens des mots sont présentés en diachronie, du sens étymologique à ceux récents. La valeur du dictionnaire accroît aussi par le grand nombre d'expressions comprises dans la structure des articles, accompagnées souvent par des définitions. La tendance puriste des auteurs est reflétée par l'indication de manière systématique de l'étymon pour les mots d'origine latine et par le contournement des indications étymologiques pour les mots non-latins.

«La graphie des mots-titre était ainsi destinée non seulement à identifier l'origine latine de beaucoup de mots (les termes ayant d'autres origines étaient écrits de manière normale, analogique), mais aussi de faciliter l'unification formelle de notre langue de culture» (Chivu 2012: 50).

Malgré les exagérations puristes qu'il reflète, *Le dictionnaire de Buda* contient plus de mille étymologies latines restées valables jusqu'à nos jours, contribuant

énormément au développement de l'étymologie roumaine (Seche 1966). L'influence du *Dictionnaire de Buda* sur la lexicographie roumaine a été énorme, presque tous les dictionnaires du XIXe siècle se sont rapportés à celui-ci comme à un étalon. Le prestige de l'ouvrage a conduit à la propagation du courant latin dans la plupart des travaux lexicographiques de la Valachie et de la Moldavie, jusqu'après 1870.

Au niveau de la macrostructure, *Le Dictionnaire de Buda* est organisé de la sorte: *Cuvânt către cititori* [Mot pour les lecteurs] (en latin et en roumain — avec alphabet latin et cyrillique, p. 3-8), *Prefața la Ortografie* [Préface pour l'Orthographe] (en latin, p. III-VIII), *Ortografia lui P. Maior* [L'Orthographe de P. Maior] (en latin, p. 1-53 p.), *Dialogul pentru începutul limbii române* [Dialogue pour le début de la langue roumaine] (en roumain, avec alphabet latin et cyrillique, p. 54-102), *Extrasul cu regulile de pronunțare* [Extrait avec les règles de prononciation] (p.103) et *Dictionarul propriu-zis* [Le dictionnaire proprement-dit] (p. 1-771). Le dictionnaire proprement-dit, étendu sur 771 pages, expose par ordre alphabétique les unités lexicales du vocabulaire général de la langue roumaine au début du XIXe siècle, et la présentation des correspondances dans les langues composantes (latin, hongrois, allemand) se réalise de manière groupée, dans le corpus des articles. L'inventaire des entrées comprend de nombreux syntagmes, noms propres, diminutifs, appellations populaires, instruments grammaticaux du type des prépositions et conjonctions, interjections et éléments d'argot. Au niveau de la microstructure, les articles du *Dictionnaire* sont organisés ainsi: mot en roumain (à initiale majuscule, orthographié avec alphabet latin), suivi des informations grammaticales, mot-titre orthographié avec alphabet cyrillique, définition en langue roumaine suivie des correspondants en latin, hongrois et allemand, selon le modèle:

Ispravnicu [gestionnaire], *m. pl. ci. ..., ci/, subst. 1*) cui sê incredințeze purtare de grije a quâruiua lucru [à qui on confie le soin d'une chose]: *actor, curator, administrator*: gomdviselō, elintézō: der Anstalter, Verrichter, Besorger, Verwalter. 2) Capitanulu séu maimarele quaruiva țenutu, mai in Țerile românești [Capitaine ou le chef d'une région dans les Pays Roumains]: *Praefectus circuli, Inspector districtus*: kapitánya egy kerületnek, vagy járásnak: der Kreișhauptmann, Kreișamtman.

Taină [secret], *f. pl. i...., subst. 1*) adeq. lucru ascunsu, misteriu [chose cachée, mystère]: *mysterium, res occulta, arcana*: titok: daș Geheimniß, die Heimlichkeit. — in taină [en secret]: *adv. clam, occulta, arcane, tecte*. titkon: heimlich, in Geheim. — de taină: *adj. arcanus, secretus, occultus*: titkos: heimlich, geheim. 2) a Biséricii [sacrement]: *Sacramentum*: Szentség: daș Sacrament.

Le dictionnaire comprend une section introductive généreuse, de 103 pages, qui inclut des informations concernant l'historique de la démarche lexicogra-

phique, les contributions des auteurs et leur conception concernant l'origine et le développement de la langue roumaine. Les textes qui sont en préface du dictionnaire exposent et justifient les intentions lexicographiques des auteurs. La manière de sélection et d'organisation du matériel lexical relève les tendances puristes de ceux-ci, d'éliminer les éléments considérés étrangers de l'origine et de l'esprit de la langue roumaine, fait reflété dans la pratique de l'indication systématique de l'étymon pour les mots d'origine latine et le contournement des indications étymologiques dans le cas des mots non-latins. Le caractère non-systématique du travail est généré, d'une part, par le grand nombre des auteurs qui ont contribué à la rédaction, et, d'autre part, par la longue période d'élaboration. L'hétérogénéité du dictionnaire est réfléchiée par les fluctuations de traitement des articles, par la manière dans laquelle l'information est structurée et par la présence dans l'ouvrage de plusieurs systèmes orthographiques qui rendent la lecture difficile.

Dans une époque où la lexicographie roumaine était très faiblement représentée, *Lexicon valachico-latino-hungarico-germanicum* [Le dictionnaire valaque-latin-hongrois-allemand] impressionne par: le nombre des entrées, le caractère encyclopédique des articles, la riche illustration des relations syntagmatiques et paradigmatiques que les mots-titre comportent, les informations étymologiques offertes, les affinités établies entre les langues romanes comprises dans le dictionnaire, la manière de traiter l'inventaire des entrées. L'ouvrage reflète les pratiques lexicographiques de l'époque, portant l'empreinte du temps et de l'espace où il a été créé. Malgré les exagérations puristes qu'il reflète, le dictionnaire s'élève au niveau des travaux lexicographiques étrangers, l'influence de l'ouvrage sur la lexicographie roumaine étant énorme: presque tous les dictionnaires du XIXe siècle se rapportent à celui-ci comme à un étalon.

Le Dictionnaire de Buda peut être consulté de nos jours en format digital aussi, grâce à une équipe de chercheurs conduite par Maria Aldea. L'édition corrigée et adaptée électroniquement pour la consultation en ligne du Dictionnaire de Buda reproduit le texte du volume imprimé en 1825, à l'imprimerie de Buda. En plus, l'édition électronique inclut: les traductions en roumain pour la Préface pour l'Orthographe et l'Orthographe de Petru Maior, un indice des mots des dialectes transdanubiens, de l'italien, français, espagnol et grec, un indice des proverbes et des auteurs/ œuvres cités et un inventaire des mots qui regroupe les mots-titre, les doublets graphiques et lexicaux des mots-titre et les mots-titre des entrées secondaires.

6. Conclusions

Corrélée avec le paysage de la lexicographie internationale, la lexicographie roumaine multilingue enregistre à son tour deux prototypes: le dictionnaire polyglotte, qui établit l'inventaire des entrées à partir du vocabulaire général de la langue, et le dictionnaire de spécialité ou terminologique multilingue, qui traite les terminologies correspondant à certains domaines scientifiques et tech-

niques particuliers. Entre les deux catégories de travaux lexicographiques il n'y a pas de limites fermes, la pratique lexicographique mettant en évidence, surtout dans l'étape de début, des contaminations, des immixtions de certaines préoccupations de terminologie dans le cadre des dictionnaires polyglottes.

Une grande partie des initiatives lexicographiques se sont arrêtées au stade de projet, mais elles restent importantes pour l'histoire de la lexicographie roumaine, par les objectifs et les conceptions lexicographiques exposés par leurs initiateurs. L'appel à des sources indirectes a cependant permis la découverte de certains traits spécifiques à la lexicographie slave-roumaine dans les premiers siècles: le manque de tout appareil scientifique, la limitation à de simples registres multilingues, où l'on indique les équivalences dans les autres langues pour le correspondant de la langue roumaine, le caractère hétérogène des inventaires des entrées, l'absence de certains critères lexicographiques pour organiser le matériel lexical, les tendances puristes présentes dans la création des listes de termes et des indications étymologiques sommaires sont quelques-unes des caractéristiques que les ouvrages lexicographiques roumains parus jusqu'à la fin du XVIIIe siècle relèvent.

La lexicographie roumaine ancienne s'est rapportée aux traditions lexicographiques de prestige (comme celle slavonne), et les ressorts pragmatiques de la parution des premiers projets lexicographiques sont étroitement liés aux besoins de compréhension des textes religieux qui circulaient, en slavonne ou en traduction. Les premiers projets lexicographiques appartiennent à des érudits conscients de l'utilité de ces instruments et dont la vision lexicographique porte l'empreinte des influences culturelles de l'époque. Les glossaires comprenaient, d'habitude, des termes usuels, nécessaires à parcourir les textes bibliques et liturgiques, et ils étaient destinés aux gens des églises et aux traducteurs, assurant la fidélité de ceux-ci envers la lettre du texte religieux, considéré saint. Sans avoir à la base la conception rigoureuse, scientifique, d'une école lexicographique, les premiers dictionnaires roumains s'avèrent être hétérogènes tant en ce qui concerne le matériel lexical inventorié, comme du point de vue de la manière de traitement lexicographique. Reflétant le degré de développement de la langue roumaine à l'époque et les tendances puristes partagées par certains auteurs, les premiers projets lexicographiques créent les prémisses pour le développement de la lexicographie roumaine moderne.

Dans les conditions où l'accès à certains travaux lexicographiques est limité à cause de plusieurs raisons, une série de projets de recherche qui favorisent l'accès au format numérique des travaux lexicographiques de référence pour l'espace roumain contribuent non seulement à l'alignement aux tendances de la lexicographie européenne moderne, mais aussi au développement des études metalexigraphiques, par les instruments de travail qu'ils mettent à portée de la main des chercheurs. Des exemples en ce sens sont offerts par les projets *Lexiconul de la Buda (1825). Ediție emendată și prelucrată electronic pentru consultare on-line* [Le dictionnaire de Buda (1825). Edition émondée et transposée en format numérique pour l'étude en ligne] et *Primele dicționare bilingve românești (secolul al XVII-*

lea). *Corpus digital prelucrat și aliniat (eRomLex)* [Les premiers dictionnaires bilingues roumains (le XVIIe siècle). *Corpus numérique transformé et aligné (eRomLex)*.]

Remerciement

Le présent article s'appuie sur les résultats d'un projet de recherche qui a visé l'élaboration d'une bibliographie analytique et critique de la lexicographie polyglotte roumaine du XVIIe au XXIe siècle. À partir d'un corpus composé de catalogues et bibliographies représentatifs pour le livre roumain, on a créé une liste d'environ 450 titres qui couvre la période comprise entre la fin du XVIIIe siècle et le début du XXIe siècle (Pricop et al. 2017).

Références

Dictionnaires

- Antal, Veszelski.** 1798. *A növény-planták országából való erdei és mezei gyűjtemény, vagy-is fa- és fűszer könyv* [O colecție din lumea plantelor de câmp și de pădure adică o carte a arborilor și plantelor]. Pest.
- Benkő, József.** 1783. *Nomina vegetabilium* [Glosarul latin-maghiaro-român]. *Magyar könyv-ház. Vol. II:* 405-432. Pozsony: Landerer Mihály.
- Cavallioti, Theodor.** 1770. [Vocabular grec-aromân-albanez]. *Protopeiria* [Prima învățătură]. Veneția.
- Crișan (Körösi), Ștefan.** 1803. *Covventario românesc-latenesc-unguresc* [Dicționarul lui Körözi literaturor român din Wasarhely], [Lexicon român-latin-maghiar], [f.e.], [f.l.].
- Ianovici, Nicolae.** 1821. *Diccionariu tru cince limbe: ellinescu, grecescu, rumanescu, nemcescu shii (=și) madsarescu* [Lexiconul elin-grec modern-aromân-german-maghiar], 1821 (ms. BAR, cota 1822).
- Iorgovici, Paul.** 1800. [Dicționar în patru limbi (român-german-francez-latin)]. [f.l.], [1800].
- Kräutner, Samuel.** 1793. Nachtrag zu des Herrn Profisor Sigerus Verzeichniss wildwachsender Siebenbürgischer offizineller Pflanzen. *Siebenbürgische Quartalschrift, Heft IV, Hermannstadt:* 351-356.
- Leonhard, Daniel Joseph.** 1818. [Vocabular german-saxon-maghiar-român]. *Lehrbuch zur Beförderung der Kenntniß von Siebenbürgen*. Hermannstadt.
- Micu-Klein, Samuil.** 1805. *Dictionarium latino-valachico-germanico-hungaricum/Dictionarium valachico-latino-germanico-hungaricum*. Buda: Types et Sumtibus Typogr. Regiae Univ. Hungaricae.
- Moscopoleanul, Daniil.** 1764. *Lexicon tetraglosson. Învățătură introducătoare*. Moschopolis.
- Neustädter, Mih.** 1795. *Consignatio specifica omnium Plantarum quae in Magno hocce Principatu sponte sua crescunt. Opinio deputationis regnicolaris systematicae in Cameralibus ordinatae circa promotionem agriculturae, manufacturarum comerciiue transilvanici depromta*. Cluj.
- Praedetis, Aurelius Antoninus.** 1793. *Dictionarii trium lingvarum germano-latina et daco-romana* (ms. autograf din 1793; Biblioteca Academiei Române, Filiala Cluj, cota 457-459).
- Sigerus, Petrus.** 1791. *Verzeichniss der in Siebenbürgen wildwachsenden offziellen Pflanzen*. *Siebenbürgische Quartalschrift, Heft III, Hermannstadt:* 315-316.
- Șincai, Gh.** 1810. *Vocabularium pertinens ad tria regna naturae*, 1808-1810, (ms. nr. 482, BAR Cluj).
- *** 1687-1701. *Dicționarul român-latin-maghiar* [Dicționarul marsilian], (ms. 116, colecția Marsigli, Biblioteca Universității Regale din Bologna).

- *** 1800. [*Dicționar francez–grec–român*], [f.l.], [1800] (ms. Biblioteca lui Gh. A. Sturza, Miclăușani), reprodus de Nicolae Iorga. *Revista istorică* VII, 1921: 106-112.
- *** 1822–1823. *Dicționarul rumanesc, latinesc și unghuresc*. Dein orenduiala excelenții sale Preosfințitului Ioan Bobb, vlădicul Fagărașului asediat și cu venitu cassi clerului typărit în doao tomuri, Cluj, cu Tipariul Tipografii Colegiumului Reformaților, prin Stephan Török, I: A–L, 1822; II: M–Z.
- *** 1825. *Lesicon românesc–latinesc–unguresc–nemțesc, care de mai mulți autori, în cursul a trideci și mai multor ani s-au lucrat seu: Lexicon valachico-latino-hungarico-germanicum quod a pluribus auctoribus decursu triginta et amplius annorum elaboratum est*. Buda: Typis et Sumtibus Typographiae Regiae Universitatis Hungaricae [Bibliografia DLR = LB].
- *** 1930. *Il Lexicon Marsilianum. Dizionario latino–rumeno–ungherese del sec. XVII*. Studio filologico e testo de Dr. Carlo Tagliavini. București: Cultura Națională.

Littérature secondaire

- Brown, Keith (Éd.)**. 2006. *Encyclopedia of Language & Linguistics*. Deuxième édition. Oxford: Elsevier.
- Chivu, Gheorghe**. 2012. *Lexiconul de la Buda*, primul dicționar modern al limbii române. *Analele Universității „Alexandru Ioan Cuza” din Iași Secțiunea IIIe. Lingvistică* LVIII: 45-56.
- Coman, Lupu**. 1999. *Lexicografia românească în procesul de occidentalizare latino–romanică a limbii române moderne (1780–1860)*. București: Logos.
- Delisle, Jean et Judith Woodsworth (Éds.)**. 1995. *Translators through History*. Amsterdam: John Benjamins.
- Gînsac, Ana-Maria et Mădălina Ungureanu**. 2018. La lexicographie slavonne-roumaine au XVIIe siècle. Adaptations roumaines d'après le *Leksikon slavenorosskij* de Pamvo Berynda. *Zeitschrift für romanische Philologie* 134(3): 845-876.
- Guérard, Françoise**. 2015. *Le dictionnaire, miroir du monde, mémoire des hommes*. Paris: Honoré Champion Éditeur.
- Kelemen, B.** 1962. Începuturile lexicografiei Românești. *Cercetări de lingvistică* VII(1): 89-92.
- Pricop, A. (Coord.) et al.** 2017. *Lexicografia poliglotă în secolele al XVII-lea–al XXI-lea: liste, vocabulare, glosare, dicționare, lexicoane, tezaure care conțin limba română. Bibliografie*. Iași: Editura Universității 'Alexandru Ioan Cuza' din Iași.
- Seche, M.** 1966, 1969. *Schiță de istorie a lexicografiei române* [Brief History of the Romanian Lexicography]. (2 vol.). București: Editura Științifică.
- Strungaru, Diomid**. 1966. Începuturile lexicografiei române. *Romanoslavica* XIII: 141-158.

Projets de recherche

- Lexiconul de la Buda (1825)*. Ediție emendată și prelucrată electronic pentru consultare on-line [Le dictionnaire de Buda (1825). Edition émondée et transposée en format numérique pour l'étude en ligne]. <https://www.bcucuj.ro/lexiconuldelaBuda/site/login.php>.
- Primele dicționare bilingve românești (secolul al XVII-lea)*. Corpus digital prelucrat și aliniat (eRomLex) [Les premiers dictionnaires bilingues roumains (le XVIIe siècle). Corpus numérique transformé et aligné (eRomLex)]. <http://www.scriptadacoromanica.ro/bin/view/eRomLex/>.

Improving the Compilation of English–Chinese Children's Dictionaries: A Children's Cognitive Perspective

Dan Zhao, *School of Foreign Studies, Guangdong University of Finance
and Economics, Guangzhou, P.R. China (ddanzhao@163.com)*

and

Hai Xu, *Centre for Linguistics and Applied Linguistics, Guangdong
University of Foreign Studies, Guangzhou, P.R. China
(Corresponding Author, xuhai1101@gdufs.edu.cn)*

Abstract: Children's dictionaries have existed for more than one thousand years in China, and play an important role in children's learning. However, many of those produced in China are deficient in the selection of the wordlist, in exemplification, and in definition. This paper aims at improving the compilation of English–Chinese children's dictionaries (ECCDs) from a children's cognitive perspective. Children's dictionaries should not only be an abridgement or simplification of dictionaries for adults, because their target user group is immature, uninformed and untrained children. Informed by some innovations in current English learner's dictionaries, this paper proposes that the making of ECCDs needs to be improved in the following aspects. Firstly, instead of lexicographers' intuition, the selection of headwords should be based on an English corpus for Chinese children. Secondly, the words used in examples should be congruent with children's limited cognitive and learning abilities. Thirdly, a multifaceted method of explanation should be provided in order to assist children in understanding the meaning of headwords.

Keywords: ENGLISH–CHINESE CHILDREN'S DICTIONARIES, USER FOCUS, CHILDREN'S CORPUS, HEADWORDS, CHILDREN'S COGNITIVE ABILITY, ILLUSTRATIONS, EXAMPLES

Opsomming: Verbetering van die samestelling van Engels–Chinese kinderwoordeboeke: Die kognitiewe perspektief van 'n kind. Kinderwoordeboeke bestaan reeds meer as 'n duisend jaar in China, en speel 'n belangrike rol in die leerproses van kinders. Baie van dié wat in China saamgestel word, is egter ontoereikend in die keuse van die woordelys, in die gebruik van voorbeelde en in die definisie. In hierdie artikel word gepoog om die samestelling van Engels–Chinese kinderwoordeboeke (ECKWe) vanuit die kognitiewe perspektief van 'n kind te verbeter. Kinderwoordeboeke behoort nie net 'n beknopte uitgawe of vereenvoudiging van woordeboeke vir volwassenes te wees nie, want hul teikengebruikersgroep is onvolwasse, oningeligte en onopgeleide kinders. Aangespoor deur sommige vernuwings in huidige Engelse aanleerderswoordeboeke, word daar in hierdie artikel voorgestel dat die samestelling van ECKWe ten opsigte van die volgende aspekte verbeter moet word. Eerstens behoort die seleksie van trefwoorde op 'n Engelse korpus vir Chinese kinders en nie op leksikografiese intuïsie gebaseer te word nie. Tweedens

behoort die woorde wat in voorbeelde gebruik word, in ooreenstemming met kinders se beperkte kognitiewe en aanleervermoëns te wees. Derdens behoort 'n veelvlakkige metode vir verduideliking verskaf te word om sodoende kinders te help om die betekenis van trefwoorde te kan begryp.

Sleutelwoorde: ENGELS–CHINESE KINDERWOORDEBOEKE, GEBRUIKERSFOKUS, KINDERKORPUS, TREFWOORDE, KINDERS SE KOGNITIEWE VERMOËNS, ILLUSTRASIES, VOORBEELDE

1. Introduction

Children's dictionaries not only have a long history in China but also play a vital role in children's learning. They are intended for the edification of children. However, children's lexicography has received less attention than adult-oriented lexicography.

Dictionaries that are designed explicitly for children have a shorter word-list consisting of the most "important" words of the language, excluding regionalisms, archaisms, slang words, etc. They also use different techniques for the explanation of meaning, including, for instance the use of illustrative examples without definitions. They are often interesting for the metalexigrapher, but oddly enough they have never been the object of serious research (Béjoint 2000: 40).

For many languages, there are few or even no children's dictionaries. For example, in Slovenia, no dictionary has been designed for the school population (Rozman 2008); and in Greece, few children's dictionaries have been published (Gavrilidou et al. 2008). One of the reasons why many publishers hesitate to allocate resources to the development of school dictionaries is the relatively poor sales and corresponding profits coming from this category of dictionaries (Tarp and Ruiz Miyares 2013). Although the situation with lexicographic publishing has changed remarkably over the last decade, with digital resources becoming almost ubiquitous for adults, paper dictionaries continue to be used in schools because children are forbidden to use electronic products in school (at least in China). Nevertheless, revenues from school dictionaries are still not significant for publishers. When at home, children tend to use digital resources instead of print dictionaries.

Children's dictionaries were found to have some apparent deficiencies (Turini et al. 2000, Verburg 2006, Rozman 2008, Gavrilidou et al. 2008, Potgieter 2012, Kosch 2013, Sene Mongaba, B. 2016). In children's dictionaries, the list of headwords is often selected at random (Verburg 2006). Definitions are too difficult for this target audience (Rozman 2008; see also De Schryver and Prinsloo 2011 on definitions in Van Dale dictionaries). Lexicographic and typographical codes are incomprehensible to children (Rozman 2008). Some of the children's dictionaries are only a revised and simplified version of the existing unabridged dictionaries, such as monolingual children's dictionaries in Italy (Turini et al. 2000) and in Greece (Gavrilidou et al. 2008), and a bilingual Afrikaans

school dictionary in South Africa (Potgieter 2012). Example sentences supply the user with little or no contextual guidance (Potgieter 2012). Above all, most of children's dictionaries fail to take into due consideration the abilities and interest of young users. As a result, they can hardly meet the specific needs of the intended target users and are often more of an obstacle than an aid.

The making of monolingual dictionaries for children thrived in China. The oldest extant monolingual school dictionary in China is *Ji Jiu Pian* (《急就篇》, *The Instant Primer*) which was compiled by Shi You (史游) in the Western Han Dynasty in the first century BC. There were some other monolingual school dictionaries published before and after it, such as *Cang Jie Pian* (《仓颉篇》) in the Qin Dynasty between 221 and 207 BC, *Fan Jiang Pian* (《凡将篇》) in the Western Han Dynasty about 150 BC, *Xun Zuan Pian* (《训纂篇》) in the Western Han Dynasty between 53 and 18 BC, and *Pang Xi Pian* (《滂喜篇》) in the Eastern Han dynasty in the first century AD. However, these children's dictionaries except for *Ji Jiu Pian* were lost. In 1953, *Xinhua Zidian* (《新华字典》, *Xinhua Dictionary of Chinese Characters*), the first modern Chinese monolingual dictionary for children, was published. It is the most influential and authoritative Chinese children's dictionary, and has the highest circulation of any dictionary in the world. It has been revised many times and reprinted more than two hundred times. The latest (i.e., 12th) edition was published in August, 2020.

By comparison, the number of bilingual dictionaries for Chinese-speaking children, mainly English–Chinese Children's Dictionaries (ECCDs), are few. For example, the Foreign Language Teaching and Research Press, one of the top publishers in China, has published 47 English–Chinese learner's dictionaries, but of these only four are bilingual dictionaries for children.

ECCDs can be divided into three main types, based on the cognitive abilities of children at different ages: dictionaries for preschool children (2–6 years old), those for elementary school children (6–12 years old), and those for children in junior high school (12–16 years old). However, our investigation into the main ECCDs shows that some current ECCDs have no clear target group of users. Some dictionaries include 'children' in their titles but do not indicate for which age bracket they are intended. An analysis of these dictionaries shows the following problems: Firstly, the number of headwords varies widely among ECCDs for the same age group. For example, *Zhongguo Xiaoxuesheng Yingyu Xuexi Cidian* (*An English–Chinese Dictionary for Primary School Learners in China*) contains approximately 4,000 headwords while *Xinbian Xiaoxuesheng Yinghan Cidian* (*The Brand-New English–Chinese Dictionary for Primary School Students*) only comprises approximately 1,500 headwords. Secondly, some words used in examples are beyond children's knowledge. For instance, in *Xiaoxuesheng Yinghan Cidian* (*An English–Chinese Dictionary for Primary School Students*), there is an example under the headword **also**: *She's a talented singer and also a fine actress*. But *talented* is neither a headword nor a word which Chinese pupils are often exposed to. If children have difficulty in understanding the words in examples, they will not use the dictionary with enjoyment and enthusiasm. Thirdly, the headwords in these dictionaries are arranged differently: some are arranged

thematically such as in *Ertong Yinghan Baike Tujie Cidian* (*My First Visual Dictionary*) while others are arranged alphabetically such as in *Tujie Ertong Yinghan Cidian* (*Illustrated Children's English–Chinese Dictionary*). Finally, the method of explanation used as in dictionaries for adults (such as only translation equivalents) may not be applicable to dictionaries for children of the three different categories, because children's understanding and cognitive abilities are very limited. Even in ECCDs, the methods of explanation should vary greatly according to the cognitive abilities of children at different ages.

Synthesizing these problems, this paper will address the following research questions:

- (1) In what ways should a children's dictionary accommodate its target user group?
- (2) How should the headword list of a children's dictionary be developed?
- (3) Should the headword list of a children's dictionary be arranged alphabetically or thematically?
- (4) How should word meanings be explained in a children's dictionary?

Section 2 discusses how to define the user group according to Piaget's cognitive development theory. Section 3 proposes that when selecting headwords, an English corpus for Chinese children should be taken into account. Section 4 compares the thematic arrangement with the alphabetical order in children's dictionaries. Section 5 discusses various ways of explaining word meanings.

2. Defining the user group according to Piaget's cognitive development theory

'In planning the preparation of a dictionary, it is of vital importance to decide as early as possible what character the dictionary should have, in type, size, users and so on' (Zgusta 1971: 221). Producers and editors need to identify the specific group of users when they commence designing a dictionary. Deciding upon the intended target user is the first stage of dictionary production in a four-stage process (Rundell 2010: 367). It is noted that 'during the last few years lexicographers have become more and more aware of the importance of the so-called user perspective — determining who the intended target user is and what his or her specific needs are with regard to the dictionary' (Potgieter 2012: 262). More and more producers and editors of ECCDs also become aware that they need to define the user group when they initiate a dictionary project.

All dictionaries are written for a specific user group, and the content and presentation must therefore be directed/aimed at that specific target group (Cheng 2001, Potgieter 2012). 'Child', a general term, usually refers to people younger than 16. In China, children are generally divided into two age groups. The first group is from 2 to 6 years old, and the second from 6 to 12 years old (Yu 1998: 960). Those aged between 12 and 18 are called teenagers. This cate-

gory is only applicable to laws in China, different from the category based on children's cognition. But the age brackets of children vary according to countries. In Britain, 'child' refers to young persons under the age of 18 according to the Children Act of 1948, while it refers to young persons under the age of 16 in the Education Act of 1994. Therefore, there is no clear dividing line according to age for the term 'children'. In terms of lexicography, young people under the age of 16 are generally considered children.

In fact, children at different ages, even if all under the age of 16, are very different in terms of their cognitive development. According to Piaget's cognitive development theory (Piaget 1969), children's cognitive development can be divided into four stages. The first stage (birth to 2 years) is the sensorimotor stage in which children experience the world through their senses and actions. Object permanence and stranger anxiety are its main developmental phenomena. The second stage (2 to 6 years) is the preoperational stage in which children start to be able to represent things with words and images. The preoperational stage is characterized by three developmental phenomena: pretend play, egocentrism, and language development. The third stage, from 7 to 11 years old, sees children able to think logically about concrete events and grasp concrete analogies. It is called the 'concrete operational' stage, and is characterized by two developmental phenomena: conservation and mathematical transformation. Piaget refers to the fourth stage (12 onwards to adulthood) as the formal operational stage. In this stage, children can think about hypothetical scenarios and process abstract thoughts. It has two developmental phenomena: abstract logic and potential for mature, moral reasoning.

In the first stage, children only sense the world by natural feeling, touch and behaviour, and most children cannot use many words or use dictionaries. Therefore, a dictionary does not make sense for them. When they reach the second stage, children's language abilities develop very quickly by matching words and images or things. In the third stage, children begin to have logical thinking ability. The cognitive and learning abilities of children in the last stage approximate those of adults.

The focus of this paper is on children in the second and third stages (2 to 6 years, and 6 to 12 years, respectively). Although both are referred to as 'children', they differ considerably in cognitive and learning abilities. The word 'children' in the titles of many ECCDs is really ambiguous or unclear, since children's dictionaries should rather have two explicit categories: for younger children (2 to 6 years) and for primary school students (6 to 12 years old). At the beginning of a children's dictionary project, it is essential for producers and editors to identify its prospective users in terms of these age groups: either younger children or primary school students. The dictionaries for children over the age of 12 also vary greatly with adult dictionaries although they have fewer difficulties in using adult dictionaries than those under the age of 12. Headwords such as swear words should be excluded; slang should be selective; definitions are less complicated; and it is easy for children to carry their dictionaries around.

3. Selecting headwords based on a children's English corpus

After determining the group of intended users, the issue of how to select the headword list is the next challenge for developers of ECCDs. What words are children most likely to be exposed to? Which words should be included and which excluded? The criteria for lexical coverage should take into account the evidence from an English corpus for Chinese children. 'The children's corpus can be used to help lexicographers make decisions about headwords' (Wild et al. 2013: 190).

The Oxford Children's Corpus (OCC), a corpus of over 30 million tokens of writings targeted at 5–14 year-olds, contains a wide range of fiction, non-fiction, children's writing, etc. (see Table 1).

Table 1: The make-up of the OCC (Adapted from Wild et al. 2013: 193)

Genre	Tokens
fiction	23,139,119
non-fiction	6,755,691
children's writing	1,421,720
unclassified	397,352
<i>Total:</i>	<i>31,713,882</i>

Period	Tokens
pre-1900	4,429,132
1900–1964	10,139,421
1965–1999	2,058,857
2000–present	14,816,110
unclassified	270,362
<i>Total:</i>	<i>31,713,882</i>

Key Stage	Tokens
1 (5–7 years old)	1802,762
2 (7–11 years old)	14,003,042
3 (11–14 years old)	7,772,753
unclassified	8,135,325
<i>Total:</i>	<i>31,713,882</i>

The subcorpus of children's writings (approximately 1.4 million tokens) is made up of material from websites where children have posted reviews, poems and stories. There are plans to expand this part of the corpus, so that the language that children use can be compared with the language to which they are exposed. Hence, typical spelling and usage errors committed by children can be identified (Wild et al. 2013: 194). Since its creation in 2006, lexicographers have used OCC to make decisions about headwords in writing dictionaries for children. Wild et al. (2013: 207-208) offer a case study of how the OCC helped to select the list of musical instruments in children's dictionaries. The first step was obtaining a list of musical instruments in order of frequency from the OCC. The raw data had to be manually checked due to some polysemous words in the list not being in the required sense. There are some surprising findings in the list, such as the high frequencies of the words *lute* and *lyre*, which are mainly found in history texts for KS3 (11 to 14 years old). This type of findings from a children's corpus can help lexicographers make decisions about which words should be included in children's dictionaries for different age groups. Similarly, the list of headwords in ECCDs should also be based on an English corpus for Chinese children. Without a corpus as evidence, an editor is likely to exclude some words to which children are mostly exposed, and include some words to which children are seldom exposed.

A children's English corpus like the OCC has not yet been built in China, though there are some small corpora of children's language for specific research purposes. A 7-million-token Chinese corpus, which was based on popular Chinese books for children aged 1 to 12, has been constructed to investigate language features of children's books (Zhi 2016). In order to compare the use of different parts of speech, Gu (2018) has also constructed a very small Chinese children's cartoon corpus (48,401 tokens) and a foreign children's animation corpus (48,900 tokens). Xu (2018) has created a small corpus, consisting of data from audio story books for children aged 3 to 6. However, these corpora will do little to ECCDs. In addition, none of them are freely accessible.

To produce a reliable headword list and select appropriate examples (see Section 5.2.2), lexicographers of ECCDs call for a balanced English corpus for Chinese children. The corpus could consist of English textbooks for children, children's English magazines, children's books in English, children's compositions in English, etc. (cf. Unstead 2009), and be balanced in terms of text types, and of writings for children and writings by children. In addition, websites, children's English cartoons and other digital texts are very essential corpus sources in the digital era.

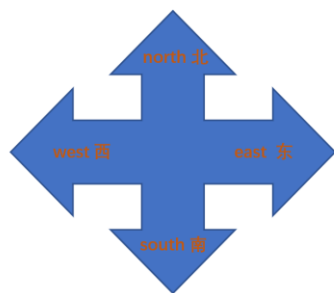
While constructing an English corpus for Chinese children, the OCC can be a reference. However, a close imitation is not advocated. In terms of data sources, the corpus for Chinese children should mainly include English novels popular in China rather than in Britain. Popular English novels in other parts of the English-speaking world should also make an appropriate contribution. In OCC, the writings by children are made up of material from websites where English children have posted reviews, poems and stories, whereas an English corpus for Chinese children should consist of compositions, test papers and

other material written by Chinese children. In OCC, there are three classified key stages: 5–7 years old, 7–11 years old, and 11–14 years old. Because of different proficiency levels, a corpus for Chinese children should have three different key stages: 2–6 years old, 6–12 years old and 12–16 years old.

4. Arranging the headwords list: thematically vs. alphabetically

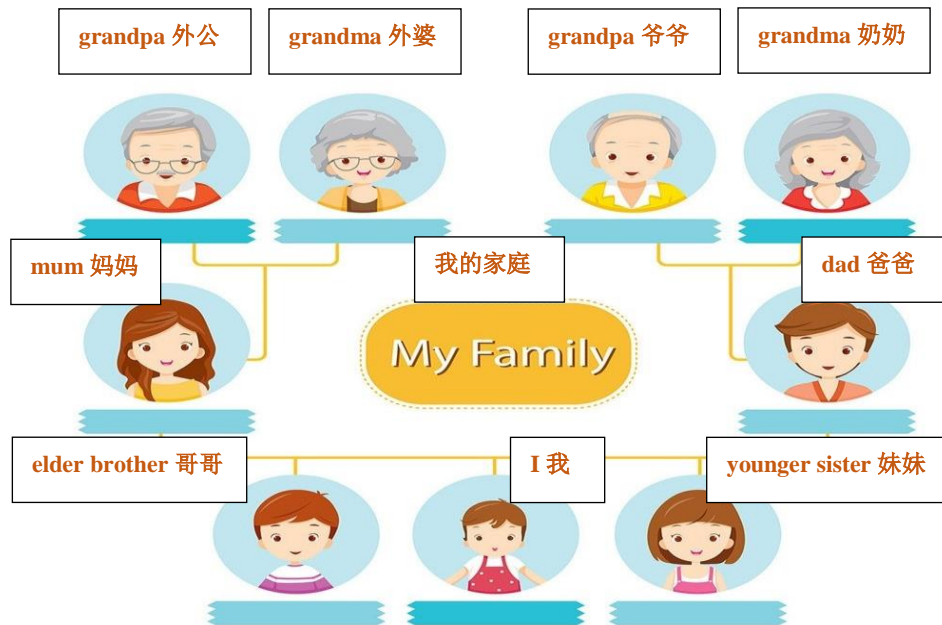
In monolingual Chinese dictionaries for primary school children, the headwords are predominantly arranged in alphabetical order, for pupils have learnt *pinyin* (the standard system of Roman spelling in Chinese) in school. An index of radicals of Chinese characters and sometimes an index of strokes of Chinese characters are also provided to help students to locate a Chinese character whose pronunciation is not known.

In ECCDs for primary school students, compilers can arrange the headword list in alphabetical order, because pupils are very familiar with the sequence of the alphabet. However, in ECCDs for younger children, it is beyond their ability to look up an English word in a dictionary arranged alphabetically. For this reason, a thematic arrangement would be more suitable. Headwords can be arranged in semantic sets, including related concepts, synonyms, etc. To help these younger children to use an ECCD, lexicographers can choose themes that will appeal to children, such as shapes, transport, colours, animals, plants, toys, and games. The words related to a theme can be shown in one picture. In this way, images will not only represent the referents but also reflect the relationship among related words. Accompanying headwords can then be arranged alphabetically, and sometimes explained with simple sentence examples. Consider the two themes — 'directions' and 'family' — in Figures 1 and 2.



east 东 dōng
north 北 běi
south 南 nán
west 西 xī

Figure 1: Thematic arrangement of words related to 'directions' (Source: <https://www.51wendang.com/doc/0d2012567949db45601fc2c8/2>)



dad	爸爸	bà·ba
elder brother	哥哥	gē·ge
elder sister	姐姐	jiě·jie
grandma	奶奶	nǎi·nai
grandma	外婆	wàipó
grandpa	爷爷	yé·ye
grandpa	外公	wàigōng
I	我	wǒ
mum	妈妈	mā·ma
younger brother	弟弟	dì·di
younger sister	妹妹	mèi·mei

Figure 2: Thematic arrangement of words related to 'family' (Source: http://bpic.588ku.com/element_origin_min_pic/18/06/09/2a41633915c03814f45a9c4dde28fb0c.jpg)

The two samples demonstrate that it is more effective for this age group to group words related to theme than to scatter them throughout a dictionary arranged alphabetically. The thematic arrangement shows the relationships among the headwords in a group. Children can clearly get to know how they

are connected to each other. This method is suitable for related words in a semantic field, such as furniture, facilities in the playground, colours, musical instruments, toys, and transport. But for some abstract nouns (e.g., *honour, loyalty, truth*) or abstract verbs (e.g., *discuss, compare, evaluate, analyse*), it will be unsuitable. Unlike an alphabetical arrangement, a thematic order can attract younger children, and promote their understanding of related concepts. Therefore, dictionaries for younger children should combine fun and learning.

However, there are some downsides to the thematic arrangement. It takes up considerable space, and pictorial illustrations are potentially expensive. In addition, words in thematic order are inconvenient to look up, and it is also difficult to fit examples (or definitions) into a highly illustrated thematic structure. Therefore, it would be a good idea to supplement it with an index of English headwords in alphabetical order, and an index of Chinese equivalents in *pinyin*.

5. Multifaceted methods of explanation

5.1 Pictorial illustrations in ECCDs for younger children

One of the vital functions of a children's dictionary is to help users understand the meaning of a headword in question. While verbal definition is a major method in a monolingual dictionary, a bilingual dictionary uses translation equivalents and/or example sentences. In ECCDs for younger children, few provide example sentences. Sole translation equivalents can hardly meet the needs of younger children. Therefore, pictorial illustration plays a significant role in this type of children's dictionaries.

Dictionaries are generally consulted rather than read as running text. However, in ECCDs for younger children, a dictionary is read rather than consulted, because the main task is to help younger children acquire English vocabulary. What they want to know is not 'why it is' but 'what it is'. Dictionaries for them should be interesting and attractive. Pictorial illustration, either photographic, digitally created or hand-drawn, serves this function. 'An illustration is a particular kind of image which is used in conjunction with a text and which decorates, illustrates, or explains the text' (Klosa 2015: 516). Lew et al. (2018: 53-54) further point out that of the three functions Klosa has discussed, explaining the text that is 'the most central function of lexicographic illustration'. Similarly, it is not decorating but explaining the text is the main function of a pictorial illustration in ECCDs. A pictorial illustration helps children to build up connection between a headword and its designatum.

In dictionaries for younger children, pictorial illustration is not a complement but a principal method of explanation. Its main function is to help younger children acquire second language vocabulary and build connection between images and words. Therefore, it is a challenge for illustrators to create suitable pictures to match children's cognitive levels, appeal to their interests,

and support their understanding of headwords. In order to meet the standards, the pictures should be representative, attractive and vivid.

Consider the following pictorial illustration of 'facial expression' in *Ertong Yinghan Baike Tujie Cidian* (*My First Visual Dictionary*).



Figure 3: Pictorial illustration of 'facial expression' in *Ertong Yinghan Baike Tujie Cidian* (*My First Visual Dictionary*)

Images 2 (*crying*), 3 (*afraid*) and 6 (*angry*) in Figure 3 have not clearly represented their corresponding meaning. Image 2 shows that the girl is sad but not 'crying': there are no tears on her face. From Images 3 and 6, it is difficult to discern whether the girls are really 'afraid' or 'angry' from their facial expressions. Also, arguably, *laughing*, *crying*, *smiling*, *shouting* are not emotions (which are by definition nouns), although *happiness*, *sadness*, *amusement* and *fear/anger* are.

To summarize, in ECCDs for younger children, pictorial illustration complements translation equivalent. Editors need to acquire some understanding of the principles of good illustration and the stages of children's development before they can select appropriate pictures for illustration. In order to avoid inappropriate pictures, illustrators can join in the work of compiling, and psychologists could be consulted if necessary. Feedback from children on draft or sample illustrations would also be very helpful in selecting the most appropriate illustrative style or approach.

5.2 Multifaceted methods of explanation in ECCDs for primary school students

5.2.1 Pictorial illustration

In ECCDs for primary school students, pictorial illustration does not play a dominant role as in ECCDs for younger children, because users at this later stage have some basic language knowledge and can recognize some basic English words. Nevertheless, when being asked to define an ideal dictionary, this

group of children still listed attractive pictorial illustrations as one of the most important features (Turrini et al. 2000). Visual content makes a dictionary more attractive (Klosa 2015: 516, Biesaga 2017: 133). This is particularly true of school dictionaries.

Due to technical, financial and other factors, pictorial illustration is not often used in a general-purpose dictionary (Biesaga 2017: 134). However, some words are more appropriately illustrated with a picture than defined verbally, such as musical instruments, household goods, computer equipment, styles of architecture, animals, plants and furniture. In terms of categories like fruit, games, vegetables, plants and clothes, it is not difficult to illustrate them with appropriate pictures, as Figure 4 demonstrates. But some pictorial illustrations in an ECCD may cause confusion. As Figure 5 shows, one of the images in the pictorial illustration at **square** is not a square but a rectangle. Similarly, learners have no idea whether the pictorial illustration at **summer** is meant to suggest 'sun', 'landscape' or something else. A defined word should be linked to a clear pictorial illustration. There are some other issues with the entries in these figures, such as an article and a plural missing in the two sentence examples: "My favorite fruit is [a] banana." at **banana**, and "Cut the cake in square [squares]." at **square**.

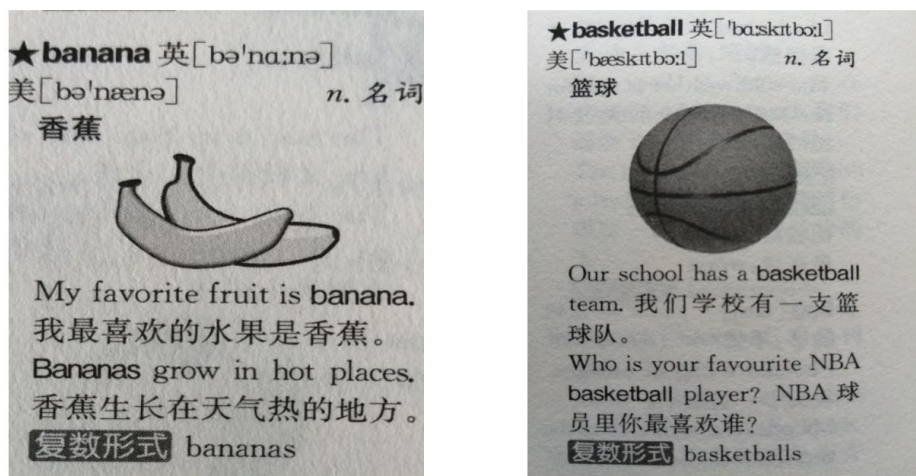


Figure 4: Pictorial illustrations of 'banana' and 'basketball' in *Xiaoxuesheng Yinghan Cidian (An English-Chinese Dictionary for Primary School Students)*

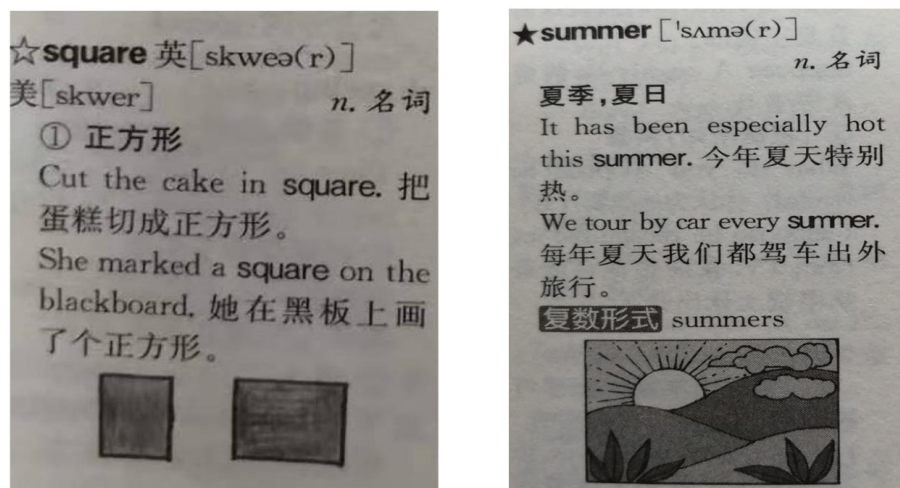


Figure 5: Pictorial illustrations of 'square' and 'summer' in *Xiaoxuesheng Yinghan Cidian* (*An English–Chinese Dictionary for Primary School Students*)

5.2.2 Examples

Unlike dictionaries for younger children, lexicographers can offer some simple examples in ECCDs for primary school students. It is commonly accepted that examples help users understand headwords, and complement grammatical information. Some principles of exemplification in learner's dictionaries have been proposed. Examples should provide the typical usage of headwords (Fox 1987), should have practical applicability (Moulin 1983), and should be easy to understand (Lemmens and Wekker 1991, Atkins 1995, Potgieter 2012), natural and concise (Cowie 1999).

As for exemplification in ECCDs for primary school students, compilers should focus on the following three aspects. Firstly, they should select, from a corpus, typical as well as simple example sentences. In terms of contexts of usage, typical examples are more informative than untypical ones, and primary school students are more likely to encounter similar sentences. In addition, examples should be easy to comprehend, because pupils do not have adequate knowledge to understand some difficult words and complicated sentence structures (Thorndike 1991: 15). Admittedly, corpus-based examples may include words and structures beyond children's proficiency. In this case, lexicographers can eliminate or revise them.

Secondly, the chosen examples should assist learners in understanding the meaning and proper use of headwords by providing typical contexts of usage. Consider how *Zhongguo Xiaoxue Yingyu Xuexi Cidian* (*An English–Chinese Dictionary for Primary School Learners in China*) and *Waiyanshe Kelinsi Yinghan Hanying*

Cidian (FLTRP COLLINS Elementary English–Chinese Chinese–English Dictionary) exemplify **miss** in (1) and (2), respectively.

- (1) *Henry has **missed** the bus.*
- (2) – *Did you get on the train? – No, I **missed** it.*
– *Goodbye! I'll **miss** you. – I'll **miss** you, too.*

For this headword, *Waiyanshe Kelinsi Yinghan Hanying Cidian (FLTRP COLLINS Elementary English–Chinese Chinese–English Dictionary)* appears to be more useful than *Zhongguo Xiaoxue Yingyu Xuexi Cidian (An English–Chinese Dictionary for Primary School Learners in China)*, since the former distinguishes the two different senses of **miss** with separate example sentences (showing the verb can be used for an inanimate object — 'the train' — as well as for a person). The brief dialogues in the former dictionary show the typical use of the headword in a way that is more comprehensible to children. In contrast, example (1) uses 'missed' in a way that gives the user little idea of the meaning of the word: one could replace 'missed' with 'painted', for example, or several other verbs, and the sentence would still make sense.

Thirdly, ECCDs should give prominence to grammatical information that children just beginning to learn English often confuse, because of the influence of Chinese grammar — the negative L1 transfer. According to Jiang (2000: 47), second language learners at the early stage are likely to copy the lemma information of the L1 counterpart into the L2 lexical entry, thus mediating L2 word use. If the grammatical usage of an English headword is different from that of its Chinese equivalent, the negative L1 transfer will come into play. A typical error committed by children is the inflection of an English predicate verb. While an English verb must agree with the number and person of the subject in a sentence, Chinese is an inflection-free language. To inhibit the transfer of L1 Chinese, lexicographers can illustrate the grammatical usage with some sentence examples. For instance,

English	Chinese
I am a student.	我 是 一个学生。(wǒ shì yīgè xuéshēng)
You are a student.	你 是 一个学生。(nǐ shì yīgè xuéshēng)
She/He is a student.	她/他 是 一个学生。(tā shì yīgè xuéshēng)
We/They are students.	我们/他们 是 学生。(wǒmen/tāmen shì yīgè xuéshēng)

Considering space for print and time for lexicographers, examples of *be* can be optimized as follows:

- I **am** / You **are** / She **is** / He **is** a student.
我**是**/你**是**/她**是**/他**是**一个学生。(wǒ **shì**/nǐ **shì**/ tā **shì**/ tā **shì** yīgè xuéshēng)
We/They **are** students. 我们/他们**是**学生。(wǒmen/tāmen **shì** xuéshēng)

We may take for granted the inflection of 'be'. However, many Chinese pupils often commit such errors, for the Chinese equivalent *shì* does not change its form whatever the subject is.

As an essential didactic tool for language acquisition, a school dictionary should provide children with some example sentences (and sometimes usage notes), which warn them against potential grammatical errors caused by L1 negative transfer. To identify pupils' grammatical errors, lexicographers can exploit resources such as an English composition corpus by Chinese children and a grammar book.

To recapitulate, examples in ECCDs should use simple words and grammatical structures to demonstrate the typical context of usage of a headword, its grammatical properties and semantic features.

6. Conclusion

Children's dictionaries should not be an abridgement or simplification of dictionaries for adult learners. Factors such as children's proficiency level and learning and cognitive abilities have an impact on the making of ECCDs. In this paper, we propose distinguishing dictionaries for younger children from those for primary school students, because children in the two periods are quite different in terms of language development and cognitive ability. The construction of an English corpus for Chinese children is imperative because corpus data could facilitate the production of a reliable and useful headword list, and the selection of typical sentence examples which demonstrate the semantic, syntactic, and collocational features of a headword. In addition, evidence provided by corpus data can be used to warn students against potential grammatical errors. Pictorial illustration complements translation equivalent in ECCDs for younger children and for primary school students. In addition, words used in example sentences should be congruent with children's limited cognitive and learning abilities.

Acknowledgements

This study was supported by the Philosophy and Social Science Fund of the 13th Five-year Plan of Guangdong Province of China (Project No. GD18XWW02) and China Scholarship Council (No. 201808440098). We are grateful to Professor Elsabé Taljard and two anonymous adjudicators for their valuable comments and suggestions. Correspondence should be addressed to Dr. Hai Xu.

References

A. Dictionaries

- Huo, Qingwen (Ed.). 2001. *Zhongguo Xiaoxue Yingyu Xuexi Cidian (An English–Chinese Dictionary for Primary School Learners in China)*. Beijing: Foreign Language Teaching and Research Press.

- Institute of Linguistics, Chinese Academy of Social Sciences (Ed.).** 1953. *Xinhua Zidian (Xinhua Dictionary of Chinese Characters)*. Beijing: The Commercial Press.
- Li, Yuzhi (Ed.).** 2003. *Xinbian Xiaoxuesheng Yinghan Cidian (The Brand-New English–Chinese Dictionary for Primary School Students)*. Beijing: The Commercial Press.
- Luo, Lie and Qinghua Xiao (Eds.).** 2017. *Xiaoxuesheng Yinghan Cidian (An English–Chinese Dictionary for Primary School Students)*. Chengdu: Sichuan Lexicographic Publishing House.
- QA International (Ed.).** 2004. *Ertong Yinghan Baiké Tujie Cidian (My First Visual Dictionary)*. Beijing: Foreign Language Teaching and Research Press.
- Yu, Shutong (Ed.).** 1998. *Xinhanying Faxue Cidian (A New Chinese–English Law Dictionary)*. Beijing: Law Press.
- Zhang, Daozhen.** 2000. *Tujie Ertong Yinghan Cidian (Illustrated Children’s English–Chinese Dictionary)*. Beijing: Foreign Language Teaching and Research Press.
- Zhang, Siying (Ed.).** 2006. *Waiyanshe Kelinsi Yinghan Hanying Cidian (FLTRP COLLINS Elementary English–Chinese Chinese–English Dictionary)*. Beijing: Foreign Language Teaching and Research Press.
- Zhong, Jiayan (Ed.).** 2011. *Zhongguo Xiaoxuesheng Yingyu Xuexi Cidian (An English–Chinese Dictionary for Primary School Learners in China)*. Changchun: Jilin Education Press.

B. Other literature

- Atkins, B. T. S.** 1995. The Role of Examples in a Frame Semantics Dictionary. Shibatani, M. and S.A. Thompson (Eds.). 1995. *Essays in Semantics and Pragmatics: In Honour of Charles J. Fillmore*: 25-42. Amsterdam: John Benjamins.
- Béjoint, H.** 2000. *Modern Lexicography: An Introduction*. Oxford/New York: Oxford University Press.
- Biesaga, M.** 2017. Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated? *Lexikos* 27: 132-151.
- Cheng, Rong.** 2001. Hanyu Xuexi Cidian Bianzuan Tedian de Tantaoyao (Features of the Compiling of Chinese Learners' Dictionary). *Cishu Yanjiu (Lexicographical Studies)* 2: 64-71.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press/Clarendon Press.
- De Schryver, G.-M. and D.J. Prinsloo.** 2011. Do Dictionaries Define on the Level of their Target Users? A Case Study for Three Dutch Dictionaries. *International Journal of Lexicography* 24(1): 5-28.
- Fox, G.** 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London/Glasgow: Collins ELT.
- Gavrilidou, M., V. Giouli and P. Labropoulou.** 2008. The Greek High School Dictionary: Description and Issues. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the 13th EURALEX International Congress, EURALEX 2008, Barcelona, Spain, 13–19 July 2008*: 515-524. Barcelona: Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra.
- Gu, Mengmeng.** 2018. *Jiyu Yuliaoku de Ertong Donghuapian Yanjiu (Corpus-based Research on Vocabularies in Children’s Animations)*. M.A. Thesis. Shanghai: Shanghai Normal University.
- Jiang, N.** 2000. Lexical Representation and Development in a Second Language. *Applied Linguistics* 21(1): 47-77.
- Klosa, A.** 2015. Illustrations in Dictionaries. Encyclopaedic and Cultural Information in Dictionaries. Durkin, P. (Ed.). 2015. *The Oxford Handbook of Lexicography*: 515-531. Oxford: Oxford Univer-

sity Press.

- Kosch, I.** 2013. An Analysis of the *Oxford Bilingual School Dictionary: Northern Sotho and English*. *Lexikos* 23: 611-627.
- Lemmens, M. and H. Wekker.** 1991. On the Relationship between Lexis and Grammar in English Learners' Dictionaries. *International Journal of Lexicography* 4(1): 1-14.
- Lew, R., R. Kazmierczak, E. Tomczak and M. Leszkowicz.** 2018. Competition of Definition and Pictorial Illustration for Dictionary Users' Attention: An Eye-tracking Study. *International Journal of Lexicography* 31(1): 53-77.
- Moulin, A.** 1983. The Pedagogical/Learner's Dictionary II: LSP Dictionaries for EFL Learners. Hartmann, R.R.K. 1983. *Lexicography: Principles and Practice*: 144-152. London: Academic Press.
- Piaget, J.** 1969. *Psychologie et Pédagogie*. Paris: Denoël.
- Potgieter, L.** 2012. Example Sentences in Bilingual School Dictionaries. *Lexikos* 22: 261-271.
- Rozman, T.** 2008. Mother-tongue's Little Helper (The Use of the Monolingual Dictionary of Slovenian in School). Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the 13th EURALEX International Congress, EURALEX 2008, Barcelona, Spain, 13-19 July 2008*: 1317-1324. Barcelona: Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra.
- Rundell, M.** 2010. Taking Corpus Lexicography to the Next Level: Explicit Use of Corpus Data in Dictionaries for Language Learners. Zhang, Y. (Ed.). 2010. *Learner's Lexicography and Second Language Teaching: Proceedings of First International Symposium on Lexicography and L2 Teaching and Learning*: 367-386. Shanghai: Shanghai Foreign Language Education Press.
- Sene Mongaba, B.** 2016. A Terminological Approach to Making a Bilingual French–Lingala Dictionary for Congolese Primary Schools. *International Journal of Lexicography* 29(3): 311-322.
- Tarp, S. and L. Ruiz Miyares.** 2013. Cuban School Dictionaries for First-Language Learners: A Shared Experience. *Lexikos* 23: 414-425.
- Thorndike, E.** 1991. The Psychology of the School Dictionary. *International Journal of Lexicography* 4(1): 15-22.
- Turrini, G., A. Paccosi and L. Cignoni.** 2000. Software Demonstration: Combining the Children's Dictionary Addizionario with a Multimedia Activity Book. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000. Vol I*: 107-110. Stuttgart: Institut für maschinelle Sprachverarbeitung, University of Stuttgart.
- Unstead, S.** 2009. A Question of Words. *Books for Keeps: The Children's Book Magazine* 175. <https://booksforkeeps.co.uk/article/a-question-of-words/>.
- Verburg, M.** 2006. The Making of *My First Van Dale*: A Pre-School Dictionary. Corino, E., C. Marelo and C. Onesti (Eds.). 2006. *Proceedings of the 12th EURALEX International Congress, EURALEX 2006, Torino, Italia, September 6-9, 2006*: 357-362. Alessandria: Edizione Dell'Orso.
- Wild, K., A. Kilgarriff and D. Tugwell.** 2013. The Oxford Children's Corpus: Using a Children's Corpus in Lexicography. *International Journal of Lexicography* 26(2): 190-218.
- Xu, Kaili.** 2018. *Jiyu Yuliaoku de 3-6 Sui Ertong Yousheng Gushi Yuyan Yanjiu (Corpus-based Research on Audio Stories Language of Children Aged 3-6)*. M.A. Thesis. Shanghai: Shanghai Normal University.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia.
- Zhi, Luyao.** 2016. *Jiyu Hanyu Ertong Duwu Bianxie de Hanyu Yuliaoku Jianshe (The Construction of Chinese Corpus Based on Chinese Children's Books)*. M.A. Thesis. Guangzhou, Guangdong University of Foreign Studies.

Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology*

Sven Tarp, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; Department of Afrikaans and Dutch, Stellenbosch University, South Africa; International Centre for Lexicography, Valladolid University, Spain; Centre of Excellence in Language Technology, Ordbogen A/S, Denmark; and Centre for Lexicography, Aarhus University, Denmark (st@cc.au.dk)*

Abstract: This is a report from the real world. It informs about the outcome of a project, which the author conducted during a months-long research stay at the Danish company Ordbogen where he integrated its research and development (R&D) team. The first part of the project was to test machine translation and find out to what extent it is usable in the compilation of bilingual lexicographical databases. The hypothesis was that the technology was not yet mature. But surprisingly, it turned out that the accuracy rate is already so high that it is worth considering how to implement it. The second part of the project aimed at further developing an idea formulated by Fuentes-Olivera et al. (2018) on how to invert a dictionary without losing semantic content. The new vision is to compile a monolingual L2 database, bilingualize it to an L2–L1 database using machine translation, and then invert the relationship between L2 lemmata and L1 equivalents using the L1 definitions of the L2 lemmata as the axis. The third part of the project was to test this idea using a specially designed ad hoc program. The program automatically uploads relevant data from existing lexicographical databases, translates L2 definitions and example sentences into L1, suggests adequate L1 equivalents, and eventually inverts the relationship between the two languages. It worked, but the methodology still needs further refinement to be implementable on a large scale. The report concludes by listing some of the remaining challenges and defining the new role of the lexicographer in this type of project.

Keywords: LEXICOGRAPHICAL R&D, INTERDISCIPLINARY COLLABORATION, DIGITAL TECHNOLOGY, BILINGUAL LEXICOGRAPHY, LEXICOGRAPHICAL DATABASES, MACHINE TRANSLATION, AUTOMATIC INVERSION, OBJECT LANGUAGE, AUXILIARY LANGUAGE, HUMAN VERSUS ARTIFICIAL LEXICOGRAPHER

Opsomming: 'n Omwenteling in tweetalige leksikografie: Die verbetering van kwaliteit en produktiwiteit met nuwe metodes en tegnologie. Hierdie is 'n

* A version of this project report was presented as a paper at the 26th Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by Stellenbosch University, Stellenbosch, South Africa, 27–29 July 2022.

verslag uit die praktyk. Daar word verslag gelewer oor die uitkoms van 'n projek, wat die outeur gedurende 'n maandelange navorsingstydperk by die Deense maatskappy Ordbogen aangevoer het waartydens hy die navorsings- en ontwikkeling- (N&O)-span geïntegreer het. Die eerste deel van die projek het die toets van masjienvertaling behels en om die bruikbaarheid daarvan in die saamstel van tweetalige leksikografiese databasisse te bepaal. Die hipotese was dat die tegnologie nog nie gevorderd genoeg was nie. Dit het egter, verrassend genoeg, geblyk dat die akkuraatheidsyfer reeds so hoog was dat dit die moeite werd was om die implementering daarvan te oorweeg. Die tweede deel van die projek het die verdere ontwikkeling van 'n idee, geformuleer deur Fuertes-Olivera et al. (2018) oor die omskakeling van 'n woordeboek sonder verlies van semantiese inhoud, ten doel gehad. Die nuwe visie is om 'n eentalige L2-databasis saam te stel, dan met behulp van masjienvertaling te omskep in 'n L2-L1-databasis, en daarna die verhouding tussen L2-lemmata en L1-ekwivalente om te skakel deur die L1-definisies van die L2-lemmata as die spil te gebruik. Die derde deel van die projek was die toets van hierdie idee met 'n spesiaal ontwerpte ad hoc-program. Hierdie program laai outomaties relevante data vanuit leksikografiese databasisse, vertaal L2-definisies en -voorbeeldsinne in L1, stel gepaste L1-ekwivalente voor, en skakel uiteindelik die verhouding tussen die twee tale om. Dit was geslaagd, maar die metodologie moet nog verder verfyn word voordat dit op groot skaal geïmplementeer kan word. Die verslag word afgesluit met die lys van sommige van die oorblywende uitdagings en met die definiëring van die nuwe rol van die leksikograaf in hierdie tipe projek.

Slutelwoorde: LEKSIKOGRAFIESE N&O, INTERDISCIPLINÊRE SAMEWERKING, DIGITALE TEKNOLOGIE, TWEETALIGE LEKSIKOGRAFIE, LEKSIKOGRAFIESE DATABASISSE, MASJIENVERTALING, OUTOMATIESE OMSKAKELING, DOELTAAL, SEKONDÊRE TAAL, MENSLIKE VERSUS KUNSMATIGE LEKSIKOGRAAF

1. Introduction

Today, the compilation and presentation of dictionaries and other lexicographical products are inconceivable without assistance from digital technologies that are constantly improving and breaking new ground. Good lexicographical craft presupposes, in one way or another, interdisciplinary collaboration with specialists from other fields. The collaboration has various dimensions. One of them is research and development (R&D), i.e., experimenting with new compilation methods and ways of presenting the final product to the target users. To be successful, the pursuit of innovation requires almost daily contact and daily exchange of views between the lexicographer, on the one hand, and information scientists, programmers, or designers, on the other hand. Few lexicographers, whether university professors or employees in publishing houses, have the opportunity to engage in this kind of interdisciplinary research, without which the discipline will have to struggle even more to find its place in the Fourth Industrial Revolution. Far too often, the two parts work in different directions with only occasional contact.

From this perspective, I consider myself extremely privileged. In 2021, my University facilitated a research stay at Ordbogen A/S, a successful Danish

company specializing in language services, digital teaching material, online dictionaries, and writing assistants. It was an extraordinary experience. It allowed me to conduct experiments and test new and old ideas during various months. From the very first day, I was co-opted by the company's ODIN Team, where most of its research and development takes place. Here, I was the only one with a background in lexicography, language didactics, and translation. The other members were information scientists, programmers, and designers who, from the perspective of their expertise, had a very different approach to my discipline. It was both challenging and stimulating.

Since 2017, I have, to some extent, collaborated in the development of Ordbogen's digital Write Assistant and published several research articles on this topic, one of them together with information scientists from the company (Tarp et al. 2017). I find the underlying philosophy and technology timely and future-oriented. I am especially fascinated by the new ways and forms in which lexicographical data can be presented to users engaged in a particular activity like L2 writing, thus supporting the basic tenets of the Function Theory (see Fuertes-Olivera and Tarp 2014). But I am also very critical of how it has been done so far. From my narrow disciplinary perspective, I have even described the lexicographical component as "Write Assistant's Achilles heel", among other things because "existing databases are highly deficient and problematic" when it comes to serving this kind of software and turning it into a high-quality product (Tarp 2019: 237-238).

There is an urgent need to prepare more appropriate databases, and to this end, develop new compilation methods and techniques that can guarantee both higher quality and higher productivity. The last point is particularly important considering that many publishers of dictionaries struggle to make ends meet. Higher productivity without compromising quality could be part of the solution, but it calls for basic research through intense interdisciplinary collaboration between relevant experts. I was thus excited when Aarhus University allowed me to have a research stay at Ordbogen. To that end, I formulated two projects, or experiments, to be conducted:

1. Using artificial intelligence to select adequate example sentences and automatically assign them to the relevant senses in a lexicographical database.
2. Using machine translation to translate L2 definitions into L1, where the translated definitions can both explain the meaning of L2 lemmata and function as semantic differentiators when bridging from L1 to L2.

The immediate objective of the two experiments was to see how far technology has come, to what extent it is already implementable, and what consequences it may have for the future relationship between man and machine, between the human and the artificial lexicographer. The following report and reflections will focus on the second of the two experiments, which resulted in a major breakthrough.

2. Testing machine translation

After decades of struggle, machine translation has improved considerably during the past few years. I am not aware of any lexicographers experimenting with this technology, but I am convinced that it is only a question of time before the discipline will adopt it at a broad scale. It is thus necessary to be at the forefront of this development, although I did not have high expectations for its immediate relevance to lexicography. At our initial meeting, Michael Walther, head of the ODIN Team, said that this technology only becomes really attractive and relevant when over 70 percent of the translations are correct and can be inserted directly into the database. Both he and I judged that it still takes some time before we can achieve this success rate.

During the whole research stay, I worked closely with Henrik Hoffmann, a talented programmer and web developer. We decided to start with Google Translate, which many scholars consider the best, or at least one of the best, translation tools available today. As an empirical basis, we chose a monolingual Spanish database which Pedro Fuertes-Olivera is compiling in Valladolid under the auspices of Ordbogen (at that moment, he had completed approximately 80,000 senses). Henrik Hoffmann then extracted 200 random definitions, which were immediately translated into English using Google Translate.

It was then my task to systematically compare source and target definitions. The result was disappointing, but as expected. About 30 percent of the translations were acceptable, though not perfect. Another 40 percent had major or minor errors that were disruptive to understanding. And the last 30 percent were straight-out incomprehensible.

At this point, it is pertinent to make some observations:

1. Translating 200 definitions out of 80,000 does not give statistically precise results, but the tendency is convincing, according to discussions with members of the ODIN Team;
2. The definitions are written in the non-natural language that characterizes most dictionary definitions. This phenomenon may present additional challenges for machine translation;
3. Spanish is generally a synthetic language, a characteristic that may give rise to particular types of problems when translating into a predominantly analytical language such as English;
4. The result cannot be generalized and directly applied to translation in the reverse language direction, i.e., from English to Spanish, or to translation between other language pairs, each of which has its characteristic features.

Either way, Google Translate is not the solution to our problem. Henrik Hoffmann therefore suggested that we test the DeepL Translator, which I had not used before. The results of this new test came as a complete surprise. Of 200 translated definitions (the same as previously), 156 (78%) were now completely correct, while 44 (22%) had major or minor errors. Most surprisingly, DeepL

correctly translated most cases where Google Translate had to give up on the particular syntactic structure of the Spanish definitions.

Valladolid-UVa	Google Translate	DeepL Translator
abacorar VERB producir una sensación de angustia o recelo	produce a feeling of anxiety or apprehension	to produce a feeling of distress or apprehension
abajar VERB mover algo o moverse uno mismo de un lugar o posición superior a un lugar o posición inferior	move something or move one of a place or a place higher position or lower position itself	to move something or to move oneself from a higher place or position to a lower place or position
índice NOUN en imprenta, símbolo tipográfico en forma de mano que apunta un lugar; se usa para señalar algo importante o para llamar la atención en un texto	in press, typographic symbol shaped hand pointing a location; It is used to indicate important something or to get attention in a text	in print, typographic symbol in the form of a hand pointing to a place; used to point to something important or to draw attention to a text
ígneo ADJECTIVE que hace referencia a las rocas, minerales o formaciones geológicas que se originaron a partir de los materiales fundidos del interior de la tierra	It referred to rocks, minerals or geologic formations that originated from molten material of the inner earth	which refers to rocks, minerals or geological formations that originated from the molten materials of the earth's interior

Table 1: Comparison of Google's and DeepL's translation performance

Table 1 contains four illustrative examples of how Google Translate and DeepL perform when translating definitions from the Spanish database (Valladolid-UVa). The definition assigned to the verb *abacorar* has a relatively simple structure, and both translation tools provide English translations that are correct and understandable, although slightly different. By contrast, the definitions of the verb *abajar* and the noun *índice* have a more complex syntactic structure. In both cases, Google Translate offers incomprehensible translations, whereas DeepL manages to decipher the Spanish syntax and provide acceptable translations into English. Finally, the adjective *ígneo* shows a minor but frequent problem where Google Translate, contrary to DeepL, cannot grasp the initial structure of the Spanish definition and provides an English text that is understandable but requires language revision to serve as a lexicographical definition.

Just in case, we had Pedro Fuertes-Olivera review the translations, and he reached the same result. In other words, the performance of DeepL looked more than promising for the use of machine translation in lexicography, but further testing was required to give a final judgment.

The target users of the definitions contained in the monolingual Spanish database are native Spanish speakers. However, since the translation of these items only makes sense if the target users are non-native speakers, we now decided to extract our empirical data from a monolingual dictionary designed for the latter. For this purpose, we chose the *Oxford Dictionary* on the Lexico platform, from which we hand-picked 200 random definitions belonging to different letters and parts of speech. These definitions are characterized by a more straightforward language with less complex syntax than those from the Valladolid-UVa database. We now used DeepL to translate them into Spanish and Danish, respectively. The result was even more convincing than previously.

Of 200 definitions translated into Spanish, 187 (93,5%) were now completely correct, while 13 (6,5%) had minor or major problems. Of these, only 4 (2%) were so serious as to require a total rewrite, whereas the rest only needed a slight reworking. The revision could, in most cases, be done using another of DeepL's functionalities (see below).

Spanish and English are languages with many native speakers, while Danish has a lot fewer. As DeepL is trained on existing texts and translations, we expected its performance to be somewhat inferior in Danish. This prediction turned out to be true, but the result was far from catastrophic. Of 200 definitions translated into Danish, 142 (71%) were correct, 48 (24%) had minor problems, while the remaining 10 (5%) were unusable. In half of the 48 translations with minor problems, the first word in the definition should either be deleted or changed from one word class or inflectional form to another. Another frequent problem was the appearance of two identical words in the translation when the source definition included two more or less synonymous words. In both cases, these are minor inconveniences, which the lexicographers revising the text can quickly and easily correct by clicking on the alternative suggestions that DeepL offers its users.

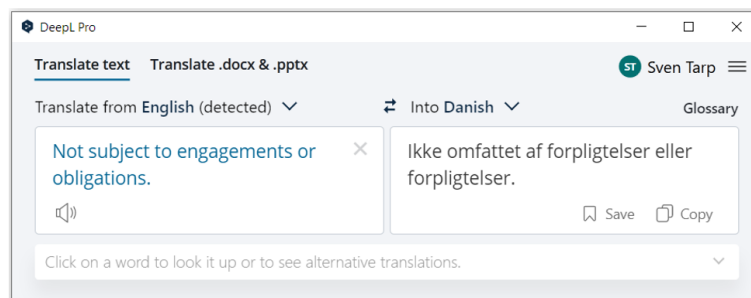


Figure 1: Using DeepL to translate from English into Danish

In Figures 1 and 2 we can see how it works. Figure 1 shows the translation of the English text segment *Not subject to engagements or obligations* that defines one of the senses of the adjective *free*. The words *engagements* and *obligations* are synonymous to some extent. However, DeepL does not grasp the subtle semantic difference and translates both words into *forpligtelser*. The Danish definition is understandable. But it is not as semantically rich as the English one, and neither is it stylistically convenient in a lexicographical context. To solve this problem, the user of DeepL can simply click on one of the translated words and get alternative solutions (see Figure 2). In the concrete case, at least two suggested alternatives (*engagementer* and *aftaler*) could replace *forpligtelser* the first time it occurs. Another click on the preferred suggestion modifies the target definition accordingly. As can be seen, DeepL's user-friendly design and functionalities make it easy to revise and, if needed, correct small pieces of text. A trained lexicographer can probably do it in a few seconds, thus saving a considerable amount of time.



Figure 2: Alternative solutions suggested by DeepL

As mentioned above, our findings are not statistically precise, as the exact percentages cannot be repeated. But the tendency is indisputable. Out of curiosity, we also had colleagues revise a few samples of the same English definitions translated into Chinese, French, and German, respectively. In all three cases, DeepL showed a very high accuracy rate. The conclusion was that we were, indeed, on the verge of a breakthrough that may have important implications for the compilation of future bilingual lexicographical databases. But what are the implications? It was now time to take advantage of this unique opportunity

to experiment with new methods for generating lexicographical data. It implied, above all, to elaborate on an old idea.

3. Reflections on bilingual lexicography

For several years, I have been critical of the way bilingual dictionaries and lexicographic databases are conceived. The very concept of a bilingual dictionary is clearly ambiguous (see Tarp 2005). As used in both academic and non-academic literature, it covers a broad range of very different dictionary types, where the only common feature seems to be that two languages are involved in one way the other. The definitions provided by well-known dictionaries of lexicography like Martínez de Sousa (1995), Bergenholtz et al. (1997), Burkhanov (1998), and Hartmann and James (1998) also vary considerably. Some scholars even regard the terms *bilingual dictionary* and *translation dictionary* as synonyms. Marengo (2003: 325), for instance, defines "bilingual dictionaries only as those dictionaries which place the two languages in contact for purposes of translation". Several terms imported from translation science have strongly influenced bilingual lexicographical terminology. In his classical book on new French dictionaries, Hausmann (1977: 58) used the terms *herübersetzende* and *hinübersetzende* (*translating into* and *from L1*, respectively) to classify L2–L1 and L1–L2 dictionaries.

I have become increasingly convinced that existing lexicographical terminology mentally blocks the necessary creativity and innovation. This is particularly true of the terms *source language* and *target language*, which also have been adopted from translation studies. They are applied almost uncritically to denote L1 and L2 in an L1–L2 dictionary and L2 and L1 in an L2–L1 dictionary, respectively, i.e., sometimes referring to one language and sometimes to the other.

Fuertes-Olivera et al. (2018) discussed the negative consequences of this terminology in the context of bilingual dictionary compilation and, instead, introduced a new vision based on the alternative terms *object language* and *auxiliary language*, where the latter is always the user's native language and the former a foreign, or non-native, language, regardless of the language direction of a dictionary or database.

The rationale is the simple one that an L1 user, as a rule, looks up in an L2–L1 dictionary to understand or know something about L2, while the same user looks up in an L1–L2 dictionary to find an L2 word or L2 expression. In both cases, the lookup is about L2, which is thus the object of the lookup. By contrast, the function of L1 is to help the user either understand L2, use L2, or bridge to L2. The practical implications are considerable. Bilingual L1–L2 dictionaries have so far been built on a selection of L1 lemmata. Instead, the new vision is to base them on a stock of L2 lemmata, which are then processed into L2–L1 and subsequently inverted into L1–L2 dictionaries. In the following, I will briefly explain how this idea has developed and what the implications are.

3.1 Scerba's translating dictionaries

Few people accomplish two or more languages to the same perfection. Most people have a first language (mother tongue or native language), which they master far better than other languages and may, therefore, be considered learners of these languages, whatever their proficiency level. As such, they may need dictionaries when they study and communicate in them. From the perspective of lexicography, the question has been which type of dictionary will best serve this purpose: a monolingual or a bilingual one? This discussion has been going on for decades.

More than eighty years ago, Scerba (1940: 341) urged L2 learners "to discard translating dictionaries as soon as possible and switch to the defining dictionary of the foreign language." By *translating dictionaries*, the Russian scholar understood dictionaries with their lemmata translated into equivalents in another language. Scerba had a distinct language-didactic approach and opposed the contrastive method used in language teaching because it could lead to "a mixed bilingualism due to numerous transfers from L1", and consequently "only goes for L2 learning at the beginner's level" (Mikkelsen 1992: 34). In Scerba's opinion, the L1–L2 "translating dictionary" was only beneficial for L2 learners at a beginner's level. To serve this user segment, he defined a set of general principles for "a special type of translation dictionary" from L1 to L2 (see Mikkelsen 1992: 27). On the other hand, he proposed to write L1 definitions in the L2 "defining dictionaries":

One could create foreign defining dictionaries in the students' native language. Of course, translations of words could also be included when this would simplify definition and would not be detrimental to a full understanding of the foreign word's true nature. (Scerba 1940: 341)

The suggestions show that Scerba was aware of the L2 learners' need to be assisted in their mother tongue, at least for a period. But he regarded it as a *malum necessarium* and advised the users to switch to monolingual L2 dictionaries "as soon as possible". For many years, I was captivated by Scerba's ideas. Today, I am more reserved and hold his disciplinary approach to be somehow misguided inasmuch as a learner's dictionary is not a learning tool in itself but a consultation tool designed to provide the best possible assistance to the L2-learning process. Scerba's ideas were generally unknown in the West until the late 20th Century, but they were applied with minor adjustments in the former Soviet Union and part of Eastern Europe. The results were apparently positive, though not unequivocally. Duda (1986), for instance, reported various critical points appearing in reviews of a Russian–German defining dictionary that had applied Scerba's principles. Among the criticisms were the user's difficulties to perform lexicalization in their native language:

The user is apparently able to understand the analysis of a word's meaning as it is given in the definition in a monolingual dictionary. It appears to be much

more difficult for the user to state the meaning based on a given definition, i.e. to perform lexicalization. (Duda 1986: 13)

Duda did not find that the suggestion substantially challenged Scerba's concept for a defining dictionary. But it required that his principle of providing definitions rather than equivalents was changed so that the latter had priority whenever possible and adequate. This adjustment was definitely a case for bilingual learners' dictionaries.

3.2 The Big Five

In the West, a different tradition developed. During various decades, monolingual dictionaries were considered the *crème de la crème* of learners' dictionaries, notably the English Big Five (Oxford, Longman, Collins, Cambridge, and MacMillan), which largely influenced Western lexicographical thinking after English became an international lingua franca. These dictionaries were undoubtedly high-standard with many innovative and sophisticated features. But just like Hollywood stars, they were iconized and turned into a sort of one-size-fits-all product for users with very different linguistic and cultural backgrounds. Their advocates followed the philosophy that dominated language didactics for many years. Learners should be forced into thinking in the foreign language and, therefore, not unnecessarily exposed to their mother tongue during the learning process. From this perspective, dictionaries were but one of many learning tools and should, by definition, be monolingual. But many language teachers also observed that only a minority of students followed their recommendation, at least outside the classroom where they consulted all sorts of bilingual dictionaries to get the required assistance. The central role of the native language for most L2 learners was clearly underestimated. In this respect, Adamska-Sałaciak and Kernerman (2016: 273) rightly state that the monolingual learners' dictionary "focuses practically on an elite minority of top-level users." Lexicographers have to act accordingly.

Over the years, the critical voices became louder and louder, thus making a new case for bilingual learners' dictionaries (see Tomaszczyk 1983, Lew and Adamska-Sałaciak 2015, among others). At first, we saw the so-called semi-bilingual dictionaries that were influenced by the then predominant language-didactic philosophy. They consisted of "traditional" L2 monolingual dictionaries that were bilingualized with the addition of equivalents in the learners' native language. The idea was to force the users to read the L2 definitions and, only in case they had problems, resort to the equivalents. It did not really work, so little by little came more "pure" bilingual L2–L1 dictionaries without definitions neither in L1 nor L2, i.e., only with L1 equivalents.

Generally, the renewal progressed without sufficiently taking account of the new digital technologies. On the one hand, the supporters of a paradigm shift acknowledged the high-quality features of monolingual dictionaries like

the Big Five and wanted to incorporate them in a new type of bilingual dictionary that would let "users have the best of both worlds" (Adamska-Sałaciak 2010: 123). On the other hand, many new dictionary projects seemed stuck in the old dichotomy between monolingual and bilingual. In an article on the recent development of learners' dictionaries, I argued that:

(...) the formal classification of dictionaries for foreign-language learners into monolingual and bilingual may be one of the major obstacles that prevent present-day lexicography from taking full advantage of the new technologies and designing the dictionaries that meet the real needs of foreign-language learners. (Tarp 2013: 426)

The introduction of digital technologies in lexicography is disruptive. It requires that the discipline be reconsidered from top to bottom, especially the compilation, storing, and presentation of lexicographical data. The fact that the results of the lexicographers' efforts are stored in databases implies that these results can be presented to the target users in differentiated and needs-adapted ways and quantities through the method of filtering (see Bothma 2011). A digital dictionary — that is, the set of lexicographical data visualized in user interfaces — can now elegantly combine features of traditional monolingual and bilingual dictionaries. The lexicographical data stored in the database and those uploaded in user interfaces are not necessarily identical. Some data can even be stored mainly or exclusively for compilation purposes.

3.3 The Valladolid experience

I mentioned previously that the replacement of the terms *source* and *target language* with *object* and *auxiliary language* has considerable practical implications. Whereas bilingual L2–L1 dictionaries, as a rule, are either bilingualized extensions of existing L2 dictionaries or built on an independently selected stock of L2 lemmata, bilingual L1–L2 dictionaries have so far been compiled from a selection of L1 lemmata. There are very few exceptions to this. One of these is a mono-directional, biscopal English–Spanish dictionary project, which Pedro Fuertes-Olivera is working on in Valladolid. Its target users are Spanish native speakers, which means that the object language is English and the auxiliary one Spanish. This vision informed the decision to apply an innovative compilation methodology in the project, as summarized by Fuertes-Olivera et al. (2018: 160):

Whereas traditional mono-directional, biscopal dictionary projects usually take their point of departure in the users' native language, the Valladolid project does the opposite. It starts with a selection and description of English lemmata including separation in senses, definitions, Spanish equivalents, grammar, etc. An automatic and simultaneous inversion is then made where the Spanish equivalents to one of more English lemmata become new lemmata whereas the English lemmata become equivalents with the brief Spanish definitions used as meaning discrimination.

The real innovation here is the use of L2 definitions written in L1 as semantic differentiators in an L1–L2 dictionary. The inversion of dictionaries — or lexicographical databases, which is a more precise term — should be done without losing semantic content. Except for some specialized dictionaries within culture-independent disciplines, this challenge has cast a shadow over virtually all previous attempts of automatic dictionary inversion. The described compilation methodology, which Ordbogen's programmers incorporated into the Dictionary Writing System in close collaboration with the project editor, should ensure this. It worked! There were, of course, various new challenges — among them, the nature of the L2 definitions, the revision of the L1–L2 part, and the lexicographer's new role — which I will discuss and further elaborate on in the following.

3.4 What is to be done?

The main drive behind the methodology used in the Valladolid project was to raise productivity without compromising quality. I regard the described methodology as an important step forward in achieving this goal. But the experience from the machine-translation experiments indicates that there is still a job to do. I thus returned to an old idea that I discussed but left unfinished in my second doctoral thesis (Tarp 2008). The idea took form during the lecture of a large body of relevant literature (some of it mentioned above) and the study of hundreds of dictionaries, among which the Swedish LEXIN project targeting immigrants from more than 20 language communities was particularly inspiring (see Gellerstam 1999). But the idea could not be fully developed at that moment due to my insufficient knowledge of and experience with digital technologies.

I realized that the research stay at Ordbogen now gave me a unique chance to refine the idea and make it directly implementable. It goes without saying that it took a great deal of reflection to connect the dots and get the complete picture. I got the inspiration not by buying pastries for five cents down at the bakery, like Jack London, but on long evening walks in the countryside around Odense. All in all, I must have walked over 150 km in these beautiful surroundings to clear my thoughts. But it was worth it! The main idea, which I presented to Henrik Hoffmann, was as follows:

1. A monolingual L2 lexicographic database is compiled, intended for L1 users. This core database serves as a *basis* for the further process. It resembles Scerba's "defining dictionary" and incorporates, at the same time, many relevant features from the Big Five.
2. The core database is bilingualized into an L2–L1 database. The conversion is done by automatically translating definitions and example sentences (and metalanguage) and assigning L1 equivalents to the respective senses of each lemma.

3. The relationship between L2 lemmata and L1 equivalents contained in the database is inverted, so that the latter become lemmata and the former equivalents. The original and now translated definitions of L2 words — together with the corresponding L2 example sentences and their L1 translations — follow these words throughout, including during their metamorphosis from lemmata to equivalents. In this process, the definitions themselves slough their skin and transmute into differentiators. The new L1–L2 relationship is not a traditional one that contrasts two languages but a bridge from L1 into L2. This difference is attributable to the definition-turned-into-differentiator that directly states the meaning of the relevant sense of the L2 equivalent without focusing on the semantic differences between the treated sense of the L1 lemma and its L2 equivalent.
4. This methodology increases productivity because much of the process is mechanical. But it also improves the overall quality of the final product since L2 words — whether they appear as lemmata or equivalents — are never detached from their original semantic and syntactic L2 universe, as frequently happens in traditional L1–L2 dictionaries.
5. As a spin-off, the idea also facilitates economy of scale. Applying the described methodology, once a core database has been compiled in one language, say English, it is relatively easy to convert it into a set of bilingual dictionaries between this language and (many) other languages like Spanish, Danish, German, French, and Chinese.

4. Testing the idea

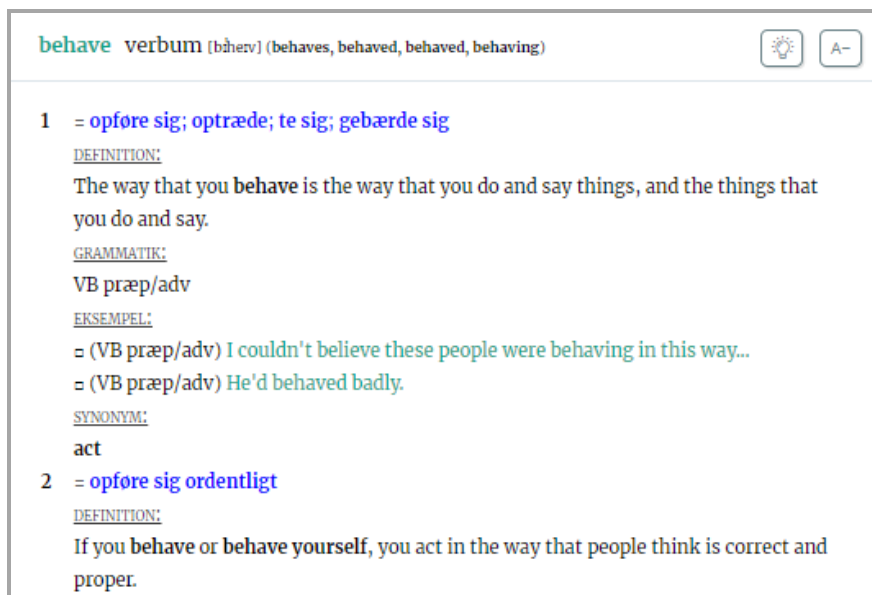
The discussion with Henrik Hoffman was difficult but highly productive. I presented the idea, he listened and had some suggestions which I opposed. He insisted that we reuse some of the lexicographical data in Ordbogen's databases, arguing that it would be more convincing. I rejected this idea because I was afraid it might jeopardize the quality of the core database and, consequently, its bilingual extensions. Both of us stood firm on our opinions, and the discussion went on for several days. Little by little, we began understanding each other's points of view and eventually reached a consensus. It was a beautiful interdisciplinary experience.

4.1 The first test

When Henrik Hoffman had grasped the idea, he immediately sat down and prepared an ad-hoc program in just two hours! To test it, he uploaded data from Politiken's English–Danish dictionary for advanced learners (*Store Engelskordbog*), which Ordbogen took over from the Politiken Publishing House after the latter's dictionary department closed down. The dictionary is a "semi-bilingual"

extension of *Collins Cobuild English Dictionary*, which includes Danish equivalents placed before (!) the English definitions, and has the metalanguage reproduced in Danish (see Figure 3). He chose this particular dictionary because it was easier to test, as it already contains Danish equivalents. When we typed an English lemma in the search field and clicked on the magnifier, the program automatically translated the definitions and example sentences using DeepL.

Figure 4 shows the result. It is not a dictionary article but only the visualization of some of the lexicographical data stored in the database. How it eventually will be presented to the users depends on the specific purpose, i.e., whether it is for a dictionary, an e-reader, or a writing assistant (see Fuertes-Olivera and Tarp 2020). The figure gives us a brief idea of the quality of the translated example sentences, which is, at least, as high as the quality of the translated definitions tested in Section 2. Contrary to the latter, these sentences represent natural language, so it is hardly a surprise. In this test, we were only interested in the program's functionality and immediate performance. We thus uploaded, studied, and saved a number of L2 lemmata with the attached data (definitions translated into L1, L1 equivalents, L2 example sentences, and their L2 translations). We then skipped the revision and correction phase and went directly to the Danish–English part, where we analyzed the data assigned to some of the inverted Danish lemmata. One of these was the reflexive verb *opføre sig*, which appeared twice as equivalent to *behave* in the English–Danish part (see Figure 5).



behave verbum [bɛhɛv] (behaves, behaved, behaved, behaving)

1 = **opføre sig; optræde; te sig; gebærde sig**

DEFINITION:
The way that you **behave** is the way that you do and say things, and the things that you do and say.

GRAMMATIK:
VB præp/adv

EKSEMPEL:
= (VB præp/adv) I couldn't believe these people were behaving in this way...
= (VB præp/adv) He'd behaved badly.

SYNONYM:
act

2 = **opføre sig ordentligt**

DEFINITION:
If you **behave** or **behave yourself**, you act in the way that people think is correct and proper.

Figure 3: Extract of article from Politiken's English–Danish dictionary

The screenshot shows a search interface for the word 'behave'. The search bar contains 'behave' and the language pair is set to 'English Danish'. The entry for 'behave' is shown as a verb. It lists the forms 'behaves, behaved, behaved, behaving'. A numbered definition follows: '1. Den måde, du opfører dig på, er den måde, du gør og siger ting på, og de ting, du gør og siger.' Below this, there is a section for 'Ækvivalenter:' (equivalents) with a bulleted list: 'opføre sig', 'optræde', 'te sig', and 'gebærde sig'. Finally, there is a section for 'Eksempler:' (examples) with two bullet points: 'I couldn't believe these people were behaving in this way... - Jeg kunne ikke tro, at disse mennesker opførte sig på denne måde...' and 'He'd behaved badly. - Han havde opført sig dårligt.'

Figure 4: Translation of sense from Politiken's dictionary

The screenshot shows a search interface for the phrase 'opføre sig'. The search bar contains 'opføre sig' and the language pair is set to 'Danish English'. The entry for 'opføre sig' is shown as a verb. It includes a numbered definition: '1. (Den måde, du opfører dig på, er den måde, du gør og siger ting på, og de ting, du gør og siger.)'. Below this, the English word 'behave' is listed. A section for 'Eksempler:' (examples) follows with two bullet points: 'I couldn't believe these people were behaving in this way... - Jeg kunne ikke tro, at disse mennesker opførte sig på denne måde...' and 'He'd behaved badly. - Han havde opført sig dårligt.' A second numbered definition is provided: '2. (Inden for videnskaben er den måde, noget opfører sig på, de ting, det gør.)'. Below this, the English word 'behave' is listed again. A final section for 'Eksempler:' (examples) has one bullet point: 'Under certain conditions, electrons can behave like waves rather than particles. - Under visse betingelser kan elektroner opføre sig som bølger i stedet for partikler.'

Figure 5: Inversion with ad-hoc program

We learned several things from this first test. First of all, that the overall idea does work but not surprisingly needs further refinement. It confirmed that special attention should be paid to the translated L2 definition, as it is the central axis around which the whole project rotates. In the test, we used Collins Cobuild's innovative "new definition" written in a natural language similar to the one that teachers use when explaining L2 words to their students. This defining technique implies that the defined word is part of the definition. One may like or dislike this type of definition (personally, I like it). But as can be seen in Figure 5, it does not work in our context because the translation defines the L1 word instead of the L2 one.

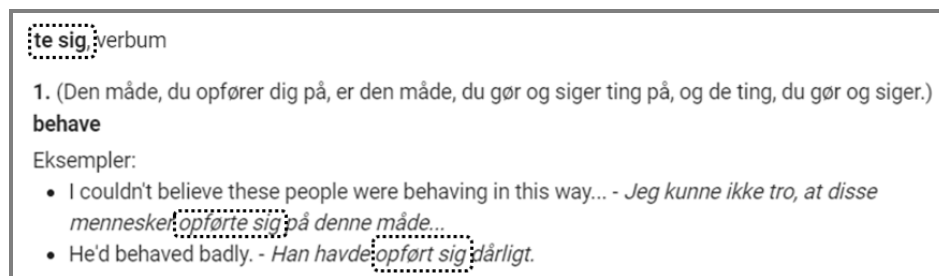


Figure 6: Discrepancy between lemma and translated example sentence

Another minor problem detected was the translation of example sentences when more than one equivalent were assigned to an L2 sense, for instance, the first sense of *behave* in Figure 4. In this case, the translations include the first of these equivalents (*opføre sig*). However, if one of the other equivalents appears as a lemma in the Danish-English part (e.g., *te sig*), there will be a certain discrepancy between the lemma and the translated example sentence (see Figure 6). It is not a big problem because the example sentence here aims at illustrating a specific syntactic property of *behave*, and the translation is merely provided to assist L1 users at a low L2 proficiency level who do not understand it. But overall it would be better to adjust it. Finally, we had no trouble admitting that we were lucky to find a learners' dictionary that provided both L2 definitions and L1 equivalents. Such dictionaries are rare, at least with the required quality. This takes us to the second test.

4.2 The second test

Henrik Hoffmann had now improved the ad-hoc program adding more functionalities. This time, we decided to upload a number of lemmata inclusive definitions from the *Oxford Dictionary of English*, which Ordbogen also licenses. The definitions were automatically translated with DeepL and could also be

corrected if there were any mistakes. The corrections, however, had to be done manually, as we did not yet have the license to build all DeepL's services into the software.

When the translations were accepted and saved, a new page appeared with two columns. The left one displayed the translated definitions, whereas the right one offered possible equivalents to the lemma. The program automatically retrieved these L1 candidates from Ordbogen's self-produced English–Danish database. Now, I finally understood the full implications of what Henrik meant when he said he wanted to reuse lexicographical data from the company's own databases. It really makes things easier. After carefully reading the definitions of the various senses of the lemma, the lexicographer can assign the suggested equivalents to the appropriate senses with a simple movement (see Figure 7) and, if necessary, add additional ones. If we had used some of the other English–Danish databases that the company manages, the number of suggested equivalents would probably be higher, thus facilitating the lexicographer's job even more. In any case, the subsequent inversion will proceed much smoother if as many adequate L1 equivalents as possible are assigned to each sense of the respective L2 lemmata.

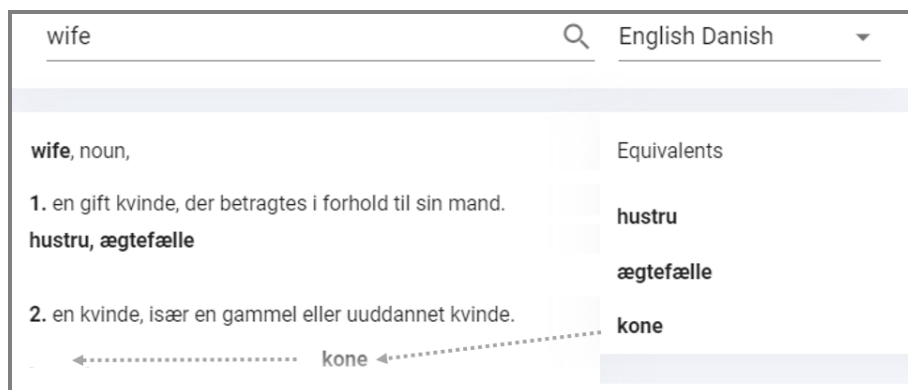


Figure 7: Assigning suggested L1 equivalents to senses

In this second test, we also learned more about the requirements for the original L2 definitions. Most of them worked perfectly, but others, like the first definition of *dragon*, were more problematic:

a mythical monster like a giant reptile. In European tradition the dragon is typically fire-breathing and tends to symbolize chaos or evil, whereas in the Far East it is usually a beneficent symbol of fertility, associated with water and the heavens.

This text is perfect as a lexicographical definition of *dragon*. But it is simply too wordy if it also has to function as a meaning differentiator in an L1–L2 diction-

ary, which is one of the possible products where the lexicographical data can be used. Pedro Fuertes-Olivera had similar challenges in his bilingual Valladolid project. The first small text segment (*a mythical monster like a giant reptile*) would serve very well as a differentiator. In this specific case, the remaining text provides additional information contrasting Eastern and Western culture. Other wordy definitions also have a similar structure with a small initial text segment followed by a longer one with more detailed information. In these cases, the problem seems to be the storage of the two parts, which does not allow them to be uploaded separately. The solution is here a case for prediction and interdisciplinary planning of databases. However, some wordy definitions do not have a similar internal structure and, thus, need to be modified accordingly. The Valladolid experience shows that it is possible to meet this requirement after some lexicographical training.

5. Some challenges

The tests show that machine translation can play a relevant role in the compilation of lexicographical databases and that the method presented in Section 3.4 to inverse the relationship between lemmata and equivalents is implementable. It suggests that overall the project is a success. It confirms the importance of never separating the L2 words from their definitions. However, the test of a relatively small amount of data does not provide answers to all questions. Many challenges remain, some of them predictable, others still to be detected. Here, I will briefly comment on two of the predictable ones.

One of these challenges is related to the stock of L1 lemmata once the inversion has taken place. For the benefit of future users, the most frequently consulted L1 lemmata must be part of this stock. The easiest way to verify if this is the case is to use log files from previous consultations. Although some L1 lemmata may be missing, the corresponding L2 senses will probably appear in the database in most cases. This shows the extraordinary importance of assigning as many L1 equivalents as possible to each sense of the L2 lemmata, preferably when preparing the L2–L1 part and alternatively when revising the L1–L2 part. Even so, some relevant single-word or multi-word L1 lemmata or their senses may still be absent. It is premature to develop the precise methods to solve these challenges based on the relatively limited experience from the above tests. It will have to wait until much more empirical data is available, probably when inverting a whole L2–L1 part. In any case, this is one of the tasks where the skilled lexicographer still has an important role to play.

Another predictable challenge is to determine to what extent existing data can be reused in a new dictionary project. There is no general answer to this question. Each publishing house will have to decide for itself. It will depend on the quality of the lexicographical data that is already available in its databases.

6. The human lexicographer's new role

As I wrote in Section 1, one of the objectives of my project at Ordbogen was to know more about the changing relationship between the human and what I call the artificial lexicographer. The latter refers to digital software that completes either compilation tasks, which the human lexicographer previously carried out, or entirely new tasks. It is not surprising at all that the human lexicographer's role in the compilation process changes. Something similar has happened throughout the history of lexicography whenever disruptive technologies have been applied. However, the current situation differs from past experiences in that the artificial lexicographer is increasingly replacing the human lexicographer and redefining his or her role. The main tendency today is that the human lexicographer dedicates more time to revising lexicographical data and less time to creative activities. This tendency seems inescapable. The project described above is no exception. Here, I will briefly list the tasks that are either new or different from what is common in other projects:

Core L2 database

- This is the part that involves the most creative work. Apart from the technological improvements that characterize all monolingual projects, the only thing new is the requirement for wordy definitions. These items should be structured in two parts to be stored separately in the database.

Bilingualized L2–L1 part

- checking translated definitions and reworking those that are erroneous or inadequate.
- checking translated example sentences and reworking those that are erroneous or inadequate.
- assigning as many L1 equivalents as possible to each sense of the L2 lemmata, either from a list of suggested candidates or using other methods, including introspection.

Inverted L1–L2 part

- checking the inversion in general, lemma by lemma.
- checking whether the most frequently consulted L1 lemmata appear after the inversion and whether their most relevant senses have L2 equivalents assigned.
- adding frequently consulted L1 lemmata when they are missing.
- assigning L2 equivalents to relevant L1 senses without equivalents.
- possibly checking translated example sentences to verify that they include the pertinent L1 lemma and not another L1 lemma (see Figure 6).

For the moment, this is all. But there may be other tasks to perform, or the ones mentioned may be slightly modified, when more experience is available.

7. Postscript

After testing the idea developed in Section 3.4, both Henrik Hoffmann and I believe it has a future. Machine translation has proven useful in lexicography, and inversion has become possible. But we are aware that there is still some work to do, both on the technological front and the lexicographical front. The experience shows that interdisciplinary collaboration is the best way to achieve fast and robust results. As stated in the introduction, I feel privileged to have had the opportunity to integrate the R&D team at Ordbogen. I urge other lexicographers to engage in similar forms of collaboration. The discipline is in great need of innovation in many aspects. The potential of the new technologies must be thoroughly explored. New ideas and hypotheses must be tested without fear of temporary setbacks. Everything useful must be incorporated to improve the compilation, storage, and presentation of lexicographical data. Nobody can do this alone.

Acknowledgements

I wish to thank the School of Communication and Culture, Aarhus University, for facilitating my research stay, and Ordbogen A/S for accepting my project and incorporating me into its R&D team.

Special thanks are due to Programmer and Web Developer Henrik Hoffmann for his patience, constructive comments, and technical assistance that allowed me to test and further develop my ideas.

Special thanks are also due to Prof. Rufus H. Gouws, Stellenbosch University, and Prof. Pedro Fuertes-Olivera, Valladolid University, for their willingness to discuss and comment on my research project, both during and after the stay at Ordbogen.

Finally, I also wish to thank Prof. Heming Yong, Guangdong University of Finance, Prof. Patrick Leroyer, and Prof. Martin Nielsen, both Aarhus University, for revising sample translations into Chinese, French, and German, respectively.

References

A. Machine translators

DeepL Translator: <https://www.deepl.com/translator>.

Google Translate: <https://translate.google.co.za/>.

B. Other literature

- Adamska-Sałaciak, A.** 2010. Why We Need Bilingual Learners' Dictionaries. Kernerman, I.J. and P. Bogaards (Eds.). 2010. *English Learners' Dictionaries at the DSNA 2009*: 121-137. Tel Aviv: K Dictionaries.
- Adamska-Sałaciak, A. and I. Kernerman.** 2016. Introduction: Towards Better Dictionaries for Learners. *International Journal of Lexicography* 29(3): 271-278.
- Bergenholtz, H., I. Cantell, R.V. Fjeld, D. Gundersen, J.H. Jónsson and B. Svensén.** 1997. *Nordisk Leksikografisk Ordbok*. Oslo: Universitetsforlaget.
- Bothma, T.J.D.** 2011. Filtering and Adapting Data and Information in an Online Environment in Response to User Needs. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.
- Burkhanov, I.** 1998. *Lexicography. A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- Duda, W.** 1986. Ein "aktives" russisch-deutsches Wörterbuch für deutsch-sprachige Benutzer? Günther, E. (Ed.). 1986. *Beiträge zur Lexikographie slawischer Sprachen*: 9-15. Berlin: Akademie-Verlag.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter.
- Fuertes-Olivera, P.A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.
- 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 Diccionarios Valladolid-UVa. *Lexikos* 28: 152-176.
- Gellerstam, M.** 1999. LEXIN — lexikon för invandrare. *LexicoNordica* 6: 3-18.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hausmann, F.J.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Niemeyer.
- Lew, R. and A. Adamska-Sałaciak.** 2015. A Case for Bilingual Learners' Dictionaries. *ELT Journal* 69(1): 47-57.
- Marelo, C.** 2003. The Bilingual Dictionary: Definition, History, Bidirectionality. Hartmann, R.R.K. (Ed.). 2003. *Lexicography. Critical Concepts II*: 325-342. London/New York: Routledge.
- Martínez de Sousa, J.** 1995. *Diccionario de Lexicografía Práctica*. Barcelona: Bibliograf.
- Mikkelsen, H.K.** 1992. What Did Scerba Actually Mean by "Active" and "Passive" Dictionaries? Hyldgaard-Jensen, K. and A. Zettersten (Eds.). 1992. *Symposium on Lexicography V. Proceedings of the Fifth International Symposium on Lexicography May 3-5, 1990, at the University of Copenhagen*: 25-40. Tübingen: De Gruyter Mouton.
- Scerba, L.V.** 1940 (1995). Towards a General Theory of Lexicography. *International Journal of Lexicography* 8(4): 314-350.
- Tarp, S.** 2005. The Concept of a Bilingual Dictionary. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 27-41. Frankfurt am Main: Peter Lang.
- 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.** 2013. New Developments in Learner's Dictionaries III: Bilingual Learner's Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 425-430. Berlin: De Gruyter.
- Tarp, S.** 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249.
- Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
- Tomaszczyk, J.** 1983. On Bilingual Dictionaries: The Case for Bilingual Dictionaries for Foreign Language Learners. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 41-51. London: Academic Press.

Mariusz Piotr Kamiński. *Defining with Simple Vocabulary in English Dictionaries.* 2021, xvi + 326 pages. ISBN 9789027208590 (Hardback), ISBN 9789027260000 (e-Book), ISSN 1388-8455. Terminology and Lexicography Research and Practice 22. Amsterdam/Philadelphia: John Benjamins. Price € 99.00 (Hardback and e-Book).

This book must be the most thorough and comprehensive up-to-date account of restricted vocabularies of the English language. In the Introduction, the author formulates six research questions, which partly aim at the design of different restricted vocabularies, and, partly focus on the effectiveness of definitions based on restricted vocabularies. Chapter 1 traces the development of restricted vocabularies back to the 17th century and outlines important stages such as Isaac Pitman's and Kaeding's shorthand systems. The author then makes a distinction between (a) the frequency-based objective approach taken by scholars such as Thorndike and Horn, (b) a more pedagogically oriented approach associated with Harold Palmer and Michael West, and (c) a logical approach attributed to Ogden and Richards, and discusses the pros and cons of each of them. The remainder of the chapter is devoted to a very detailed outline and critical discussion of Michael West's famous and influential *General Service List*. He comes back to the GSL after outlining a number of word list projects that were designed for pedagogical purposes in Chapter 2, which also deals with human-computer communication projects and Wierzbicka's theory of a semantic metalanguage.

While the first two chapters deal with the design of restricted vocabularies from various perspectives, Chapter 3 directs the focus on the issue of lexicographical definition. The author provides a very knowledgeable survey of defining policies in English lexicography ranging from Dr Johnson to modern learners' dictionaries. Kamiński draws an interesting parallel between the rise of monolingual foreign language teaching and the need for simple definitions and discusses the controlled defining vocabulary approach taken, e.g. in the 1st edition of the *Longman Dictionary of Contemporary English*, the research it instigated and the effects on other dictionaries. At the end of this chapter, the research questions already outlined in the introduction are taken up and formulated in more detail.

Chapters 1 to 3 provide the basis for the empirical analyses that are being described in the second part of the book. This involves a quantitative comparison of selected editions of learners' dictionaries between 1987 and 2015 (OALD, LDOCE, COBUILD, CALD, MEDAL and MWALED), finding, for example, that "the vast majority (87%) of the vocabulary items shared by the early lists has survived into the recent lists" (p. 169). In the framing of the book, the findings of the quantitative analysis are presented in a slightly complicated manner: "One of the aims of this chapter was to test the hypothesis that the RVLs of definitions in the latest editions of learners' dictionaries vary considerably (H1). This hypothesis cannot be substantiated by the findings of the analysis, as the

vocabulary loads are comparable (3,000–4,000 word families)" (p. 168) While in many experimental designs an approach in terms of postulating a hypothesis that will be falsified or verified certainly makes sense, here it has a slightly pseudo-scientific ring to it, especially if the hypothesis formulated seems rather counter-intuitive, as in this case.

This quantitative analysis is complemented by a qualitative one, which is based on 51 entries in various learners' dictionaries, where Kamiński comes to the interesting conclusion that there have been improvements in the restricting defining vocabularies and that these are most noticeable in the editions published around the turn of the century. He then goes on to show that at least in some native-speaker dictionaries, notably the COD, similar improvements towards greater comprehensibility can be observed.

Chapter 6 extends the scope of analysis to the user. In order to test the effect of the various defining policies, experiments were carried out with a substantial number (215) of university students of English, who were native speakers of Polish (and an additional 35 with neither English nor Polish as L1). The results show a clear advantage of definitions that are based on controlled defining vocabularies over those that are not (p. 254). The final conclusions chapter provides a summary of the preceding analyses in the form of a list of useful guidelines for the designing of defining vocabularies and the phrasing of definitions in monolingual dictionaries, which will certainly be of great value for future lexicographic projects (not only of English).

On the whole, this book is a very thorough and systematic analysis of the defining approaches of a large number of English dictionaries. Although the amount of statistical evidence is impressive, a few more examples of direct comparisons of actual dictionary entries might have supported the case even more strongly. In any case, this book is a valuable contribution to research on controlled vocabularies in lexicography providing a very useful account of previous research on the topic and adding relevant empirical data.

A final point of critique, however, concerns the title, because the present volume is not only relevant to lexicographers, but surely also of great interest to applied linguists and foreign language teachers.

Thomas Herbst

*Department of English and American Studies
Friedrich-Alexander-Universität Erlangen-Nürnberg*

Erlangen

Germany

(thomas.herbst@fau.de)

Differentiated Treatment of Cultural Items in Lexicographical Products: A Necessary Adaptation to the Digital Environment

Qian Li, *Centre for Lexicographical Studies, Guangdong University
of Foreign Studies, China (lqchristina@gdufs.edu.cn)*

and

Sven Tarp, *Centre for Lexicographical Studies, Guangdong University
of Foreign Studies, China; Department of Afrikaans and Dutch, Stellenbosch
University, South Africa; and Centre for Lexicography, Aarhus University,
Denmark (st@cc.au.dk)*

Abstract: The paper focuses on the lexicographical treatment of cultural objects. It argues that second-language learning requires second-culture learning and that digital technologies call for new solutions to both old and new challenges. As an example, it takes traditional Chinese musical instruments and starts with a critical analysis of their treatment in five Chinese–English dictionaries for both foreign learners and native speakers. It continues with some reflections on media convergence and its consequences for lexicography and reaches the conclusion that the one-size-fits-all dictionary must be replaced with a variety of lexicographical products on different platforms. Lexicographers' focus must therefore move from the dictionary to the database that supports these products. This leads to a discussion of equivalent and explanation types and the need to prepare four different database fields for equivalents and two for explanations. To exemplify this, the paper presents a lexicographical database with equivalents, explanations, and other types of culturally relevant items. It then uses a few examples to show how these lexicographical items stored in the database can be selectively employed on different platforms and adapted to specific user needs. The paper links directly to sound files and video clips with some of the discussed instruments. Finally, the paper provides some conclusions and perspectives for further improving the cultural dimension of learners' lexicography.

Keywords: MEDIA CONVERGENCE, CHINESE–ENGLISH DICTIONARIES, LEARNERS' DICTIONARIES, E-READERS, DIGITAL DEVICES, LEXICOGRAPHICAL DATABASES, EQUIVALENT TYPES, EXPLANATION TYPES, CULTURAL ITEMS

Opsomming: Die gedifferensieerde hantering van kulturele items in leksikografiese produkte: 'n Noodsaaklike aanpassing by die digitale omgewing.

In hierdie artikel word daar op die leksikografiese hantering van kultuurobjekte gefokus. Daar

word aangevoer dat die aanleer van 'n tweede taal die aanleer van 'n tweede kultuur vereis en dat digitale tegnologie nuwe oplossings vir beide ou en en nuwe uitdagings vra. As voorbeeld word tradisionele musiekinstrumente gebruik en daar word eerstens 'n kritiese analise van hul hantering in vyf Chinees–Engelse woordeboeke vir beide vreemdetaalleerders en moedertaalsprekers gedoen. Daarna word verslag gedoen oor mediakonvergensie en die gevolge daarvan vir die leksikografie, en die gevolgtrekking word gemaak dat die universeel-geskikte woordeboek met 'n verskeidenheid leksikografiese produkte op verskillende platforms vervang moet word. Daarom moet die fokus van die leksikograaf verskuif van die woordeboek na die databasis wat hierdie produkte ondersteun. Dit lei tot 'n bespreking van ekwivalente en verklarende tipes en die behoefte daaraan om vier verskillende databasisvelde vir ekwivalente en twee vir verklarings te skep. Ter illustrasie word 'n leksikografiese databasis met ekwivalente, verklarings, en ander tipe kultuurrelevante items aangebied. Daarna word 'n paar voorbeelde gebruik om aan te toon hoe hierdie leksikografiese items wat in die databasis gestoor word, selektief op verskillende platforms gebruik kan word en vir spesifieke gebruikersbehoefte aangepas kan word. Hierdie artikel is direk gekoppel aan klankleërs en videogrepe van die bespreekte instrumente. Laastens verskaf die artikel 'n paar gevolgtrekkings en perspektiewe om die kulturele dimensie van aanleerdersleksikografie te verbeter.

Sleutelwoorde: MEDIAKONVERGENSIE, CHINEES–ENGELSE WOORDEBOEKE, AANLEERDERSWOORDEBOEKE, E-LESERS, DIGITALE TOESTELLE, LEKSIKOGRAFIESE DATABASISSE, EKWIVALENTE TIPES, VERKLARENDE TIPES, KULTURELE ITEMS

There is an interactive relation between language and culture and familiarity with many aspects of this relation can be regarded as necessary for the best possible comprehension of language and the appropriate use of language in typical situations of use.

(Gouws 2020: 3)

1. Introduction

Lexicography is at a crossroads. The continuous introduction and application of new digital technologies imply that many practical and theoretical aspects of the millennial discipline need to be reconsidered. New solutions are required to both old and new challenges. One of these challenges is the treatment of culture in lexicographical products. Today, there is a big and growing body of literature dealing with this topic as reflected in the contributions by Zgusta (1989), Stark (1999), Sánchez (2010), Bergenholtz and Nielsen (2013), Kwary and Miller (2013), Hallsteinsdottir (2015), Klosa (2015), Miller et al. (2017), and Nomdedeu Rull (2020), among many others.

This paper focuses on Chinese lexicography, where a similar tendency characterizes the academic production, especially the one concerned with dictionaries for foreign learners of Chinese. Examples of this are articles by Tseng (2003), Xue (2017), Ma (2018), Xue and Tarp (2018), Huang (2020), Kang (2020), Li and Xia (2020), and Zhang and Mi (2020).

The lexicographical interest in culture is well-founded. Language and culture are interwoven. The former cannot be fully understood and correctly used without knowledge of the latter. The implications for second-language learning are evident. Each society expresses itself in a distinct culture with its particular customs and habits, beliefs and morals, arts and laws, social organization and administrative practices, production methods and artefacts, knowledge and skills, etc. All these phenomena develop over time and project themselves into the respective languages and their vocabularies. To be successful, second-language learning therefore presupposes a certain amount of second-culture learning. The more, the better. This holds particularly true for language learning when the distance between the learners' home culture and the foreign culture is as big as the one that exists between Chinese and Western cultures. Zhang and Mi (2020: 64-65) shrewdly observe:

Chinese culture-bound vocabulary is a key element in cultivating second language learners' language proficiency. Most learners of Chinese as a second language (CSL) are adults. Before learning Chinese, they already possess a relatively mature knowledge system and are less impressionable. In learning the second language, their primary challenging obstacle is overcoming the strong interference from their established knowledge of their mother tongue. This interference can only be eliminated, gradually, by activating the cultural mechanism contained in the Chinese vocabulary.

Lexicography can largely contribute to filling the cultural gap. This requires that its practical products target the intended users' real information needs when consulting these products. Mistaken approaches such as an irrelevant distinction between linguistic and encyclopedic data are not helpful. Nomdedeu Rull (2020: 41), for instance, reports that Spanish lexicographers have regarded the monolingual dictionary as "a reference tool with linguistic data only", and that this misconception "has caused cultural data to be excluded from this type of dictionary because they are considered encyclopedic". Users in need of cultural information have been referred to other reference works such as encyclopedias. Another Spanish scholar (Fuertes-Olivera 2013: 1069) describes this practice as "ill-conceived, irrelevant and potentially confusing". Similar problems can be found in other lexicographical traditions. They are particularly harmful when the lack of pertinent data prevents learners from "activating the cultural mechanism contained in the [...] vocabulary", to which Zhang and Mi (2020) referred above.

More than three decades ago, Zgusta (1989: 3) discussed the cultural information, which learners may need during second-language learning and summarized that:

a good part of this information is of encyclopedic character; be this as it may, it belongs to what the learner has to learn.

From a user-oriented perspective, the important thing is that the lexicographi-

cal data meet the learner's real needs. No relevant data should be excluded, whatever category they belong to. The printed dictionary format certainly has some limitations in this respect. Digital technologies, by contrast, open up new and better ways of meeting user needs. If fully exploited, they will have a disruptive impact on the lexicographical product as they allow a more focussed, flexible, and personalized service that also includes new data categories presented in items like videos and soundtracks. Current Chinese lexicographers often refer to the inclusion of such items as *media convergence*; see Zhang (2019).

In the following, we will take this idea as the basis and elaborate on some of its possible consequences for the lexicographical treatment of cultural objects. As we saw above, culture is a very broad notion that contains a variety of different challenges, each of which requires a special approach. We will, therefore, focus on the assistance, which dictionaries and other lexicographical products offer to foreign learners of Chinese who need information about a specific class of cultural objects. To exemplify this, we have chosen traditional Chinese musical instruments, which are not only interesting as "culture-related objects" unique to Chinese culture (Li and Xia 2020: 125), but also in terms of how and when they are used and what customs are related to their usage.

Section 2 will analyze and discuss the treatment of these musical instruments in various Chinese–English dictionaries. Section 3 will summarize the findings whereas Section 4 will discuss some of the challenges posed by the media convergence age. Section 5 will follow up with some principles for the lexicographical treatment of cultural items and will then exemplify these principles in a concrete proposal. Finally, Section 6 will provide some conclusions and perspectives for future lexicographical projects.

2. Treatment in Chinese–English dictionaries

With a music history of over 7,000 years, Chinese culture is proud of its wide range of ancient, traditional musical instruments. When foreign learners of Chinese meet the term for a musical instrument in their contact with Chinese culture, dictionaries seem to be the first helper they resort to. Whether dictionaries can offer them appropriate and sufficient information should be the primary concern of lexicographers. To find out if this is the case, we will analyze the treatment of ten musical instruments, all of them ancient and typical of Chinese culture. Some of these instruments are still played in daily life, whereas others are displayed in museums as antiques. The selected musical instruments are:

- (1) 编钟 (bianzhong)
- (2) 古琴 (guqin)
- (3) 箜篌 (konghou)
- (4) 笛子 (dizi)
- (5) 埙 (xun)

- (6) 箏 (zheng)
- (7) 排箫 (paixiao)
- (8) 二胡 (erhu)
- (9) 琵琶 (pipa)
- (10) 梆子 (bangzi)

This Section aims at analyzing how the above musical instruments are treated lexicographically, and whether this treatment is beneficial for foreign learners who have comprehension problems when reading Chinese texts. To this end, we have selected the following major Chinese–English dictionaries:

- *A Chinese–English Dictionary* (ACED)
- *A New Century Chinese–English Dictionary* (NCCED)
- *The Chinese–English Dictionary* (TCED)
- *New Age Chinese–English Dictionary* (NA)
- *ABC Chinese–English Comprehensive Dictionary* (ABC)

These five dictionaries can be described as synchronic, general, and monoscopical. They are among the most commonly used Chinese–English dictionaries in China. Prior to the study, we also consulted two Chinese–English learners' dictionaries (*A Chinese–English Dictionary for Chinese Language Learning* and *A Chinese–English Dictionary for Foreign Learners*), which focus specifically on foreign learners of Chinese. However, these learners' dictionaries are relatively small-sized and do not treat most of the selected Chinese musical instruments. This study, therefore, focuses on the mentioned general Chinese–English dictionaries, which are intended for both Chinese native speakers and non-native learners of Chinese. In some cases where these dictionaries refer to specific Western musical instruments, we also quote from two monolingual English dictionaries (*Merriam-Webster Dictionary* and *Collins English Dictionary*). In the following, we will look at the treatment of the ten selected terms, one by one.

2.1 Treatment of 编钟 (bianzhong)

The five dictionaries offer the following definitions of *bianzhong* (see Figure 1, click [here](#) to download its sound and [here](#) to download a video clip):

- a set of bells, chimes (ACED)
- set of bells/chimes; serial/chime bells (NCCED)
- chime (TCED)
- chime bells; serial bells; carillon *see illustration* (NA)
- ancient musical instrument with 16 bells (ABC)

All five selected dictionaries treat the musical instrument *bianzhong* as a sublemma under the lemma 编 (*bian*). ACED provides a brief explanation followed by a translational equivalent (*chimes* in plural), whereas TCED only offers *chime* (in singular). ABC does not furnish any equivalent but only a brief explanation with the additional information that the musical instrument consists of 16 bells. NCCED gives two groups of equivalents with different modifiers (*set of*, *serial*, and *chime*).

NA differs from the other four dictionaries in that it shows three translational equivalents. The two first (*chime bells* and *serial bells*) are similar to the translational equivalent *chimes*, whereas the third one (*carillon*) refers to a Western musical instrument. Collins defines *carillon* as *a set of bells usually hung in a tower and played either by keys and pedals or mechanically*, and Merriam-Webster as *a set of fixed chromatically tuned bells sounded by hammers controlled from a keyboard*. The Chinese *bianzhong* and the Western *carillon* have common features in terms of manufacturing techniques and playing modes. It is, nonetheless, questionable whether the adoption of the translational equivalent *carillon* is appropriate and helpful to the users.

In addition, NA uses the indicator *see illustration* to refer its users to an illustration that shows the shape of the musical instrument. The reason for this is probably that *bianzhong*, as an ancient musical instrument with big serial bells, is no longer played in current daily life. It is usually exhibited in museums or just played by a few professional traditional music orchestras for a few cases. The illustration allows learners to be quickly informed about its unique shape and features.



Figure 1: Bianzhong

2.2 Treatment of 古琴 (guqin)

Four of the five dictionaries define *guqin* (see Figure 2, click [here](#) to download its sound and [here](#) to download a video clip):

- *guqin*, a seven-stringed plucked instrument in some ways similar to the zither (ACED)
- *guqin*; heptachord (a seven-stringed plucked instrument similar to the zither) (NCCED)
- *guqin*, a seven-stringed plucked instrument in some ways similar to the zither (TCED)
- zither-like seven-stringed instrument (ABC)

While NA does not list 古琴 (*guqin*) as a lemma, NCCED offers the equivalent *heptachord* to define *guqin*. *Heptachord*, defined as *an instrument with seven strings* in Collins, is marked as an obsolete word in this dictionary. It is a question whether the adoption of an obsolete word in the definition will meet users' need. Three dictionaries give almost the same treatment to this musical instrument. They provide the Chinese pinyin *guqin* as the transliterated equivalent followed by a brief explanation to show the instruments' form (*seven-stringed*) and playing mode (*plucked*). The fourth one, ABC, only offers a brief explanation. All explanations describe the instrument as *similar to the zither* or *zither-like*. This allows us to compare the treatment of *guqin* in the Chinese–English dictionaries to the treatment of *zither* in Merriam-Webster and Collins:

a stringed instrument having usually 30 to 40 strings over a shallow horizontal soundboard and played with pick and fingers. (Merriam Webster)

a zither is a musical instrument which consists of two sets of strings stretched over a flat box. You play the zither by pulling the strings with both hands. (Collins)

The definitions in the monolingual English dictionaries provide relatively more details about *zither*, including its shape (*over a shallow horizontal soundboard* or *stretched over a flat box*), its form (*30–40 strings*), and its playing way in details (*played with pick and fingers* or *pulling the strings with both hands*). This information may benefit dictionary users who want to get a clear idea of the instrument.

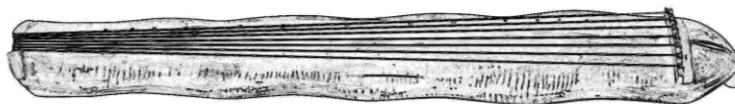


Figure 2: Guqin

2.3 Treatment of 箜篌 (konghou)

All dictionaries define *konghou* (see Figure 3, click [here](#) to download its sound):

- an ancient plucked stringed instrument (ACED)
- *konghou* [an ancient Chinese stringed instrument played by plucking] (NCCED)
- *konghou*, an ancient plucked stringed instrument (TCED)
- ancient plucked stringed instrument (with five to twenty-five strings) (NA)
- ancient harp (ABC)

Two dictionaries (NCCED and TCED) adopt the transliteration approach and offer the Chinese pinyin *konghou* of the lemma 箜篌 as the transliterated equivalent. Furthermore, three dictionaries (ACED, TCED, and NA) present almost the same explanation (*an ancient plucked stringed instrument*) that does not identify any distinctive features of *konghou*. Users may find it difficult to distinguish between *guzhin* and *konghou* when they only read the explanation of these two instruments as *plucked stringed instrument*. The explanations offered by the three dictionaries do not help dictionary users understand what exactly *konghou* is.

ABC provides the explanatory equivalent *ancient harp* to define *konghou*. *Konghou* and *harp* are similar in their shape and fundamental playing mode. *Harp* is defined as follows in the two monolingual dictionaries:

a plucked stringed instrument consisting of a resonator, an arched or angled neck that may be supported by a post, and strings of graded length that are perpendicular to the soundboard. (Merriam Webster)

A harp is a large musical instrument consisting of a row of strings stretched from the top to the bottom of a frame. You play the harp by plucking the strings with your fingers. (Collins)

Webster gives a technical description of its components, while Collins explains both the playing mode and the basic structure of the instrument. Both dictionaries present a picture of *harp* as well.



Figure 3: Konghou

2.4 Treatment of 笛子 (dizi)

Dizi (see Figure 4, click [here](#) to download its sound and [here](#) to download a video clip) is defined as follows:

- *dizi*; bamboo flute (ACED)
- bamboo flute (NCCED)
- flute; bamboo flute (TCED)
- Chinese (eight-holed) bamboo flute (NA)
- bamboo flute (ABC)

All selected dictionaries have chosen *bamboo flute* as the translational equivalent for this musical instrument. The TCED adds *flute* and NCCED pinyin *dizi* as the first equivalent, whereas NA contains more details (*Chinese, eight-holed*). Flutes also exist in other countries. The Chinese bamboo flute differs from the Western flute in that the former has a bamboo membrane on the second hole from the left end to form a resonator. The membrane helps create a wave in the flute, and its vibration makes the sound clearer and brighter. The mere provision of equivalents does not allow users to distinguish between the Chinese bamboo flute and the Western flute.

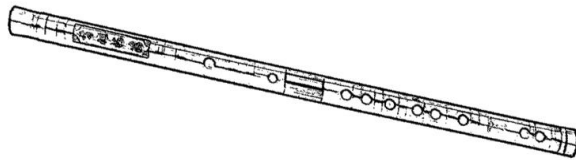


Figure 4: Dizi

2.5 Treatment of 埙 (xun)

The five dictionaries define *xun* (see Figure 5, click [here](#) to download its sound) as follows:

- an ancient egg-shaped, holed wind instrument (ACED)
- oval earthen wind instrument with six holes (NCCED)
- an ancient egg-shaped holed wind instrument (TCED)
- ancient Chinese wind instrument, made of clay with one to six holes and shaped like an egg (NA)
- ancient wind instrument (ABC)

ACED and the TCED include the same definitions of *xun* in terms of its shape

(egg-shaped, holed) and its playing mode (*wind instrument*). NCCED and NA furthermore inform about the material (*earthen, clay*) and formal details (*with one to six holes*). ABC only offers a short definition with no details about shape and playing mode. These data allow users to get an overall picture of *xun*, but they do not explain how this wind instrument is played.

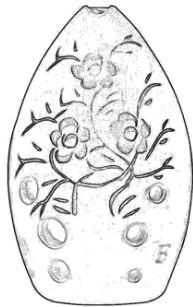


Figure 5: Xun

2.6 Treatment of 箏 (zheng)

Zheng (see Figure 6, click [here](#) to download its sound and [here](#) to download a video clip) is defined as follows in the five dictionaries:

- *zheng*, a 21- or 25-stringed plucked instrument in some ways similar to the zither (ACED)
- *zheng*, [a 21- or 25-stringed plucked instrument somewhat similar to the zither] (NCCED)
- *zheng*, a 21- or 25-stringed plucked instrument in some ways similar to the zither (TCED)
- Chinese zither with 21 or 25 strings *see illustration* (NA)
- 21–25 stringed plucked instrument similar to the zither (ABC)

Guqin and *zheng* are both ancient Chinese musical instruments and look very similar (see Figures 2 and 6). The selected dictionaries also treat them in the same way by depicting their form and playing mode. According to the dictionaries, the only difference between the two instruments is the number of strings, 7 strings for *guqin* and 21–25 strings for *zheng*. There are, however, other differences. People still play *zheng* in daily life, whereas *guqin* generally is exhibited in museums and rarely played in common life. Besides, they are played with different fingers due to their different string numbers, and their sounds are easily distinguishable. The information which the user can retrieve from the selected dictionaries does not fully reveal these differences.

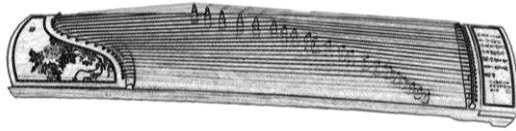


Figure 6: Zheng

2.7 Treatment of 排箫 (paixiao)

Four of the dictionaries define *paixiao* (see Figure 7, click [here](#) to download its sound):

- panpipes; Pan's pipes (ACED)
- pan pipe (NCCED)
- panpipe (TCED)
- panpipes (ABC)

Only four of the five dictionaries include *paixiao* as a lemma, and they all explain it with the translational equivalent *panpipe* (in singular or plural, in one or two words). ACED furnishes both *panpipes* and *Pan's pipes* as equivalents. Collins labels *panpipe* as American English. Merriam-Webster, which represents American English, treats *Pan's pipes* under the lemma *Pan*, the ancient Greek god of woods and shepherds. The definition of *Pan's pipes* is the same as that of *panpipe*:

a wind instrument consisting of a series of short vertical pipes of graduated length bound together with the mouthpieces in an even row — often used in plural

Webster depicts relevant details that cover the form and structure of this musical instrument. It also offers an illustration that shows how to play it. The selected Chinese–English dictionaries should have provided similar items concerning *paixiao* to present a more revealing picture to foreign learners.

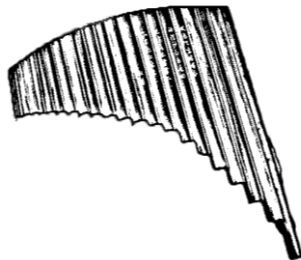


Figure 7: Paixiao

2.8 Treatment of 二胡 (erhu)

Erhu (see Figure 8, click to download [pronunciation](#), [sound](#), or [video](#)) is defined in the five dictionaries as follows:

- *erhu*, a two-stringed bowed instrument with a lower register than *jinghu* (ACED)
- *erhu*, a two-stringed bowed instrument with a lower register than *jinghu* (NCCED)
- *erhu*, a two-stringed bowed instrument with a lower register than *jinghu*; *urh(h)een*; *urhien*; *erhu fiddle* (TCED)
- *erhu*, a two-stringed Chinese fiddle, see also "胡琴" *huqin* (NA)
- *fiddle* (ABC)

ACED and NCCED define it with the pinyin equivalent *erhu* followed by a brief explanation that describes its form (*two-stringed*), basic playing mode (*bowed*), and sound features (*with a lower register than jinghu*).

TCED is distinctive from others in that in addition to the above brief explanation and the pinyin equivalent *erhu*, it furnishes another three equivalents, i.e. *urh(h)een*, *urhien*, and *erhu fiddle*. *Urh(h)een* is included in Webster as a lemma and defined as follows, and *urhien* is marked as the variant form of *urheen* in Webster:

a Chinese fiddle consisting of two strings usually of silk tuned a fifth apart, stretched across a small mallet-shaped hollow block, and fastened at the other end to tuning pegs set in a long stick

In Webster, it is also noted in the entry that *urheen* is the modification of Chinese (Pekingese) *êr-hsien* which literally means "two strings". Despite certain similarities, *Urh(h)een* and *erhu* are two different Chinese musical instruments in terms of its historical origins, features of shapes, playing techniques and sound features (Wang 2008). To adopt *urh(h)een/urhien* as the equivalents of *erhu* without any further discrimination between the two instruments would result in confusion among dictionary users.

ABC offers *fiddle* as an equivalent, whereas NA depicts *erhu* as *Chinese fiddle* and TCED as *erhu fiddle*. It is necessary to explore whether a fiddle can be used as the prototype instrument for foreign learners of Chinese to figure out what *erhu* is. When treated as a musical instrument, Merriam-Webster defines *fiddle* as a violin. Collins explains that *some people call violins fiddles, especially when they are used to play folk music*, and adds a picture of a fiddle. The shapes, forms, and playing modes of *erhu* and *fiddle* are clearly distinct. It does not seem reasonable to define *erhu* as *Chinese fiddle* without providing further data to support the term *Chinese fiddle*. In addition, it is worthwhile to question whether the term *erhu fiddle* is coined appropriately to define *erhu* in the TCED.

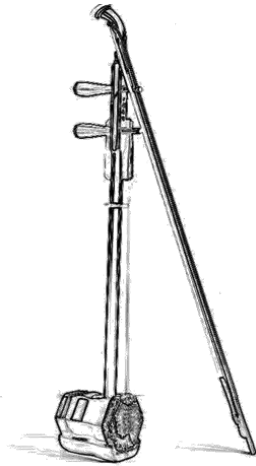


Figure 8: Erhu

2.9 Treatment of 琵琶 (pipa)

Pipa (see Figure 9, click [here](#) to download its sound and [here](#) to download a video clip) is defined as follows:

- *pipa*, a plucked string instrument with a fretted fingerboard (ACED)
- *pipa* [Chinese string musical instrument with a fretted fingerboard plucked historically with a plectrum, but now mainly with the fingers] (NCCED)
- *pipa*, a plucked string instrument with a fretted fingerboard; 4-stringed Chinese lute (TCED)
- a plucked string instrument with a fretted fingerboard *see illustration* (NA)
- lute-like stringed instrument with fretted fingerboard (ABC)

ACED contains a brief explanation of *pipa* that includes its playing mode (*plucked*) and form feature (*fretted fingerboard*). TCED adds another item (*4-stringed Chinese lute*). This piece of information may allow foreign learners of Chinese to get a clearer image of *pipa* since *lute* is similar to *pipa* in many respects. NA is the only dictionary that refers the user to an illustration of *pipa*. ABC's definition of *pipa* is similar to the one in NA. Among the five dictionaries, NCCED distinguishes itself because it describes the change of the playing way (*historically with a plectrum, but now mainly with fingers*).



Figure 9: Pipa

2.10 Treatment of 梆子 (bangzi)

Finally, the five selected dictionaries provide the following definitions of *bangzi* (see Figure 10, click [here](#) to download its sound):

- wooden clappers with bars of unequal length (ACED)
- wooden clappers (NCCED)
- wooden clappers (TCED)
- *bangzi*--clappers of unequal length, made of jujube wood (NA)
- wooden clapper of unequal length (ABC)

Four of the selected dictionaries (ACED, NCCED, TCED, ABC) offer almost the same equivalent (*wooden clappers*) to define *bangzi*. ACED provides further details like *bars of unequal length*. NA also informs about its material (*made of jujube wood*). *Bangzi* is mainly played with other instruments together in a clapper opera which is popular in the north of China. The definitions, however, do not clarify how it is played and when and where it is played.

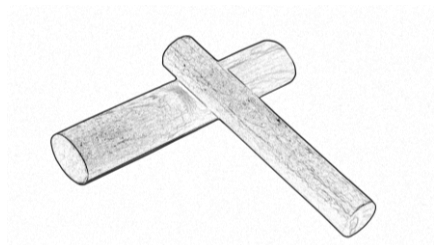


Figure 10: Bangzi

2.11 Summary

The above analysis of the treatment of ten Chinese musical instruments reveals that most of the definitions provided in Chinese–English dictionaries consist of a pinyin or a translational/explanatory equivalent plus a brief explanation. The explanation mainly depicts the instruments' form and general playing mode. Given the unique features of traditional Chinese musical instruments, short definitions are problematic as they do not provide an overall picture. Users interested in the cultural component require additional information, for instance:

- how to play the instruments?
- what are the typical sounds they produce?
- when were they first introduced in Chinese culture?
- how did they change from ancient times to the modern age?
- in which social contexts are they used?
- are there special customs related to their usage?

In brief, as a distinctive group of culture-related objects, Chinese musical instruments should be defined in terms of their form, shapes, and sizes, playing mode and techniques, history, and customs related to their usage. A short explanation of the instruments is far from satisfactory to meet the need of foreign learners interested in Chinese culture and language.

3. Some reflections on cultural items in lexicography

The dictionaries analyzed in the previous Section were all bidirectional, i.e., simultaneously aimed at both native speakers and foreign learners of Chinese. Such a broad target group is problematic, as very different user competencies, needs, and expectations have to be balanced. Chinese native speakers will typically consult Chinese–English dictionaries to solve problems related to text production in English (or translation of Chinese texts into English). Foreign learners of Chinese, by contrast, will typically use this type of dictionary when they look for assistance to either understand an existing Chinese text or formulate a correct one in this language.

Chinese native speakers generally have a broad knowledge of their national culture. If they require additional information on any topic, they will, as a rule, consult monolingual Chinese dictionaries or encyclopedias. On the other hand, when they fall short of English equivalents to refer to cultural objects and phenomena in their homeland, they will get the best assistance in Chinese–English dictionaries. In case no adequate equivalents exist in the English language, they will expect alternative solutions. It could be translational and transliterated equivalents accompanied with small notes informing them

that these words are either imprecise or unfamiliar to most Westerners and, thus, require additional explanation.

The linguistic and cultural competencies and needs of foreign learners of Chinese differ from the ones mentioned above. Most of these learners are adults with a good knowledge of their home culture and relatively little knowledge of Chinese culture that differs considerably from their own. To grasp the cultural component underpinning Chinese texts, they may need specific types of information that native speakers already have. As Zhang and Mi (2020: 74) rightly comment, this calls for much more than "a simple provision of equivalents" and also "involves the translation of cultural elements". Equivalents — like *panpipes* and *wooden clappers* — may be part of the solution, but only if the foreign learners can be expected to know them. This implies that transliterated equivalents — like *guqin*, *konghou*, *dizi*, *zheng*, *erhu*, and *pipa* — should be avoided, as they do not provide any meaningful information to foreign learners with limited knowledge of Chinese culture and musical instruments in general. Data overload of this kind may rather have a disturbing effect on the consultation process, as pointed out by Gouws and Tarp (2017: 397-398).

4. Treatment in the age of media convergence

The concept of a bilingual dictionary is ambiguous. As can be seen, foreign learners and native speakers of Chinese have different, and sometimes also contradictory, expectations for Chinese–English dictionaries. Not even the expectations of foreign learners are always the same. Learners in general never need information in general. Their lexicographical needs are always concrete and invariably related to the challenges and problems, which typically occur in specific contexts during the learning process, notably when engaging in decoding and encoding activities (text reception and text production). In this sense, the relevant types of user needs vary considerably from one activity to another, and so should the response provided by lexicography in the media convergence age.

It seems that all bilingual dictionaries discussed in Section 2 have been designed as one-size-fits-all solutions to different user needs. As such, they represent compromises between more optimal solutions. There is a sort of trade-off here. Some relevant items are included, while others are excluded for the benefit of the overall product. It implies that users consulting these dictionaries in specific types of activity sometimes have to navigate through superfluous data and other times are without enough relevant data. The one-size-fits-all approach is one of the biggest challenges to be overcome for digital lexicography to move forward. There is an urgent need to develop lexicographical products that adjust to the different situations where user needs occur. Lexicography has not only moved from print to digital. It has also developed beyond classic stand-alone dictionaries, whether print or digital, and now inte-

grates products like e-readers, writing assistants, and learning tools; see Fuertes-Olivera et al. (2018).

The growing variety of platforms from which users can access lexicographical data requires a differentiated treatment of these data, among them the cultural ones. The basic principle must be that users are offered the amount of lexicographical data required to meet their needs in each concrete context and that irrelevant data are avoided. This entails, among other things, that some of the data that traditionally characterize the stand-alone dictionary format become redundant in certain contexts. Kwary (2011: 47), for instance, provides an example where users who click on the term *account payable* activate a pop-up window that contains no other lexicographical data than the Indonesian equivalent *utang dagang*. This equivalent is all they need to solve their immediate comprehension problem. Fuertes-Olivera and Tarp (2020: 281) straightforwardly recommend "the elimination of the lemma" in writing assistants because users already know which word(s) they have clicked on to get assistance. Tarp (2022) suggests that e-readers provide differentiated lexicographical support to readers engaged in incidental and intentional learning, respectively. He argues that assistance to the former should be the default solution from which additional data can be accessed to serve the latter. Huang and Tarp (2021: 87) put forward a similar solution when discussing the challenges detected in a learning app for Chinese learners of English:

The main idea is that the pop-up window should only include items that can be justified by the immediate user needs. Thus, it merely consists of a speaker icon, a meaning discriminator followed by two equivalents, and a signifier (>). The central item is the definition (or equivalents) that directly assists understanding of the course text. [...] Finally, it provides a widely used signifier that affords access to the whole article ...

Lexicographers should provide maximum support to the big variety of information tools existing today, including digital dictionaries. This calls for their focus to move from the dictionary as such to the lexicographical database, where the relevant data are stored and from which they can be uploaded to the different tools in carefully metered doses. Surprisingly, most scholars who make concrete proposals for the lexicographical treatment of cultural and other items focus exclusively on their presentation in dictionaries; see, for instance, Tseng (2003), Sánchez (2010), Klosa (2015), Xue (2017), and Huang (2020). This approach will not be able to overcome the problematic one-size-fits-all tradition. We have only found one contribution on culture (Kwary and Miller 2013) that focuses on the underlying database and discusses its design. But even this database only aims at serving a specific dictionary, albeit a more advanced one with hypermedia integrated. It is, however, questionable whether it can provide optimal support to other tools. Fuertes-Olivera and Tarp (2020: 277), for instance, argue that "future lexicographers should consider the inclusion of both short (one-line) and longer definitions into their databases" if they aim at

sustaining writing assistants. This calls for meticulous preparatory work. Lexicographers should have a clear idea of the specific data types required by the different tools. Otherwise, it may have negative retrospective consequences for their database. The following Section will, thus, discuss some of the requirements for the inclusion of cultural items into bilingual lexicographical databases that serve foreign learners of Chinese.

5. Cultural items required in lexicographical databases

A bilingual database serving foreign learners who make look-ups from various platforms will generally have dozens of data fields attached to each lemma. In this regard, it is more diverse than databases sustaining the one-size-fits-all solutions criticized above. We will here concentrate on the main fields reflecting the cultural dimension, that is, the ones concerned with the treatment of equivalents, explanations, and non-textual items.

Before discussing each of these data categories, it is necessary to clarify in which contexts the lexicographical data stored in a Chinese–English database could assist foreign learners. The most typical situation would be when learners do not understand — or have doubts about — words and expressions appearing in Chinese texts, that is, text reception. But also assistance to text production could be relevant when learners already know a Chinese word and just want to confirm its precise meaning before using it. In addition to these two communicative situations, learners may need or wish to specifically explore the cultural dimension related to the learning process, that is, a so-called cognitive situation; see Tarp (2008). Finally, and less important for language learning in the narrow sense of the word, is L2–L1 translation when learners for some reason have to translate Chinese words or texts into their own language.

All this suggests that a well-designed Chinese–English database for foreign learners should have four main functions, each of which requires specific data categories as well as differentiated treatment of these categories, something further modified by the respective platforms from where the database is accessed. In the following sections, we will take this into consideration when discussing equivalents, explanations, and non-textual items.

5.1 Equivalents

Although there is no complete agreement on how to do it, lexicographers usually distinguish between various types of equivalents with different characteristics. Adamska-Salaciak (2011: 4), for instance, lists four types (cognitive, translational, explanatory, and functional), of which the first three are relevant for the current study. Surprisingly, the discussion of equivalence is more often than not influenced by reflections and principles imported from translation

science despite being irrelevant to most functions in bilingual lexicographical products, unless these are designed as one-size-fits-all solutions or exclusively aim at assisting translation. When it is a case of understanding an L2 text or confirming the meaning of an L2 word to be used, it has, for example, no relevance whether the equivalent provided is insertable in an L1 text, as required by many lexicographers. Existing lexicographical literature undoubtedly represents an important source of inspiration. But even the most well-established and deep-rooted principles and solutions cannot uncritically be taken over and applied when designing a database adjusted to the digital media. They have to adapt to the new environment.

From this perspective, the *cognitive equivalent* in a Chinese–English learner's dictionary can be defined as a word or expression that already exists in the English language and faithfully represents the meaning of a Chinese word or expression. This degree of equivalence is, by definition, relatively rare for cultural terms. The five Chinese–English dictionaries analyzed in Section 2, for instance, only provide truly cognitive equivalents (*panpipes* and *wooden clappers*) to two out of ten Chinese musical instruments. Contrary to the other types of equivalents, the cognitive equivalent serves all three communicative functions of a Chinese–English lexicographical database, that is, to explain a Chinese item, to confirm the meaning of a Chinese item, and to translate a Chinese item into English. Because of these properties, the cognitive equivalent should have its own field in the database.

In this essential aspect, the proposed database differs qualitatively from almost all traditional ones where different types of equivalents are assigned to one and the same data field. If the database only included one equivalent field, it would be technically impossible for the tool to distinguish between the equivalents and upload the ones that correspond to the respective user situations and platforms. This would considerably reduce the quality of the database. When no cognitive equivalent is available, the field reserved for this specific data type should, therefore, be left empty and other fields prepared for alternative solutions.

One of these complementary fields should be reserved for an *explanatory equivalent* that can briefly explain the meaning of the L2 item and assist L2-text reception and L2-text production as described above. The explanatory equivalent usually takes the form of a paraphrase like *an ancient egg-shaped, holed wind instrument* (see Section 2.5) and *a 21- or 25-stringed plucked instrument in some ways similar to the zither* (see Section 2.6). These explanatory equivalents represent to an acceptable degree the meaning of two Chinese musical instruments (*xun* and *zheng*), although an additional explanation is required to give the learner the full cultural picture and respond to the six questions listed at the end of Section 2.

However, some explanatory items — like Chinese string musical instrument with a fretted fingerboard plucked historically with a plectrum, but now mainly with the fingers (see Section 2.9) — are too long to comply with Fuertes-

Olivera and Tarp's (2020: 277) request that future databases include "both short (one-line) and longer definitions". The two authors propose that the former also serve as meaning discriminators in writing assistants; for instance, when *Write Assistant* suggests various L2 equivalents or *Grammarly* offers diverse synonyms or quasi-synonyms. In such cases, too long paraphrases (or explanatory items) may result in data overload and hamper the consultation process, as argued by the two authors. A shorter explanatory equivalent — e.g., Chinese string musical instrument with a fretted fingerboard — should, therefore, be prepared and further details — like plucked historically with a plectrum, but now mainly with the fingers — included in a supplementary explanation.

Until now, we have not discussed the *translational equivalent*. Its lack of explanatory power is well-known, although two or more cumulative equivalents of this type may give the user a certain idea of the meaning of an L2 term. It is, nonetheless, questionable whether this solution works for cultural objects. The two cases where the Chinese–English dictionaries have used cumulative equivalents (e.g., *chime bells*; *serial bells*; *carillon* in Section 2.1 and *flute*; *bamboo flute* in Section 2.4) to depict the meaning of Chinese musical instruments are not convincing. In this respect, we agree with Zgusta (1984: 148), who prefers an explanation "in cases where the mere translational equivalent, exact as it may be, might remain only poorly intelligible". There is no reason to use translational equivalents to explain meaning when the database already offers cognitive and/or explanatory equivalents. They should instead be reserved for situations where learners need assistance to translate Chinese words or texts into English and where no cognitive equivalents exist. It requires a separate data field (combined with a usage note field) from which this type of equivalents can be uploaded when — and only when — needed for translation purposes.

In Section 2, we saw how the selected dictionaries provided *transliterated equivalents* to six of the ten musical instruments treated. This equivalent type has the advantage that it presents the exact meaning of the Chinese cultural item and the disadvantage that it is unknown to most foreign learners. In spite of being a quasi-cognitive equivalent, we do not recommend it be presented as default in "normal" Chinese–English learners' dictionaries. But because of its semantic properties, it could be a perfect choice in translations if accompanied by a usage note the first time it appears in a text. It should, therefore, also have its separate field (combined with a usage note field) in the lexicographical database.

Whether a translational or a transliterated equivalent should be used in connection with translation depends on various factors. Each of them has its advantages and disadvantages. And there could also be subjective criteria involved. In any case, this decision could be left to the designers of the specific tools expected to use the lexicographical data. It is then the task of lexicographers and designers of databases to make provisions for the different usages.

5.2 Explanations and other items

Equivalents are necessary but not sufficient to deal with cultural terms in bilingual learner's dictionaries. Foreign learners need more detailed information to get a good understanding of traditional Chinese musical instruments and their usage. The role of explanations is to enhance the cultural information provided by the cognitive and explanatory equivalents. In an article on music dictionaries, Bergenholtz and Bergenholtz (2007: 411) propose the preparation of three types of explanations: a short default one, a longer one "with historical and other kinds of details", and an optional one "with examples from music history". Their proposal includes many relevant things for the topic discussed here. But the two authors have a music dictionary for a particular user group in mind, namely native laypeople and semi-experts especially interested in music. Our proposal focuses on a different user group with different lexicographical needs and, therefore, only envisages two types of explanations, a short one and a longer, supplementary one. In our proposal, the short one should mainly give a brief description of the musical instrument and the way it is played. The long one should then provide further cultural details like its history, the social contexts where it is used, customs related to its usage, etc. When relevant, the definitions should be supported by non-textual items like images (illustrations and photos), video clips, and sound files. Each of these items can show, illustrate, or express particular aspects and phenomena where written text falls short. They should, thus, have their own fields in the database and either be presented as default or activated by clicking on the respective icons.

As can be observed, there is a sort of continuum from the cognitive equivalent over the explanatory equivalent and the short explanation to the long explanation. There are no sharp dividing lines between the four categories. It is, nonetheless, essential to establish such dividing lines in order to prepare the lexicographical database for a variety of purposes. In a "normal" learner's dictionary, the provision of a cognitive or explanatory equivalent together with a short explanation could be the default solution, from where interested users can access the long explanation with a simple click. In a learner's dictionary focussing on the cultural aspect, the long explanation could be part of the default solution. In an e-reader, the cognitive or the explanatory equivalent could be the default solution that stimulates incidental learning, whereas readers interested in intentional learning could get the required information by clicking through to the short and long explanations. And so on. This kind of differentiated presentation of lexicographical data would be impossible without a conceptual distinction between the various equivalent and explanation types and their subsequent storing in separate fields in the database.

Figure 11 shows the relevant database fields filled in with cultural data assigned to the Chinese lemma 二胡 (erhu) (there is no cognitive equivalent in this case). Although other data categories like collocations may also have some cultural relevance (see Huang 2020: 116-120), it only contains the fields reserved for the ones discussed above.

DATA FIELDS	LEXICOGRAPHICAL CONTENT
LEMMA	
CHINESE CHARACTER	二胡
PINYIN	èrhú
PRONUNCIATION	Audio
EQUIVALENTS	
COGNITIVE	
EXPLANATORY	Chinese bowed musical instrument
TRANSLATIONAL	Chinese fiddle
USAGE NOTE	<i>Chinese fiddle</i> does not represent the full meaning of 二胡. It is recommended that a clarifying note be added the first time it appears in a text.
TRANSLITERATED	<i>erhu</i>
USAGE NOTE	<i>Erhu</i> faithfully represents the meaning of 二胡 but may be unknown to many native English speakers. It is recommended that a note explaining its meaning be added the first time it appears in a text.
EXPLANATIONS	
SHORT	traditional Chinese two-stringed bowed instrument with a low but mellow sound which can be played as a solo instrument as well as in orchestras
LONG	<p><i>Erhu</i> has evolved from the ancient musical instrument <i>xiqin</i> in Tang Dynasty (618-907). This instrument, which is believed to originate from the <i>Xi</i> people in northeast China, became popular among different areas of ancient China. As a result, a variety of similar bowed instruments called the “<i>Huqin</i> instrument family” saw the light. Today, <i>erhu</i> is the most popular member of this instrument family.</p> <p><i>Erhu</i> consists of a long vertical neck at the top of which there are two big tuning pegs. At the bottom, there is a small sound box covered with python skin on the front end. Two strings stretch from the pegs to the base, and a small loop of string is placed around the neck. <i>Erhu</i> is often described as a Chinese fiddle, but its sound differs somewhat from that of a fiddle. In addition, <i>erhu</i> has only two strings, while the fiddle has four. People usually play <i>erhu</i> sitting down, with the soundbox placed on the top of the left thigh and the instrument neck held vertically.</p> <p><i>Erhu</i> can be used in both traditional and contemporary music activities. It can be played alone or accompanied by other traditional musical instruments. Today, <i>erhu</i> is one of the main instruments in regional music ensembles, Chinese opera ensembles, and large orchestra. It also appears as soundtrack in movies and TV series.</p>
NON-TEXTUAL ITEMS	
IMAGE	Image
VIDEO	Video clip
SOUND	Sound file

Figure 11: Suggested data fields in the lexicographical database

5.3 Differentiated presentation

This section provides five examples of how the items stored in the above database can be selectively employed in different tools and adapt to the foreseen user needs. The first three examples are from a hypothetical Chinese–English dictionary for foreign learners. Figure 12 shows a default article from such a dictionary. The selection and presentation of data correspond to consultations where foreign learners primarily look for assistance to understand a Chinese text or, alternatively, to confirm the meaning of Chinese words they want to use in L2-text production. The article contains the pertinent cultural data without slipping into data overload: an explanatory equivalent, a short explanation, and an illustration. Learners who require additional information can then click on "MORE" to display the supplementary explanation (see Figure 13) or on one of the two icons to either listen to the music or see the video. Finally, users who need to translate 二胡 into English can get proper assistance in a pop-up window by clicking on "OPEN" in the bottom bar (see Figure 14).

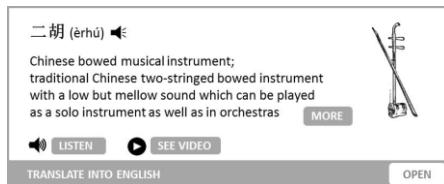


Figure 12: Default article in Chinese–English dictionary for foreign learners

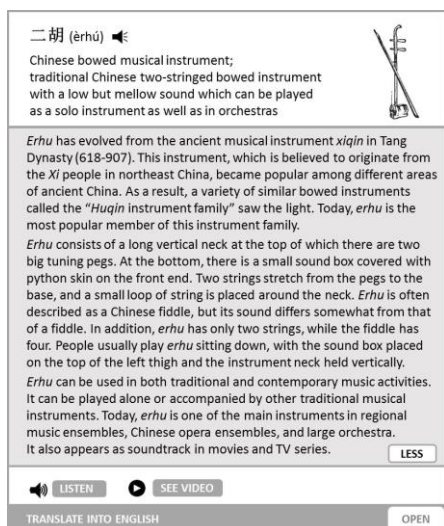


Figure 13: Extended cultural explanation

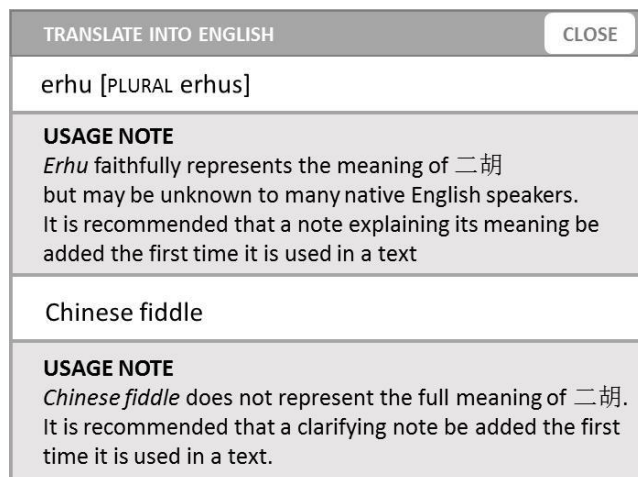


Figure 14: Pop-up window providing assistance to L2–L1 translation

Foreign learners increasingly read Chinese texts on digital devices like e-readers and tablets. If carefully integrated into these devices, the cultural items stored in the Chinese–English database may assist both incidental and intentional learning during the reading process. From a lexicographical perspective, the former presupposes immediate, contextualized, and unobtrusive assistance with a minimum of lexicographical data, whereas the latter requires easy access to additional data; see Tarp (2022). Figure 15 shows a proposed pop-up window that visualizes when readers, who do not understand 二胡, click on these characters. Its limited and instantly displayed content — an explanatory equivalent and a signifier (▼) — facilitates text comprehension without disturbing the reading flow. In this way, it also stimulates incidental learning. If some learners, for one reason or another, decide to interrupt the reading flow to get more detailed cultural information about 二胡, they can easily access additional data by clicking on the signifier. Figure 16 shows the result where motivated users can scroll down to find all relevant details, thus moving on to intentional learning with a clear cultural component.



Figure 15: Default pop-up window in digital device used to read Chinese texts

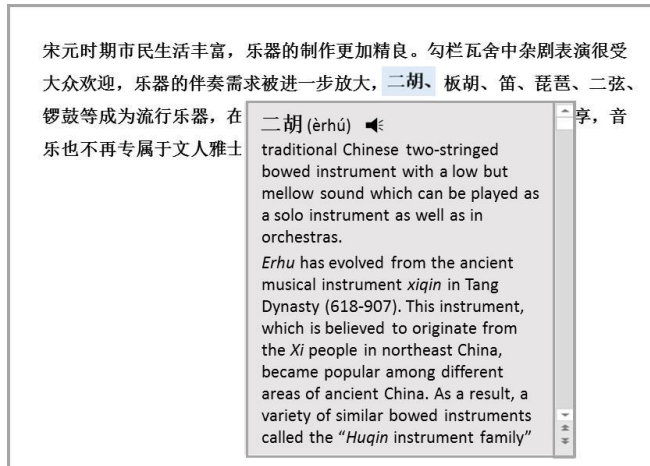


Figure 16: Supplementary cultural details in digital device used to read

6. Conclusions

In the introduction, we noted that lexicography needs new solutions to both new and old challenges. The critical analysis of five Chinese–English dictionaries and their treatment of ten traditional musical instruments exposed some of these challenges. The one-size-fits-all approach, together with an undifferentiated interpretation of user needs, was targeted as a major culprit. Too much focus on translation is also problematic. Foreign users' needs invariably relate to the situations and platforms where they occur, and the needs occurring in connection with translation are only a minority. Much more relevant for learners is the assistance to decoding and encoding Chinese texts and their immersions in contemporary and traditional Chinese culture. We have argued that this requires a new conceptual approach to equivalents and explanations with consequences for their storing and presentation to the end-users. From this perspective, we have shown that the database, if well-designed, is the natural centre of a whole web of lexicographical products that allows a differentiated and user-adapted treatment of cultural items, among others.

The preparation of lexicographical databases does not only require adequate linguistic knowledge but also specialized knowledge of culture. It is not always a happy marriage. Bergenholtz and Bergenholtz (2007) analyzed the treatment of musical instruments in various dictionaries and found that the definitions are often lacking or directly erroneous. They recommend that this part of the lexicographical work is conducted by, or in collaboration with, subject-field experts. In our case, we have based the explanations of *erhu* on Wikipedia and confirmed them in authoritative sources like Feng (2006) and Wang (2008). Interdisciplinarity, in its different expressions, is vital for contemporary lexicog-

raphy to create quality products that adapt to foreign learners' linguistic and cultural needs. The traditional musical instruments discussed in this paper require a particular treatment that may differ from that of other cultural objects and phenomena. But many of the above reflections can be generalized and relevant to them in one way or another.

Finally, as an expression of media convergence, the paper offers access to video clips and sound files that may enhance the cultural dimension as well as the readers' benefit from the text. But the quality of these items is not always as desired. Copyright and too high costs have, in some cases, led to the choice of lower quality items. This is yet another challenge to be addressed by lexicography in the years to come.

Acknowledgments

Thanks are due to the Key Program of National Social Science Fund of China (20AYY012) and Guangdong Well-known Overseas Professorship Program for supporting and funding this work.

Special thanks are due to the Center for Lexicographical Studies at Guangdong University of Foreign Studies, China, for appointing Sven Tarp as Yunshan Chair Professor, thus making the collaboration between the two authors of this article possible.

Special thanks are also due to Li Lexuan for preparing the illustrations of the ten traditional Chinese musical instruments treated in this article.

References

A. Dictionaries

ABC Chinese–English Comprehensive Dictionary. 2003. Shanghai: Chinese Grand Dictionary Press.

A Chinese–English Dictionary. 2010. 3rd edition. Beijing: Foreign Language Teaching and Research Press.

A New Century Chinese–English Dictionary. 2004. Beijing: Foreign Language Teaching and Research Press.

Collins English Dictionary Online. <https://www.collinsdictionary.com/dictionary/english/>. Accessed 10 January, 2022.

Merriam-Webster Dictionary Online. <https://www.merriam-webster.com/dictionary/>. Accessed 10 January, 2022.

New Age Chinese–English Dictionary. 2002. Beijing: The Commercial Press.

The Chinese–English Dictionary. 2010. 3rd Edition. Shanghai: Shanghai Translation Publishing House.

B. Writing assistants

Grammarly. <https://www.grammarly.com/>.

Write Assistant. <https://www.writeassistant.com/en/>.

C. Other literature

- Adamska-Salaciak, A.** 2011. Between *designer drugs* and *afterburners*: A Lexicographic-Semantic Study of Equivalence. *Lexikos* 21: 1-22.
- Bergenholtz, H. and I. Bergenholtz.** 2007. On the Timelessness of Music Dictionaries. *Lexikos* 17: 407-415.
- Bergenholtz, H. and S. Nielsen.** 2013. The Treatment of Culture-bound Items in Dictionaries. Gouws, R., 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*: 469-481. Berlin/Boston: De Gruyter.
- Feng, X.** 2006. *The Echo of Traditional Musical Instruments*. Shanghai: Shanghai Conservatory of Music Press.
- Fuertes-Olivera, P.A.** 2013. Electronic Encyclopedias. Gouws, R., 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*: 1069-1081. Berlin/Boston: De Gruyter.
- Fuertes-Olivera, P.A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.
- 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.** 2020. Metalexigraphy, Dictionaries and Culture. *Lexicographica* 36: 3-9.
- Gouws R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Hallsteinsdottir, E.** 2015. The Notion of Cross-linguistic and Cross-cultural Equivalence in the Field of Phraseology. *International Journal of Lexicography* 28(3): 385-390.
- Huang, F.** 2020. The Treatment of Cultural Words in Chinese Foreign Dictionaries — A Case Study on "粽子". *Lexicographica* 36: 109-122.
- Huang, F. and S. Tarp.** 2021. Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvement. *Lexikos* 31: 68-92.
- Kang, S.** 2020. Thoughts on Chinese Learner's Dictionary of Media Convergence: A Case Study of the Contemporary Chinese Learner's Dictionary. *Journal of Ludong University* 37(2): 1-7.
- Klosa, A.** 2015. Illustrations in Dictionaries. Encyclopaedic and Cultural Information in Dictionaries. Durkin, P. (Ed.). 2015. *The Oxford Handbook of Lexicography*: 515-531. Oxford: Oxford University Press.
- Kwary, D.A.** 2011. Adaptive Hypermedia and User-oriented Data for Online Dictionaries: A Case Study on an English Dictionary of an English Dictionary of Finance for Indonesian Students. *International Journal of Lexicography* 25(1): 30-49.
- Kwary, D.A. and J. Miller.** 2013. A Model for an Online Australian English Cultural Dictionary Database. *Terminology* 19(2): 258-276.
- Li, Q. and Y. Xia.** 2020. Equivalent Varieties for Culture-specific Words in Chinese-English Dictionaries. *Lexicographica* 36: 123-138.
- Ma, S.** 2018. On the Translation of Culture-specific Words in Chinese-English Dictionaries Based on Frame Semantics: A Case Study of Words Related to *keju*. *Lexicographical Studies* 2: 57-67.

- Miller, J., D.A. Kwary and A.W. Setiawan.** 2017. Koalas, Kiwis and Kangaroos: The Challenges of Creating an Online Australian Cultural Dictionary for Learners of English as an Additional Language. *Lexikos* 27: 310-345.
- Nomdedeu Rull, A.** 2020. How to Select and Present Cultural Data: A Challenge to Lexicography. *Lexicographica* 36: 39-57.
- Sánchez, A.** 2010. The Treatment of Cultural and/or Encyclopaedic Items in Specialised Dictionaries for Learners. Fuertes-Olivera, P.A. (Ed.). 2010. *Specialised Dictionaries for Learners*: 107-129. Berlin/New York: De Gruyter.
- Stark, M.** 1999. *Encyclopedic Learners' Dictionaries. A Study of their Design Features from the User Perspective*. Tübingen: Max Niemeyer.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.
- Tarp, S.** 2022. A Lexicographical Perspective to Intentional and Incidental Learning. (To be published.)
- Tseng, T.H.T.** 2003. The Cultural Component in an Ideal Monolingual English Learner's Dictionary. Ming, L., H. Hue-Feng and Z.J. Zhuang. (Eds.). 2003. *Aspects of English Language Teaching and Research*: 215-226. Suzhou: Suzhou University Press.
- Wang, P.** 2008. *A Comparative Study between Fujian Nanyin Erxian and Chinese Erhu*. Unpublished M.A. Thesis. Fuzhou: Fujian Normal University.
- Xue, M.** 2017. Representing the Cultural Dimension of Meaning in Learner's Dictionaries — From the Perspective of Chinese EFL Learners in L2 Reception. *Lexikos* 27: 578-596.
- Xue, M. and S. Tarp.** 2018. Towards Chinese Learner's Dictionaries for Foreigners Living in China: Some Problems Related to Lemma Selection. *Lexikos* 28: 384-404.
- Zhang, Y.** 2019. On the Innovation of Dictionary Compilation and Publication in the Context of Media Convergence. *Research on Language Strategies* 6: 79-89.
- Zhang, Y. and H. Mi.** 2020. Enhancing the Role of Culture-specific Constructs in Chinese (–English) Dictionaries for International Learners. *Lexicographica* 36: 59-87.
- Zgusta, L.** 1984. Translational Equivalence and the Bilingual Dictionary. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 147-154. Tübingen: Max Niemeyer.
- Zgusta, L.** 1989. Idle Thoughts of an Idle Fellow; or, Vaticinations on the Learners' Dictionary. Tickoo, M.L. (Ed.). 1989. *Learner's Dictionaries: The State of the Art*: 1-9. Singapore: SEAMEO Regional Language Centre.

The Mental Lexicon in Lexicography: The *Diccionarios Valladolid-UVa*

Pedro A. Fuertes-Olivera, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa; International Centre for Lexicography, University of Valladolid, Spain; and Centre of Excellence in Language Technology, Ordbogen A/S, Odense, Denmark (pedro@emp.uva.es)*

Abstract: This article analyzes the possibility of making dictionaries that take into consideration the mental lexicon, i.e. words do not work in isolation; instead, they are dynamic constructs that are activated, stored, processed and retrieved gradually. For that, it proposes several general lexicographical and methodological ideas and illustrates them referring to their implementation in the *Diccionarios Valladolid-UVa*: (a) dictionary types are a thing of the past; (b) words are not only holistic products but also processes that are always on the move; consequently their descriptions in dictionaries must be as complete and precise as possible; (c) dictionaries must be equipped with dynamic search system, e.g. systems for allowing human and machine-users search and retrieve *a la carte*, e.g. in a speaking situation; (d) there must be a huge number of words and other data types for describing each meaning and usage of each lemma, thus favoring the creation of patterns and the learning process associated with Artificial Intelligence (AI); (e) designing and making online dictionaries is a cooperative process in which lexicographers and several types of experts must participate; (f) the main task of lexicographers is the preparation of lexicographical data, which can be used in many different forms, formats and usages, being the making of dictionaries one of them.

Keywords: E-LEXICOGRAPHY, ARTIFICIAL INTELLIGENCE, MENTAL LEXICON, WORDS AS PRODUCTS, WORDS AS PROCESSES, LEXICOGRAPHICAL METHODOLOGY

Opsomming: Die kognitiewe leksikon in die leksikografie: Die *Diccionarios Valladolid-UVa*. Hierdie artikel analiseer die moontlikheid om woordeboeke wat die kognitiewe leksikon in ag neem, te skep, m.a.w. woorde funksioneer nie in isolasie nie; intendeel, hulle is dinamiese konsepte wat geleidelik geaktiveer, gestoor, geprosesseer en onttrek word. Met hierdie doel in gedagte word verskeie algemene leksikografiese en metodologiese idees aangebied en geïllustreer deur na hul toepassing in die *Diccionarios Valladolid-UVa* te verwys: (a) woordeboek-tipes behoort tot die verlede; (b) woorde is nie net holistiese produkte nie, maar ook prosesse wat gedurig beweeg; gevolglik moet hul beskrywings in woordeboeke so volledig en presies moontlik wees; (c) woordeboeke moet toegerus word met dinamiese soekstelsels, bv. stelsels wat soektogte en onttrekkings deur menslike en masjien-gebruikers *a la carte* toelaat soos in 'n gespreksituasie; (d) daar moet 'n groot aantal woorde en ander datatipes vir die beskrywing van elke betekenis en gebruik van 'n lemma wees om sodoende die skep van patrone en die aanleerproses wat met Kunstmatige Intelligensie (KI) geassosieer word, te steun; (e) die ontwerp en skep van aanlyn woordeboeke is 'n koöperatiewe proses waaraan leksikograwe en verskeie soorte kundiges moet deelneem;

(f) die hooftaak van leksikograwe is die voorbereiding van leksikografiese data, wat in baie verskillende vorms, formate en toepassings gebruik kan word, met die skep van woordeboeke as een daarvan.

Sleutelwoorde: E-LEKSIKOGRAFIE, KUNSMATIGE INTELLIGENSIE, KOGNITIEWE LEKSIKON, WOORDE AS PRODUKTE, WOORDE AS PROSESSE, LEKSIKOGRAFIESE METODOLOGIE

0. Introduction

This paper revolves around two broad issues. The first one considers the lexicon from a cognitive point of view. Such an approach has shown its centrality, e.g. in Natural Language Processing (NLP), and its dynamicity, i.e. its functioning within dynamic networks that illustrate that words do not work in isolation; instead, their working somehow reproduces the modularity, parallelism, dynamicity and high connectivity of the human brain (Section 1, below). This means that words are, on the one hand, products, i.e. holistic entities, and, on the other hand, processes, i.e. entities on the move. As products, they are codified and can be stored in knowledge resources, e.g. dictionaries. As processes, they are always changing, modifying and/or adapting their meanings, forms and usages to different environments.

Research, e.g. Indefrey and Levelt (2004), has maintained that starting from meanings, the speaker initially activates *lemmata*, i.e. abstract lexical forms devoid of lexical information, and then *phonological forms*, i.e. sounds, syllable, phonemes. Such a process connects the first issue with the second one, which considers whether we can devise an ecosystem to support word finding, i.e. word access at the moment of speaking and writing, e.g. by relying on Computer Science to overcome the challenges this transformation poses for traditional lexicography (Section 2, below).

This paper, then, assumes that we can design and make dictionaries that facilitate the brain processes used when humans are employing languages, the so-called "mental-lexicon" (Aitchison 2003; see Section 1, below), and that Computer Science, especially Artificial Intelligence (AI) may help us in the above-mentioned task. It illustrates them with ideas and data from the design and making of the *Diccionarios Valladolid-UVa* (Section 3). This lexicographical project has been discussed in lexicographical environments that are basically related with the user as reader (Tarp and Fuertes-Olivera 2016; Fuertes-Olivera 2019; Fuertes-Olivera and Tarp 2020; Fuertes-Olivera, Tarp and Sepstrup 2018; Fuertes-Olivera and Esandi-Baztan 2020). In this paper, I will focus on lexicographical decisions that aim at suiting the speaker/writer when they need products that can be easily converted into processes. This user can be a human or a machine, as I believe that the future of lexicography is in the preparation of lexicographical data that will be used in a myriad of usages, forms, formats and purposes (Fuertes-Olivera and Tarp 2020).

1. The mental lexicon

Jackendoff (2002) claims that the mental lexicon is a kind of dictionary that, firstly, contains individual speakers' word representations, which are described in terms of their meanings, pronunciations, formal and functional characteristics, and so on; and, secondly, deals with how those words are activated, stored, processed, and retrieved by each speaker. This view assumes that the mental lexicon or *brain dictionary* is an individual's construct that is constantly changing and growing as new words are learned and old words are forgotten. As such, users of the brain dictionary activate their search gradually, i.e. depending on individual user needs and situations.

The above view opposes the approach mostly present in *traditional dictionaries*, i.e. repositories of words that have a physical or digital form. Traditional dictionaries also contain descriptions of the meaning, form, pronunciation, syntactic characteristics and so on of words but, to the best of my knowledge, most of them take for granted that their words are holistic entities and as such can only be activated, stored, processed and retrieved holistically, i.e. in a way that seems to be different from the modular, dynamic, parallel and highly connective system used by our brain (Emmorey and Fromkin 1988).

Research (Aitchison 2003) has also shown that the brain dictionary is not organized alphabetically. Instead, its organization seems to be connected with the existence of neural circuits. These are subjected to processes such as *spreading activation*, a concept proposed in semantic network theory (Forster and Chambers 1973; Marslen-Wilson 1987) to refer to a "hypothetical mental process that takes place when one of the nodes in the semantic network is activated" (Traxler 2012: 84).

Within the tenets of semantic network theory, three main approaches to the activation of the brain dictionary have been proposed: (a) priming; (b) neighborhood effects; and (c) frequency effects. Priming basically defends the activation of related words once a particular word is searched for; for instance, if we have "euro", priming will activate "European Union" (Hoey 2005). Neighborhood effects (Andrews 1989) refer to the activation of all similar "neighbors" of a target word, i.e. items that are highly confusable with the target word, e.g. for the word "game", its neighbors will be "came, dame, fame, lame, name, same, tame, gale, gape, gate, and gave," (*Wikipedia*, The Mental Lexicon). Frequency effects suggest that words that are frequent in an individual's language are recognized faster than words that are infrequent (Forster and Chambers 1973).

On the other hand, traditional dictionaries tend to list all their words alphabetically, which suits the reader but not the writer or speaker. Furthermore, dictionaries tend to offer their users complete descriptions of their words and do not usually include any system for helping their potential users to search only for what they need in a particular usage situation. In addition, they typically give full lexicographical status to single-word lexemes, i.e. they only offer a complete lexicographical description of words such as "bank" but not of

"extended units of meaning", i.e. phraseological expressions such as "cry all the way to the bank", although recent research has shown that they are crucial in language processing and must be, therefore, lemmatized (Rundell 2018; Fuertes-Olivera 2019).

In sum, the mental lexicon starts from concepts, i.e. meanings, whereas the traditional dictionary approach adopts a topological view, which starts from forms and basically corresponds to off-line processing deliberately searching in a lexical resource, e.g. a dictionary. The next section offers a brief description of some attempts aiming at reconciling the working of our mental lexicon with the design and making of novel reference resources that might be more in line with how our brain works.

2. Computer science and lexicography

In the field of lexicography, Computer Science has mostly focused on the design of Natural Language Processing (NLP) tools which can facilitate information extraction, information retrieval, named-entity recognition, parsing, chunking, part-of-speech tagging, word sense disambiguation, and so on. The tools help in "the representation of linguistically expressible knowledge in language understanding, the use of knowledge for several sorts of commonsense reasoning, and knowledge accumulation" (Espinosa-Anke 2017: 4).

Projects such as *FrameNet* (Ruppenhofer et al. 2018), typically identified as *knowledge resources*, are examples of the joint work of computer scientists and lexicographers. Such resources are useful "because they can store meanings of words and phrases, relations of any kind (e.g. syntactic, syntagmatic, semantic or ontologic) holding among them, and also descriptions about entities or commonsense facts" (Espinosa-Anke 2017: 3).

Knowledge resources are classified into three categories: *structured knowledge resources*, e.g. dictionaries or knowledge bases, *unstructured knowledge resources*, e.g. statistical models derived from text corpora, or *semi-structured knowledge resources*, e.g. *Wikipedia* (Hovy et al. 2013). In what follow, I will focus on structured knowledge resources, i.e. manually-crafted resources, because they are considered to represent knowledge at the highest level of quality. Hence, it can be hypothesized that if these are the highest quality resources, they will offer the best potentialities for overcoming some of the challenges NLP is facing, in particular those concerned with meaning extraction and elimination of contextual ambiguity.

Research, e.g. Clark et al. (2012); Hovy et al. (2013); Espinosa-Anke (2017), classifies structured knowledge resources into three types: (a) lexicographical resources; (b) lexical databases and thesauri; (c) knowledge bases. Lexicographical resources such as dictionaries are typically human readable. They consist of a list of words and their associated *senses* or *meanings*, usually accompanied by an array of lexicographical data, i.e. data that have been prepared by a human lexicographer with the aim of offering a (complete) descrip-

tion of the meaning and usage of the *lemmas*, i.e. the entry words of dictionaries. For the purpose of this article, it is adequate to indicate that the data included in dictionaries are typically accessed and retrieved holistically, that such a characteristic does not suit the working of our brain and, consequently, this lexicographical method must be changed assuming that lexicographers aim at improving these knowledge resources by using NLP methodology in lexicography.

Lexical databases and thesauri, e.g. *Roget's Thesaurus* (see Kirkpatrick 1987) and *WordNet* (Miller 1995; Fellbaum 1998) represent *senses* by grouping them into sets of (cognitive) semantic relations, called *synsets* in *WordNet*). They employ an onomasiological ordering, i.e. the lemmas are arranged by their meanings and consequently users access them through their semantics. This means that word forms such as "bank", which can have several semantic relations, are found in several sections of the lexical databases and thesauri. For instance, the *Roget's Thesaurus* (Kirkpatrick, 1987) uses a topical distribution, e.g. "bank" is connected semantically to *height, support, obliquity, edge, laterality, land, store, lending* and *treasury*. Research (Zock and Bilac 2004; Zock and Schwab 2008) has shown that such resources, especially online lexical databases and thesauri, are especially useful for production purposes. In this paper, I will also add some comments to the above reflections and will hypothesize that the basics of these resources, i.e. semantic relations such as synonymy, antonymy, hyponymy and meronymy are especially relevant for meaning extraction and disambiguation, thus increasing the potential use of NLP for lexicographical purposes. For instance, they can help the formation of network hubs in the human brains assuming that lexicographers can create some kind of pattern between the meaning and the semantic relations of a particular lemma (see 4, below).

Ré et al. (2014: 1) define a knowledge base (KB) as "a relational database together with inference rules, with information extracted from documents and structured sources". Espinosa-Anke (2017: 8) adds that in general, "we expect KBs to be graph-like data structures where each node represent an entity or concept (e.g. *Nintendo* or *hope*), and where edges between nodes may express *WordNet*-like semantic relations, but also ontologic relations such as *is-based-in* or *is-CEO-in*". They are especially useful for representing knowledge in a network form, especially in terminological knowledge bases (TKBs), which are in the forefront of research in this field. This can be achieved by giving definitions a reticular form, which consists of two stages, "first, to de-contextualise the terms and, second, to retain only the contexts that can be used to code knowledge in a network form" (Condamines 2018: 338).

Condamines (2018: 343) also looks at the future of TKBs and makes three observations, two of which are relevant for this paper. Firstly, TKBs are being substituted by *ontologies*, i.e. explicit specifications of conceptualizations (Gruber 1993; Roussey et al. 2018). Ontologies are being constructed by applying machine learning methods on very large corpora, usually the entire Web, "in order to spot new patterns and new triplets". She adds that with machine-

learning methods, the main aim "is not to build a precise representation of the knowledge but, rather, to detect enough regularities to assume that some couples of terms have a constant and relevant relationship. In these cases, the most important application is to improve information retrieval". Secondly, the use of natural language processing tools, which are being used more and more, has emphasized that the so-called *term*, i.e. the linguistic representation of a concept, can be used as a key for entering specialized texts. Such a key may rely on patterns (a top-down method) or distributional contexts (a bottom-up method).

The above paragraphs have shown a mounting interest in several fields. In the next two sections, I will discuss possible ways of reinforcing this connection by defending a relationship between some lexicographical ideas and practice with Artificial Intelligence methods.

3. Lexicographical philosophy for designing and making online structured knowledge resources: The *Diccionarios Valladolid-UVa*

This section describes some of the main lexicographical ideas underlying the design and making of the *Diccionarios Valladolid-UVa*. For space reasons, I will only focus on very general principles that may influence the use of IA methods in lexicography (see Section 4). The first idea is explained in terms of the tenets of the *function theory of lexicography* (e.g. Bergenholtz and Tarp 2003; Tarp 2008; Fuertes-Olivera and Tarp 2014). Dictionaries are information tools dealing with "things", "facts" and "languages". The advent of online lexicography has reinforced this idea whose practical application is that *lexicographers do not need to design and make different dictionary types*, i.e. monolingual, bilingual, general, specialized, abridged, semi-abridged, learner's and so on. In the digital environment, lexicographers can, and in my opinion should, deal with all the words they can find, describe them in the most precise way, and adapt and store them in Dictionary Writing Systems (DWS) that facilitate different types of searches and retrieval. For practical purposes, this idea implies considering three specific decisions: (a) selection of the headword or lemma list; (b) selection of the empirical sources; (c) selection of the data types to be included in the DWS as well as its *grammar* i.e. specifications about the structure of the dictionary (Kilgarrieff 2006), and *homepage* with specific search and retrieval systems.

Following current practice in lexicography, I consider the selection of the headword or lemma list to be an *ongoing process*, i.e. a process that is never finished. As such, lexicographers must decide on the method for selecting the initial lemma list and its continuous amplification. Since the advent of the *Cobuild Dictionary* (Sinclair 1987), lexicographers have mostly defended a corpus-based approach to headword selection, i.e. the words to be included must be *basically* extracted from corpora on the basis of their frequency and/or keyness. My proposal is different: the selection is a process that needs taking into consideration its inception and continuous development. Its initial stage aims at selecting the words that users *really* look up, as research has discovered that

many of the words lemmatized in existing dictionaries — some researchers claim that almost 80%; see Bergenholtz and Norddahl 2014 — have never been looked up (Trap-Jensen et al. 2014). The *Diccionarios Valladolid-UVa* have followed this methodology and initially selected two lists of single-word lemmas, one for English and one for Spanish. The initial headword lists of the *Diccionarios Valladolid-UVa* were selected at Ordbogen A/S headquarters, a Danish language technology company with whom we have been designing and making our lexicographical projects since 2014 (Fuertes-Olivera 2019).

The Danish company used big data analytics for around two months. The process comprised several stages and was based on an analysis of around one million daily searches in several dictionaries, e.g. an English–Spanish/Spanish–English dictionary, an English–German/German–English dictionary, an English monolingual dictionary, a Spanish monolingual dictionary, and so on. Around 80% of the searches could be matched, i.e. they could be interpreted with the aim of identifying the most popular dictionary articles in both languages. After two months of work with the logfiles of the searches — they amounted to more than 60 million logfiles — IT staff at Ordbogen A/S were able to produce the above-mentioned lists, each comprising around 20,000 single words. These are the words most searched for the period under analysis. The editor of the project systematized them and decided on their amplification, i.e. the process used for adding more lemmas to the initial lemma list. From now on, I will only refer to the Spanish list and the Spanish dictionary of the project.

Systematization means that all the members of the lists must be converted into a *unit of inclusion*, e.g. a lemma in traditional lexicography. Following standard practice, the editor *initially* converted the list into 16,678 single-word lemmas and these were included in the DWS in their canonical form, e.g. the infinitive of the verb, but adapted to an online process of searching (see Section 4, below). In January 2022, the project had completed the lexicographical description of around 10,000 of the initial single-word lemmas. This resulted in around 60,000 meanings or senses (around 6 meanings per lemma). This means that polysemous words are abundant and need some special treatment for making them adequate for disambiguating purposes (see Section 4 below).

The quantity of the meanings included offers some clues on the general philosophy of the project. For illustrative purposes, I will compare the lexicographical data of 25 single-word lemmas with their treatment in the *Diccionario de la Lengua Española* (RAE), which is the dictionary designed and made by the Royal Spanish Academy:

1. *ábaco* (abacus);
2. *abajo* (down); (downstairs);
3. *abalorio* (glass bead);
4. *abanderado* (standard-bearer), (champion), (linesman);
5. *abanderar* (register);
6. *abandonado* (deserted), (abandoned);

7. *abandonar* (leave), (abandon), (desert), (give up), (withdraw), (pull out), (resign), (retire), and so on;
8. *abanico* (fan), (range);
9. *abaratar* (reduce), (lower), (cut) and so on;
10. *abarcar* (cover), (cope with), (embrace), (circle), (take in);
11. *abastamiento* (provisions);
12. *abastecer* (supply);
13. *abastecimiento* (supply);
14. *abasto* (supply), (basic provisions);
15. *abatir* (shoot down), (bring down), (knock down), (pull down), (demolish), (fell), (cut down), (bow), (lower), and so on;
16. *abdicación* (abdication);
17. *abdomen* (abdomen);
18. *abecedario* (alphabet);
19. *abeja* (bee);
20. *abejorro* (bumblebee);
21. *aberración* (aberration);
22. *abertura* (opening), (hole), (slit);
23. *abeto* (fir);
24. *abiertamente* (openly);
25. *abierto* (open), (undone), (split), (openminded), and so on.

The comparison only aims at illustrating some of the key differences between the two lexicographical projects. For space reasons, I will only focus on differences that may be connected with the possible use of AI in lexicography (see Section 4). These lemmas have 153 meanings (around six meanings per lemma) in the *Diccionarios Valladolid-UVa* and 114 meanings (around four and a half meanings) in the *Diccionario de la Lengua Española* (RAE). This difference is relevant and will be explained below.

Amplification is also an *on-going process*. It is initially concerned with the words and expressions that are related with the lemmas of the initial lemma list. In the *Diccionarios Valladolid-UVa*, an expression or "extended unit of meaning" (Rundell 2018) is a linguistic unit formed by two or more orthographical words that expresses a concept and is used as a unit within a sentence. Such a unit is converted into an "extended-unit-of-meaning-lemma" and included in the lemma list if it is still in use, e.g. by being in around 5% of the Google mini-texts used as sources (see below) and in four out of seven existing dictionaries that we also look up during the process of compilation: *Diccionario de la Lengua Española* (RAE); *Diccionario del Español Actual* (Seco et al. 2011); *Diccionario Español-Inglés* (Collins); *Diccionarios.com*; *Lexico Spanish* (Oxford); *SpanishDict*; and *WordReference* (*Spanish*; *Spanish-English*).

The lemmatization of expressions is based on the tenets of *semantic network theory* (see Section 1, above). This theory affirms that humans *mostly* use meaning networks in their daily linguistic interactions. Hence, all the expressions that can be identified during the process of description of the initial lemmas are

lemmatized in the *Diccionarios Valladolid-UVa*. For instance, we have lemmatized *pájaro bobo* (penguin or tropical bird), which was found when I was describing the adjective *bobo* (stupid). Spanish dictionaries typically include *pájaro bobo* as an expression at the end of the dictionary articles for *pájaro* (bird) or as a meaning of the adjective *bobo*. The lemmatization of *pájaro bobo* facilitates searching and retrieval, as will be explained below (see Section 4).

One or more of the 10,000 single-word lemmas already completed are present in around 30,000 "extended-unit-of-meaning-lemmas" (i.e. each single-word lemma is in around 3 extended-units-of-meaning-lemmas). Their lexicographical description has amounted to around 40,000 more meanings, (around 1,25 meaning per expression). For instance, the abovementioned 25 lemmas are part in one or more 83 new extended-unit-of-meaning-lemmas (e.g. *en abierto*; *el que mucho abarca poco aprieta*) also included in the *Diccionarios Valladolid-UVa*. The *Diccionario de la Lengua Española* (RAE) only lemmatizes single-word lemmas, and consequently there are no extended-unit-of-meaning-lemmas in this dictionary, which nests them at the bottom of a dictionary article, usually accompanied with definition and, sometimes, some grammar information. Of the 25 words under analysis, the *Diccionario de la Lengua Española* includes 49 expressions (e.g. *echar abajo*), i.e. almost half of those included in the *Diccionarios Valladolid-UVa*. As before, I will comment on the different numbers between both dictionaries below.

The above figures illustrate an interesting difference between the "single-word-lemma" and the "extended-unit-of-meaning-lemma": extended units of meaning tend to be monosemic entities, and this tendency increases when the number of words forming part of the expression also increases. In other words, the use of extended-unit-of-meaning-lemmas tend to eliminate polysemy and, hence, meaning ambiguity. It seems evident that the larger the number of extended units of meaning included in the structured knowledge resource the less meaning ambiguity in it.

By "related words" I mean the words that stem from the initial single-word-lemmas due to grammar rules. In Spanish, these *basically* affect some nouns, adjectives, adverbs, and verbs. For instance, *abanderado* is a male noun and its related word is *abanderada* (female noun). In traditional Spanish dictionaries such as the *Diccionario de la Lengua Española*, this process of amplification only exists for lemmatizing some manner adverbs, i.e. they are formed by adding *-mente* to the base of an adjective, e.g. *abiertamente*. For the rest of related words, Spanish dictionaries use constructs such as *abanderado, ra* that do not exist in real linguistic interactions (Fuertes-Olivera and Tarp 2022) or do not lemmatize them at all. For instance, the related words of the verbs *abanderar, abandonar, abaratar, abastecer* and *abatir* (they are *abanderarse, abandonarse, abaratar, abastecerse, and abatirse*; they are reflexive or pronominal verbs) and the related word of the adjective *abierto* (i.e. a noun, which is nominalized by putting an article before it, e.g. *un abierto, el abierto, unos abiertos, los abiertos*) are not lemmatized in the *Diccionario de la Lengua Española*.

However, in the *Diccionarios Valladolid-UVa*, this process of amplification is totally active and works with nouns, adjectives, adverbs and verbs. For instance, of the abovementioned 25 single-word lemmas we have included 11 more single-word lemmas: *abanderado* (adjective), *abanderada* (noun), *abastos* (plural noun), *abejorra* (noun), *abierto* (noun), *abierta* (noun), *abanderarse*, *abandonarse*, *abaratarse*, *abastecerse*, and *abatirse* (reflexive or pronominal verbs). These 11 lemmas contain 35 meanings (around 3 meanings per lemma).

The application of this amplification policy means that the *Diccionarios Valladolid-UVa* not only contains a much larger stock of lemmas and meanings but also that it is much more useful for NLP as all relevant word strings, no matter how many words they contain, are lemmas and are described in full. In other words, amplification also offers some clues on another lexicographical idea that underlies the design and making of the *Diccionarios Valladolid-UVa*: *the lexicographical process must be as complete and precise as possible*. The rationale for such a philosophy is twofold: (a) it offers a better description of the language and (b) it facilitates searching and retrieving. Hence, it might be better prepared for using NLP tools, as I will show below (see Section 4). This idea, which is also the philosophy of semi-structured knowledge resources such as *Wikipedia*, eliminates the traditional conception of dictionaries as finished products, subjected to the publication of different editions, and limited, in one way or another, to a particular topic, variety, user's needs, situation, and so on. In sum, the making of dictionaries is a never-ending process that must *constantly* calibrate amplification and the finding of adequate empirical sources.

In today's world, I think that the Internet is the best empirical source for lexicographical work. In other words, the internet is a *lexicographical corpus*, defined by Fuertes-Olivera (2012: 51) as "any collection of texts where lexicographers can find inspiration for completing the dictionary structures they need when making a dictionary". Going a step further, I add that such a lexicographical corpus not only is adequate for making dictionaries but also for any knowledge resources that can be imagined. Consequently, time is ripe for using the Internet to understand the meaning and usage of a particular word or expression in a way that reduces, even eliminates if possible, the "creation and maintenance effort". In this project, we use "Google minitexts", i.e. the two to three lines Google retrieves when making a particular search, for an initial analysis of the meaning and usage of lemmas (Tarp and Fuertes-Olivera 2016). If we find relevant information in them, we click on the homepage and analyze the text or part of it. With this method, it takes around 15 minutes for finding out relevant meanings and linguistic characteristics of most lemmas, especially of extended-unit-of-meaning-lemmas and single-word-lemmas that have up to 7 different meanings, i.e. around 85% of all the lemmas described so far. For instance, only 6 out of 36 single-word lemmas (i.e. the 25 initial lemmas and the 11 created by amplification) contain more than 7 meanings (16%).

The "Google-minitext" method does not properly work with lemmas that have a lot of meanings, e.g. the adjective *abierto* (opened) has 22 meanings and

the verb *hacer* (make, do) has 55 meanings. In such a situation, which currently amounts to around 15% of the lemmas finished so far, we use a "guided search method". It consists in searching in the Internet if the meaning(s) previously found in the consulted dictionaries can be confirmed, i.e. are still used. This method implies the construction of "search strings" formed by the lemma (in quotation marks) plus two or three keywords extracted from the definitions found in the consulted dictionaries. For instance, the search string "abierto" + billete (ticket) + vuelta (return) retrieves more than 3 million hits. Just in the first twenty we can easily confirm the meaning of *abierto* referring to a ticket whose return date is not fixed yet. Such a meaning is a figurative or metaphorical extension of its base meaning. These results were obtained with several different browsers, which indicated that this meaning of *abierto* is still in daily use and that the results are not affected by the search history or cookies of the browser.

The "guide search method" explains the third main lexicographical idea behind the design and making of this dictionary. It can be summarized by saying that *all existing dictionaries, encyclopedias, glossaries as well as grammar books, usage books and the like, should be consulted for inspiration, but not for copying and pasting.*

Finally, the selection of the data types or lexicographical data to be included is basically a cooperative process. *Cooperation is, then, another important idea underlying the Diccionarios Valladolid-UVA.* Cooperation implies the joint work of lexicographers, IT people and experts, e.g. in web design tools. All of them must jointly decide the number of data types they need for describing each lemma and the characteristics of the DWS which must be used. Existing Spanish dictionaries usually use between three and six different data types. The *Diccionario de la Lengua Española*, for instance, always has etymology, abbreviations for indicating part of speech of the lemma and a definition, usually a short one or a synonym. In addition, for many lemmas, it also has expressions (if there are) and one or two clause or sentence examples. For instance, for *abejorro*, the dictionary includes its origin (it comes from "De abeja"), three meanings, two of them described with a short sentence and one with a synonym, and the abbreviation "m", for "masculine" (Example 1):

abejorro

De abeja

1. m. Insecto himenóptero, semejante a la abeja pero más grande, de cuerpo velludo, generalmente negro y con bandas amarillas, que produce un zumbido al volar y vive en enjambres poco numerosos.

2. m. **escarabajo sanjuanero.**

3. m. Persona de conversación pesada y molesta.

Example (1): *abejorro* in the *Diccionario de la Lengua Española* (RAE)

In the *Diccionarios Valladolid-Uva* there are up to 25 possible data types for each Spanish lemma, being the typical one described with fourteen data types: part of speech; inflections, meaning, antonym, synonym, related words, phrase sentence, example sentence, diastatic and/or diaphasic mark (for lemma, antonym, synonym and related words) and diatopic mark for meaning. In addition, some lemmas also have ten more data types: (a) a photo, e.g. for animals, plants and objects; (b) alternative inflections and orthography; (c) part of the conjugation of a verb, (d) proscriptive notes, which are used for recommending between options, e.g. orthographic options, (e) link to a conjugation table, e.g. a verb; (f) link to an external text, e.g. *Wikipedia*; (g) grammar note, (h) usage note, (i) phrase note, which explains the syntactic pattern of an extended-unit-of-meaning-lemma, and (j) synonymy note, which explains possible specific uses of a synonym, e.g. it is only used in Argentina. Example (2) shows *abejorro*, as it is now in the DWS of the *Diccionarios Valladolid-Uva*:

abejorro

noun

<un abejorro, el abejorro, unos abejorros, los abejorros>

meanings

1. insecto parecido a la abeja perteneciente a la familia de los ápidos; tiene el cuerpo más gordo y puede llegar a los 3 centímetros de largo; tiene el cuerpo cubierto de vello oscuro y una trompa muy desarrollada, que emite un zumbido intenso al volar; vive en enjambres poco numerosos debajo del musgo o de las piedras; en el enjambre solo hay una hembra, que es la que fecunda; se alimenta del polen y néctar de las flores

Synonyms for this meaning:

- *abejarrón*
- *abejón*
- *Bombus* <formal>

Phrase sentences for this meaning

- cámaras capaces de captar el vuelo del abejorro con un nivel de detalle espectacular
- diferencia entre el abejorro y la abeja carpintera
- las picaduras de abejorros
- los abejorros, que son bien gorditos y peludos
- si una flor apetitosa está solitaria o concurrida por otros abejorros

Example sentences for this meaning

- El abejorro de tierra o *Bombus terrestris*, es uno de los tipos de abejorros más empleados en la agricultura intensiva, debido a su alto nivel de polinización.
- El pelo grueso actúa como aislante, manteniendo al abejorro a una temperatura adecuada.

Related words for this meaning

- abeja
- abejorro carpintero
- abejorro común
- abejorro cuco

Photo for this meaning:



2. insecto de la familia de los escarabajos; tiene el cuerpo de color marrón oscuro y élitros pardos; puede llegar a los 3 centímetros de largo; roe las hojas de las plantas cuando es adulto y sus raíces en estado de larva; emite un zumbido intenso al volar

Synonyms for this meaning

escarabajo sanjuanero
Melolontha melolonta <formal>

Phrase sentences for this meaning

- el caparazón pardo del abejorro
- los abejorros que decidieron abandonar los árboles en los que se encontraban para invadir prados, jardines y herbazales
- los élitros del abejorro
- los huevos del abejorro bajo los pastos o el césped

Example sentences for this meaning

- Las antenas de estos abejorros se caracterizan por poseer laminillas terminales, capaces de plegarse como varillas de un abanico y formar una maza.

Related words for this meaning:

- abejorro carpintero
- abejorro común
- abejorro cuco

Photo for this meaning



3. en sentido figurado, hombre cuya conversación resulta aburrida, pesada y causa molestia <informal>

Phrase sentences for this meaning

- al abejorro que no aguanta nadie
- mejor ser abejorro que mosca cojonera
- que es un abejorro y un pesado
- un abejorro dando la tabarra

Example sentences for this meaning

- El vecino es un abejorro, como te vea te enrolla hablando de cosas que no te importan.

Related words for this meaning

- abejorra

4. en sentido figurado, persona (hombre o mujer) cuya conversación resulta aburrida, pesada y causa molestia <informal>

Phrase sentences for this meaning:

- no quedar con esos abejorros, sus conversaciones son demasiado cargantes
- se largó en cuanto empezamos a hablar de abejorros

Example sentence for this meaning

- Son como los abejorros: no callan ni debajo del agua.

Example (2): *abejorro* in the DWS of the *Diccionarios Valladolid-UVa*

Comparing examples (1) and (2) offer several conclusions that are relevant for the use of AI in lexicography (see Section 4):

- In the *Diccionarios Valladolid UVa* there are around 400 words for describing *abejorro* lexicographically, whereas the *Diccionario de la Lengua Española* uses fewer than 50 words, i.e. the *Diccionarios Valladolid-UVa* uses almost 12 times more words for describing *abejorro* than the *Diccionario de la Lengua Española*.
- The *Diccionarios Valladolid-UVa* also uses photos for describing physical meanings, e.g. animals in the lemma *abejorro*.
- The *Diccionarios Valladolid-UVa* does not use abbreviations.
- The *Diccionarios Valladolid-UVa* offers a very precise description of meanings and forms, e.g. each definition of each lemma goes with inflections, part of speech, semantic relations, varieties and so on. In other words, most of the lexicographical data are attached to each specific definition.
- Each meaning is independently described.

Using such a large quantity of lexicographical data for describing each lemma influences the design and characteristics of the DWS used for compiling the knowledge resource. In the *Diccionarios Valladolid-UVa*, we are working with an in-house DWS designed by the joint work of IT people at Ordbogen A/S and the editor of the project (Fuertes-Olivera 2019). The DWS of the *Spanish* part of the *Diccionarios Valladolid-UVa* has 30 slots: 25 of these contain the lexicographical data previously commented (see example 2, above). In addition, there are two slots for ordering lemmas and meanings, one slot for internal communication, one slot for administrative purposes, e.g. knowing who has been working in the description of the lemma, and one slot for internal searching, e.g. for searching for "figurative meanings".

4. Using Artificial Intelligence in lexicography

Artificial Intelligence is a wide-ranging branch of Computer Science concerned with building smart machines capable of performing tasks that typically require human intelligence. Russel and Norvig (2010: viii) claim that AI is "the study of agents that receive percepts from the environments and perform actions". In the last five years, there have been several proposals for using AI in lexicography. Plakhotniuk (2018), for example, claims that the collaboration of AI and e-lexicography basically concerns two aspects: (a) improving the data extracted from existing dictionaries and (b) eliminating constraints, e.g. editorial constraints, for digitalizing printed sources. In this section, I will focus on the first aspect and will maintain that the improvement needs not only more lexicographical data (e.g. 12 times more words for describing the word *abejorro* in the *Diccionarios Valladolid-UVa* than in *Diccionario de la Lengua Española*, see examples 1 and 2, above), but also better created, systematized and ordered,

e.g. by trying to reproduce the way the mental lexicon works (see Section 1, above). This means the adoption of three main methodological approaches to dictionary making.

Firstly, our data are prepared for adopting the "closure criteria", which mean that "everything wtwwhat [sic] occurs on the right side of a dictionary must be listed on the left side of the same dictionary" (Dembitz et al. 2005: 1). In other words, it is easy to create a list of word types extracted basically from definitions and phrase and example clauses. Furthermore, such a huge number of lexicographical data for each meaning of each lemma is in line with the so-called "middle ground", i.e. working with big data and good data (Hovy et al. 2013), and allow the creation of "multiple alignment, i.e. treating words in context and comparing their contextual usage metrically" (Dembitz et al. 2005: 2).

Secondly, the Spanish dictionary must be equipped with a search system which will allow users to retrieve *a la carte*, i.e. different data in different situations and for different users. The system will offer users the search button ENCONTRAR UN TÉRMINO (FIND A TERM). This button will allow "users who are uncertain of the exact form of the term to be searched for, or who want to explore the data of a particular term field, to generate their own searches and search strategies by using Boolean operators" (Fuertes-Olivera and Leroyer 2014). For instance, using the search string "+ cost OR gasto-" (Figure 1), users retrieve a series of texts, all of which are clickable and adequate for retrieving the dictionary article in which such texts are, e.g. being part of the phrase or example sentences.

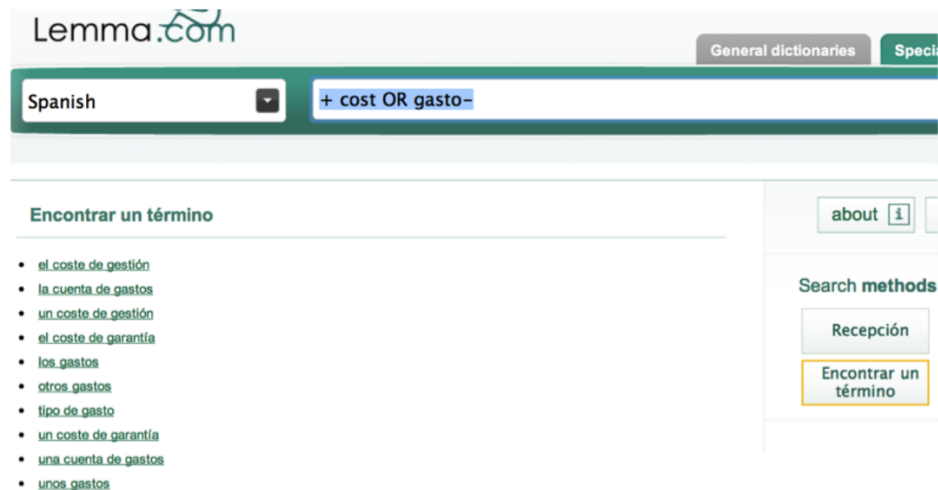


Figure 1: Retrieving texts when searching + cost OR gasto-

Thirdly, the data must be formalized for showing "consistency of inner semantic relations" (Plakhotniuk 2018: 78). These are adequate for creating patterns for machine learning, whose aim, according to Condamines (2018: 343) "is not to build a precise representation of the knowledge but, rather, to detect enough regularities" which will allow us to find constant and relevant relationships, e.g. for eliminating meaning ambiguity and reproducing the processes associated with how words are activated, stored, processed and retrieved by speakers, who never use them in isolation but in contexts. For this, the data stored in the DWS of the *Diccionarios Valladolid-UVa* contains the following:

1. Inflections for nouns and adjectives and conjugations for verbs, both for single-word lemmas and unit-of-meaning-lemmas. These will allow users to retrieve data in any form. For instance, the search engine of the *Diccionario de la Lengua Española* does not work with search strings such as "habríamos querido" (a perfect conditional form of *querer*). This search string will be found in the *Diccionarios Valladolid-UVa*. In other words, users do not need to know the canonical form of the lemma for searching.
2. Very precise definitions (Fuertes-Olivera and Esandi-Baztan 2020); they group semantically similar senses, thus allowing the search engine to search for strings such as that of Figure 1, e.g. the search string "*serpiente* (serpent) + *venenosa* (poisonous) + *americanismo* (Americanism)" will retrieve all the poisonous serpents that are living in South America, whereas the string "*serpiente* (serpent) - *venenosa* (poisonous) + *americanismo* (Americanism)" will retrieve those that are not poisonous. Furthermore, all definitions are self-sufficient, i.e. neither recursive definitions nor synonyms are used for defining each meaning of each lemma. For instance, the second definition of *abejorro* in examples (1) and (2) refers to the same animal; in the *Diccionario de la Lengua Española*, users are given a synonym and linked to a different dictionary article, whereas in the *Diccionarios Valladolid-UVa* the meaning is precise and users have all what they need in the dictionary article. In a similar vein, every time a word related with the lemma is used in the definition, this word is also defined, typically after formulae such as "que es" (that is) or "es decir" (i.e.). For instance, the lemma *avicultura* (aviculture), which is used in the definition of the adjective *avícola* (poultry), is also defined in *avícola* after "que es" (that is) (example 3):

avícola

referido a o relacionado con la avicultura, que es una técnica, actividad, etc. que se ocupa de la cría de aves y el aprovechamiento de sus productos

Example (3): Definition of *avicultura* in the entry for *avícola*

3. semantic networks between definitions and semantic relations, especially with synonyms and, less frequently, antonyms and related words (see example 2). This means that definitions *explicitly* differentiate between similar meanings, e.g. between literal and figurative meanings, animal, things or human beings functioning as actors, and so on. Each of these meanings always goes with up to three synonyms and/or antonyms. The synonyms are replaceable, e.g. in all the phrase and example clauses used in the dictionary article. For instance, in the DWS of the *Diccionarios Valladolid-UVa*, there are three meanings for the Spanish verb *aullar*. In the *Diccionario de la Lengua Española*, this verb only has one meaning and its description is recursive "dar aullidos". Example (4) shows the three meanings and its semantic relations in the *Diccionarios Valladolid-UVa*:
 1. emitir un animal sonidos agudos, tristes y prolongados (animals emit high-pitched, sad and long sounds)
synonym:
 - a. gemir (whine)
 - b. mugir (moo, bellow)
 2. emitir una persona sonidos agudos, tristes y prolongados (persons emit shrill, sad and long sounds)
synonym:
 - a. gritar (shout)
 - b. vociferar (yell)
 3. en sentido figurado, emitir una cosa sonidos agudos, tristes y prolongados (figuratively, something emits high, intense and long sounds)
synonym:
 - a. bramar (roar)
 - b. ulular (howl)

Example (4): Creation of a semantic network between definitions and semantic relations

In example (4), there are three meanings: two literal (the default criterion in Spanish dictionaries) and one figurative, being the actor of the process the main difference between the three meanings: they are respectively an animal, a person, and an object or abstract actor such as the wind. This difference is reinforced with the selection of synonyms *gemir* (whine) and *mugir* (moo and bellow) for animals, *gritar* (shout) and *vociferar* (yell) for human beings, and *bramar* (roar) and *ulular* (howl) for objects or abstract things such as the wind. For instance, *el viento aulla* (the wind howls) is

correct Spanish, whereas *el viento vocifera* (the wind yells) is nonsensical and never used (for instance, I did not find any hit of "viento vocifera" in Google Books. Spanish 2019 (Google Books Ngram Viewer). In sum, these semantic networks are useful for meaning disambiguation (and also for creating different dictionary types, e.g. a dictionary of synonyms containing the meanings and its antonyms and synonyms).

4. similarities, e.g. those formed by the gender of a noun and its reference to a man, woman or person, e.g. the meaning 3 of *abejorro* (example 2, above) starts with *hombre* (man), whereas the meaning 4 does it with *persona* (person), i.e. one refers to a male (and it also has its counterpart *abejorra* (woman); see the discussion on related words above), whereas the other meaning is generic and refers to human beings in general (some other generic also include institutions, organizations, companies, countries, etc. (Fuertes-Olivera and Tarp 2022).
5. phrase and sentence examples (see example 2, above) for each meaning of each lemma. There are *always* from three to six of them for all content words and expressions and between one and two for function words and expressions. They illustrate grammar, usages, e.g. indicating contractions ("del"), singular and plural forms ("abejorro" and "abejorros"), and meanings, e.g. the phrase and example sentences of the meaning one of example (2) confirm the four main attributes of the meaning of this insect: "the insect lives among flowers", "the insect is fatty", "it is dark brown with yellow lines", and "these insects are used in intensive farming".
6. photos for all material beings, objects and things, e.g. animals and instruments. These are not only very useful for describing their meanings in a perfect way but also for differentiating material meanings from abstract ones, most of which are figurative. For instance, in *autopista* (motorway), the DWS contains two meanings: the literal one ("a highway designed for fast traffic, with controlled entrance and exit and so on") goes with a photo of a motorway, whereas the figurative meaning ("an easy way to achieve something without much work") goes without photo but with the indication that this meaning is figurative.

In sum, the data types are all perfectly formalized, standardized and adequate for (a) proposing an interdisciplinary approach to dictionary making, one which meets the needs "of creators of intellectual information systems and dictionaries for humans and machine-based users" (Plakhotniuk 2018: 78), and (b) training the system and hence allowing AI methods reproduce our mental lexicon.

5. Conclusion

This paper has analyzed the possibility of making dictionaries that take into consideration the mental lexicon, i.e. words do not work in isolation; instead,

they are dynamic constructs that are activated, stored, processed and retrieved gradually. This possibility demands the design and making of dictionaries that are very different from the static structured knowledge resources that now exist. In my view, these dictionaries of the future demand new lexicographical thinking, especially one that analyzes the possibility of using AI for solving complex problems such as disambiguating meanings and allowing users search in speaking situations. My proposal, which is illustrated with the *Diccionarios Valladolid-UVa*, is based on several general ideas and specific lexicographical practice, all of which view AI as an adequate methodology for designing and making the dictionary of the future:

- *Lexicographers do not need to design and make different dictionary types, i.e. existing dictionary typologies do not suit AI as humans do not segment their brains into the categories typically used in today's lexicographical work.*
- *The lexicographical process must be as complete and precise as possible, e.g. with the inclusion of photos for describing material objects, inflections and conjugated forms, and so on. This favors searching and retrieving assuming the "closure criteria" and searching and retrieving *a la carte*, i.e. many different possibilities of searching and retrieving.*
- *All existing dictionaries, encyclopedias, glossaries as well as grammar books, usage books and the like, should be consulted for inspiration, but not for copying and pasting, e.g. using the Web as a lexicographical corpus facilitates the process of compilation in around 85% of the lemmas, offers real language use and allows lexicographers to equip their meaning descriptions with phrase and example clauses that help disambiguate meaning and create multiple alignments. Existing resources can help complete description by facilitating the use of "guided searches", which must be employed in very specific situations, e.g. when we have to describe highly polysemous lemmas.*
- *Cooperation is a must and no adequate structured knowledge resource can be implemented without the joint work of lexicographers and experts, e.g. IT and web experts, i.e. dictionaries are no longer the realm of linguists and their making is much more than describing the grammar and meaning of isolated words.*
- *Words should also be considered processes that are always on the move, e.g. as they can have different forms and meaning, we need systems that allow users retrieve them in different usages, forms, formats and purposes.*
- *The Dictionary Writing System must be an in-house system created for specific lexicographical projects and equipped for favoring the creation of patterns, e.g. those formed with semantic networks, and the working of words in contexts, e.g. a large number of phrase and example clauses adequate for AI methodology.*
- *All the lexicographical work must be formalized and standardized, e.g. adequate for human and machine-based users.*

Acknowledgments

Thanks are due to the editor of the issue and to two anonymous reviewers, whose comments and suggestions have greatly improved the first draft of the article.

References

- Aitchison, J. 2003. *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: Blackwell.
- Andrews, S. 1989. Frequency and Neighborhood Effects on Lexical Access: Activation or Search? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(5): 802-814. doi:10.1037/0278-7393.15.5.802.
- Bergenholtz, H. and B. Norddahl. 2014. The Ideal Number of Lemmas in an Ideal Accounting Dictionary. *Hermes, Journal of Language and Communication in Business* 53: 143-150. doi:10.7146/hjlc.v27i53.20988.
- Bergenholtz, H. and S. Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196. 10.7146/hjlc.v16i31.25743.
- Clark, M., Y. Kim, U. Kruschwitz, D. Song, D. Albakour, S. Dignum, U. Cerviño Beresi, M. Fasli and A. de Roeck. 2012. Automatically Structuring Domain Knowledge from Text: An Overview of Current Research. *Information Processing and Management* 48(3): 552-568. <https://doi.org/10.1016/j.ipm.2011.07.002>.
- Collins. *Diccionario Español-Inglés*. <https://www.collinsdictionary.com/es/diccionario/ingles-espanol> [Accessed May 5, 2022].
- Condamines, A. 2018. Terminological Knowledge Bases. Pedro A. Fuertes-Olivera (Ed.). 2018. *The Routledge Handbook of Lexicography*: 335-349. London: Routledge.
- Dembitz, Š., Lj. Jojić and J. Pavlek. 2005. Artificial Intelligence in Lexicography: Croatian Encyclopaedic Dictionary Example. *The 16th International DAAAM Symposium: Intelligent Manufacturing & Automation: Focus on Young Researchers and Scientists, University of Rijeka, 19-22 October 2005, Opatija, Croatia: 1-2*. <https://bib.irb.hr/datoteka/231539.Dembitz-jojic-pavlek.pdf> [Accessed May 5, 2022].
- Diccionarios.com. *Diccionario Gratuito: Español*. <https://www.diccionarios.com/> [Accessed May 5, 2022].
- Emmorey, K.D. and V.A. Fromkin. 1988. The Mental Lexicon. Newmeyer, F.J. (Ed.). 1988. *Linguistics: The Cambridge Survey*: 124-149. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511621062.006>.
- Espinosa-Anke, L. 2017. *Knowledge Acquisition in the Information Age: The Interplay between Lexicography and Natural Language Processing*. [Ph.D. Dissertation]. Barcelona: Universitat Pompeu Fabra. <http://hdl.handle.net/10803/404985> [May 5, 2022].
- Fellbaum, C. (Ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Forster, K.I and S.M. Chambers. 1973. Lexical Access and Naming Time. *Journal of Verbal Learning and Verbal Behavior* 12(6): 627-635. [https://doi.org/10.1016/S0022-5371\(73\)80042-8](https://doi.org/10.1016/S0022-5371(73)80042-8).
- Fuertes-Olivera, Pedro A. 2012. Lexicography and the Internet as a (Re-)source. *Lexicographica* 28: 49-70. <https://doi.org/10.1515/lexi.2012-0005>.

- Fuertes-Olivera, Pedro A.** 2019. Designing and Making Commercially Driven Integrated Dictionary Portals: The *Diccionarios Valladolid-UVa*. *Lexicography* 6(1): 21-41. <https://doi.org/10.1007/s40607-019-00056-8>.
- Fuertes-Olivera, Pedro A. and M.A. Esandi-Baztan.** 2020. Integrating Terminological Resources in Dictionary Portals: The Case of the *Diccionarios Valladolid-UVa*. *Lexikos* 30: 90-110. <https://doi.org/10.5788/30-1-1598>.
- Fuertes-Olivera, Pedro A. and P. Leroyer.** 2014. User-generated Exploratory Search Routes: ENCONTRAR UN TÉRMINO in the Accounting Dictionaries. Ruppenhofer, J. and G. Faass (Eds.). 2014. *Proceedings of the 12th Edition of the Konvens Conference*: 86-95. Hildesheim: University of Hildesheim. http://opus.bsz-bw.de/ubhi/volltexte/2014/289/pdf/konvens_proceedings.pdf.
- Fuertes-Olivera, Pedro A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783110349023>.
- Fuertes-Olivera, Pedro A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicographically-assisted Writing Assistant. *Lexicographica* 36: 257-286. <https://doi.org/10.1515/lex-2020-0014>.
- Fuertes-Olivera, Pedro A. and S. Tarp.** 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating the Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132.
- Fuertes-Olivera, Pedro A, S. Tarp and P. Sepstrup.** 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the *Diccionarios Valladolid-UVa*. *Lexikos* 28: 152-176. <https://doi.org/10.5788/28-1-1460>.
- Google Books Ngram Viewer.** <https://books.google.com/ngrams> [Accessed: May 5, 2022].
- Gruber, T.R.** 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5(2): 199-220. <https://doi.org/10.1006/KNAC.1993.1008>.
- Hoey, M.** 2005. *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge. <https://doi.org/10.4324/9780203327630>.
- Hovy, E., R. Navigli and S. P. Ponzetto.** 2013. Collaboratively Built Semi-structured Content and Artificial Intelligence: The Story so Far. *Artificial Intelligence* 194: 2-27. <https://doi.org/10.1016/j.artint.2012.10.002>.
- Indefrey, P. and Wilhem J.M. Levelt.** 2004. The Spatial and Temporal Signatures of Word Production Components. *Cognition* 92(1-2): 101-144. [10.1016/j.cognition.2002.06.001](https://doi.org/10.1016/j.cognition.2002.06.001).
- Jackendoff, R.** 2002. *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Kilgarriff, A.** 2006. Word from the Chair. G.-M. de Schryver (Ed.). 2006. *DWS. Proceedings of the Fourth International Workshop on Dictionary Writing Systems, 5 September 2006, Turin, Italy (Pre-EURALEX 2006)*: 7. Pretoria: (SF)². <https://tshwanedje.com/publications/dws2006.pdf>.
- Kirkpatrick, B. (Ed.).** 1987. *The Authorized Roget's Thesaurus of English Words and Phrases*. London: Penguin.
- Marslen-Wilson, W.D.** 1987. Functional Parallelism in Spoken Word-recognition. *Cognition* 25(1-2): 71-102. [https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9).
- Miller, G.A.** 1995. *WordNet: A Lexical Database for English*. *Communications of the ACM* 38(11): 39-41. <https://doi.org/10.1145/219717.219748>.
- Oxford Lexico Spanish.** <https://www.lexico.com/> [Accessed: May 5, 2022].
- Plakhotniuk, Ye.** 2018. Lexicography: From Art to Science (Paradigmatic Perspective). *Studia Philologica* 11: 74-80. http://nbuv.gov.ua/UJRN/stfil_2018_11_14.

- Ré, C., A.A. Sadeghian, Z. Shan, J. Shin, F. Wang, S. Wu and Ce Zhang. 2014. Feature Engineering for Knowledge Base Construction. *arXiv* preprint *arXiv:1407.6439*. <https://arxiv.org/pdf/1407.6439.pdf>.
- RAE (Real Academia Española). *Diccionario de la Lengua Española*. [Accessed: May 7, 2022. <https://www.rae.es/>].
- Rousse, C., N. Hernandez and H. Zargayouna. 2018. Domain Ontologies. Pedro A. Fuertes-Olivera (Ed.). 2018. *The Routledge Handbook of Lexicography*: 217-234. London: Routledge.
- Rundell, M. 2018. Searching for Extended Units of Meaning — and What To Do When You Find Them. *Lexicography* 5: 5-21. <https://doi.org/10.1007/s40607-018-0042-1>.
- Ruppenhofer, J., H.C. Boas and C.F. Baker. 2018. FrameNet. Fuertes-Olivera, Pedro A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 383-398. London: Routledge.
- Russel, Stuart J. and P. Norvig. 2010. *Artificial Intelligence. A Modern Approach*. 3rd edition. Boston: Prentice Hall.
- Seco, M., O. Andrés and G. Ramos. 2011. *Diccionario del español actual*. 2nd edition. Madrid: Aguilar.
- Sinclair, J. (Ed.). 1987. *Collins COBUILD English Language Dictionary*. London: Collins ELT. *SpanishDict*. <https://www.spanishdict.com/>. [Accessed: May 5, 2022].
- Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge*. Tübingen: Niemeyer. <https://doi.org/10.1515/9783484970434>.
- Tarp, S. and Pedro 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: 273-295.
- Trap-Jensen, L., H. Lorentzen and N.H. Sørensen. 2014. An Odd Couple — Corpus Frequency and Look-up Frequency: What Relationship? *Slovenščina 2.0: Empirical, Applied and Interdisciplinary Research* 2(2): 94-113. <https://doi.org/10.4312/slo2.0.2014.2.94-113>.
- Traxler, M.J. 2012. *Introduction to Psycholinguistics. Understanding Language Science*. Oxford: Wiley-Blackwell.
- Wikipedia. **The Mental Lexicon**. https://en.wikipedia.org/wiki/Mental_lexicon [Accessed: January 5, 2022].
- WordNet: An Electronic Database. <https://wordnet.princeton.edu/> [Accessed: January 22, 2022].
- WordReference. <https://www.wordreference.com/> [Accessed: May 5, 2022].
- Zock, M. and S. Bilac. 2004. Word Lookup on the Basis of Associations: from an Idea to a Roadmap. *ElectricDict '04: Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries. Association for Computational Linguistics, Geneva, Switzerland, 29 August 2004*: 29-35. <https://dl.acm.org/doi/10.5555/1610042.1610048>.
- Zock, M. and D. Schwab. 2008. Lexical Access Based on Underspecified Input. *Coling 2008: Proceedings of the Workshop on Cognitive Aspects of the Lexicon (COGALEX 2008)*: 9-17. Manchester, UK: Coling 2008 Organizing Committee. <https://www.aclweb.org/anthology/W08-1902.pdf>.

Corpus-based Headword Selection Procedures for LSP Word Lists and LSP Dictionaries

Milica Vuković-Stamatović, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro* (vmilica@ucg.ac.me)

and

Branka Živković, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro* (brankaz@ucg.ac.me)

Abstract: In compiling both Language for Specific Purposes (LSP) word lists for foreign language learners and LSP dictionaries, the headword-selection process is of paramount importance. LSP word lists and LSP dictionaries will function effectively if they contain appropriate terms and register items, i.e. the lexical items that end users need. In this paper, we first present corpus-based LSP word lists, with special emphasis on how they were compiled. In the process, the make-up and size of the specialised corpus are important, as is the choice of the headword selection methods used. Among the possible criteria are word frequency, keyness, specialised occurrence, range, and dispersion, as well as some non-corpus linguistic methods that are more rarely applied. A greater variety of methods is used for compiling headword lists for LSP dictionaries, and of the corpus linguistic methods, frequency is typically solely applied. The article compares headword selection procedures for LSP word lists and LSP dictionaries before discussing how they can mutually inform one another.

Keywords: LANGUAGE FOR SPECIFIC PURPOSES, LSP WORD LIST, LSP DICTIONARY, CORPUS LINGUISTICS, HEADWORD LIST, TERMS, HEADWORD SELECTION

Opsomming: Korpusgebaseerde lemmaseleksiëprosedures vir TSD-woorde-lyste en -woordeboeke. In die samestelling van beide Taal vir Spesifieke Doeleindes-(TSD-)woorde-lyste vir vreemdetallearders en TSD-woordeboeke is die lemmaseleksiëproses van kardinale belang. TSD-woorde-lyste en TSD-woordeboeke sal effektief funksioneer indien hulle toepaslike terme en registeritems, m.a.w. die leksikale items wat eindgebruikers benodig, bevat. In hierdie artikel word korpusgebaseerde TSD-woorde-lyste eerste bespreek, met besondere klem op hul samestelling. In hierdie proses is die samestelling en grootte van die gespesialiseerde korpus, asook die keuse van die lemmaseleksiëmetodes wat gebruik word, belangrik. Onder die moontlike kriteria is woordfrekwensie, sleutelstatus, gespesialiseerde voorkoms, omvang en verspreiding, asook enkele nie-korpus-linguistiese metodes wat minder gereeld toegepas word. 'n Groter verskeidenheid metodes is gebruik vir die samestelling van lemmalyste vir TSD-woordeboeke, en van die korpuslinguistiese metodes is slegs frekwensie tipies toegepas. Lemmaseleksiëprosedures vir TSD-woorde-lyste en TSD-woordeboeke word in die artikel vergelyk voordat daar bespreek word hoe

hulle mekaar wedersyds van inligting kan voorsien.

Sleutelwoorde: TAAL VIR SPESIFIEKE DOELEINDES, TSD-WOORDELYS, TSD-WOORDEBOEK, KORPUSLINGUISTIEK, LEMMALYS, TERME, LEMMASELEKSIE

1. Introduction

Word lists have many purposes in the process of teaching and learning a foreign language: they can be used as resources for vocabulary learning (Khani and Tazik 2013; Yang 2015), guidelines for designing curricula and courses, as well as for selecting reading and listening materials (Wang, Liang and Ge 2008; Jin et al. 2013), and guidelines for teachers in organising their explicit vocabulary teaching (Khani and Tazik 2013). The selection of headwords for inclusion in certain word lists has become an important strand of applied research in the field of foreign language teaching and learning in general, and language for specific purposes (LSP) in particular. As vocabulary sizes attained by native speakers are never attained by a vast majority of foreign language learners, the rationale guiding this type of research is to produce word lists of the sizes which are manageable for them to learn from. Word lists should provide language learners with the most useful words they need for a particular language function they are pursuing, for instance, attending university studies in a foreign language or reading research articles from a particular specialist field in a foreign language. Some of these functions are related to LSP contexts and for them, consequently, LSP word lists are produced. Most of them are, in fact, English for Specific Purposes (ESP) word lists, given that English is the language which is most widely taught as a foreign language around the world.

In the past, both general and LSP word lists used to be compiled manually, typically relying on the compiler's intuition and, more rarely, on an authentic corpus of a very limited size by today's standards (West 1953; cf. Gilner 2011). However, over the past two decades, they have principally been derived from vast authentic corpora of general or specialised texts, which are carefully constructed having particular types of foreign language learners in mind, and then scanned for words meeting certain criteria or a combination of criteria, such as the frequency of occurrence, distribution, range, or keyness (Coxhead 2000; Coxhead and Hirsh 2007; Brezina and Gablasova 2013; Browne et al. 2013a, Gardner and Davies 2014, etc.). The choice of the criteria and the related "cut-off" points (for instance, how frequent a word has to be to be included in a certain word list) are informed by the target users' needs and involve a number of decisions during the compilation of the list. As corpora and software solutions evolve, so do the different methods for selecting those words. In this paper we will discuss various word lists intended for LSP learning, with a focus on how they were compiled.

Selection of headwords for any dictionary, including specialised dictionaries, is also governed by the needs of its end users (Fuertes-Olivera and Arribas-

Baño 2008), i.e. what should be taken into account are different types of users, user situations and user needs (Tarp 2008), according to the theory of lexicographic functions (Bergenholtz and Tarp 1995; Tarp 2008). In principle, there are four main methods of selecting headwords for dictionaries — these assume relying on the existing dictionaries, grammar and etymology, canonical literary texts, or corpora (Esandi-Baztan and Fuertes-Olivera 2020). The fourth method, compiling headword lists based on corpora, has been an option for the past few decades and is now widely used in the process of making general dictionaries. However, as Bowker (2010: 166) notes, the use of corpus linguistic methods has been rather slow to take hold in the creation of specialised dictionaries. When it comes to the methods and procedures of compiling corpora for the purpose of creating LSP dictionaries as a type of specialised dictionaries, one may only rarely find detailed accounts regarding this issue (cf. Khumalo 2015; Đurović 2021; Kruse and Heid 2021). Also, typically, few details are also presented relating to the corpus-linguistic procedures employed as part of the process of selecting headwords from specialised corpora — most studies only briefly note that it is the frequency criterion that was applied (cf. Rundell and Kilgarriff 2011), without delving into the type of details that are provided by various specialised word-list compilers (cf. Lei and Liu 2016; Todd 2017; Dang 2018, etc.). In addition, in these accounts, further corpus-linguistic procedures for headword selection beyond simple frequency are only sometimes mentioned in LSP dictionary research and projects (cf. Khumalo 2015; Đurović 2021; Kruse and Heid 2021).

In this paper we compare corpus-based headword selection procedures used for producing LSP word lists and LSP dictionaries, bearing in mind that there are some similarities (although, also, important differences) between these two types of lexicographic products. We focus on the steps in headword selection that are based on corpus bearing in mind the important place that corpora currently have in their creation. The premise from which we depart is that the two fields can mutually inform and contribute to one another in terms of the corpus-based headword selection procedures.

We will first present an overview of word lists, with a special focus on LSP word lists and how they are produced (section 2), after which we discuss LSP dictionaries and how headwords are selected for them (section 3). Section 4 compares headword selection for LSP word lists and LSP dictionaries.

2. Word lists

This section first provides a brief overview of general and academic word lists, after which the focus is narrowed down to discipline-specific or LSP word lists.

Reviews of word lists used for the purposes of foreign language teaching and learning typically start by presenting West's General Service List (GSL) (1953) (cf. Coxhead 2000; Coxhead and Hirsh 2007; Gardner and Davies 2014; Dang and Webb 2016; Dang, Coxhead and Webb 2017; Dang 2018; McQuillan 2020, etc.).

Although West's list was not generated using computer software, it was based on an authentic word corpus of 5 million words representing General English. About 2,000 word families¹ were manually extracted and suggested to be the first words to be learned by any English language learner (they were mostly chosen according to the frequency criterion). This word list was very influential in English Language Teaching (ELT) and was used widely for decades (Nation 2013; Coxhead 2018). The emergence of the computer solutions providing data on a word's frequency and coverage in a corpus showed why — it turned out that West's list covered about 80% of the words used in most general English texts, or 4 in every 5 words. As English has about 70,000 word families (Nagy and Anderson 1984; Nation 2013), this word list proved to be a very useful resource (Coxhead 2000; Nation 2013).

In the ensuing decades, other English words were built too (for instance, Champion and Elley 1971; Praninskas 1972; Lynn 1973; Ghadessy 1979; Xue and Nation 1984, etc.), however, the next word list which can match the influence of the GSL, the Academic Word List (AWL), came only in 2000 (Coxhead 2000). Its influence lies not only in how widely it was used in ELT, but the methodology of its compilation also set standards for many of the ensuing word lists (among them, Fraser 2007; Konstantikis 2007; Wang, Liang and Ge 2008; Khani and Tazik 2013; Valipouri and Nassaji 2013; Hsu 2013; Minshall 2013; Hsu 2014; Liu and Han 2015; Yang 2015; Lei and Liu 2016, etc.). The AWL contains 570 word families which are common in academic writing. To produce the list, Coxhead compiled a corpus of 3.5 million words of academic texts. The words were extracted according to the following criteria: (1) specialised occurrence (the words had to be outside high-frequency general words (outside the GSL in this case)), (2) frequency, (3) dispersion (the words had to occur in all the corpus's subsections while featuring a certain frequency in all of them, and they also had to occur in at least half of the academic disciplines involved in the corpus) (Coxhead 2000). These carefully weighed and strict criteria ensured that the word list would have a substantial coverage in any academic corpus, not just in the one it was derived from (Coxhead 2000). Indeed, the AWL's coverage of 10% in the corpus of its origin held strongly in many other academic corpora compiled later — for instance, it featured 10.07% in the academic medical corpus (Chen and Ge 2007), 11.17% in the academic applied linguistics corpus (Vongpumivitch et al. 2009), 9.96% in academic chemistry corpus (Valipouri and Nassaji 2013), etc. These impressive results confirmed that any future word list would have to be carefully made, so as to be as useful as possible in a variety of similar language contexts.

One of the rare issues that may be contended against the AWL is the relatively small corpus it was derived from taking into account that it aims to be a general academic word list, an issue which the ensuing general lists have been trying to overcome. The dated GSL needed to be replaced and two new GSLs were offered to both research and instructional purposes in 2013. Brezina and Gablasova (2013) based their New GSL, containing about 2,500 lemmas, on a combined corpus of samples from 4 different corpora, together making 12 bil-

lion words. The lemmas from each of the 4 corpora were selected based on the criterion of the Average Reduced Frequency (this measure is obtained from the absolute frequency of the word and its distribution in the corpus (Savický and Hlaváčová 2002)), and then the 4 lists were compared for overlaps — the shared items entered the New GSL. The same year, Browne, Culligan and Phillips (2013a) used a 273-million-word section of the Cambridge English Corpus to derive their list of about 2,800 lemmas based on the frequency criterion. Both lists outperform the old GSL in modern corpora, typically by a few percentage points.

Browne, Culligan and Phillips (2013b) also created the New AWL, containing 963 lemmas, by excluding the words already contained in the NGSL. Another replacement for the AWL was offered by Gardner and Davies (2014), who used a 120-million-word corpus (an academic subsection from COCA), to produce a list of about 3,000 lemmas (the Academic Vocabulary List, or the AVL). They did not exclude any group of words, but employed the keyness criterion solely: the authors took into account the ratio of words in their academic corpus, compared to a non-academic corpus. Newman (2016) and Hernandez (2017) found that the AVL outperforms the old AWL, while not much data is available on how the NAWL performs against other similar lists.

Other researchers have investigated whether lists such as the AWL might be created for other languages. Cobb and Horst (2004) studied the vocabulary profile of French and determined that the high-frequency vocabulary of this language is in fact more frequent than the high-frequency vocabulary of English (2,000 most frequent French words reach a 90% coverage in most texts they examined), which excludes the need for creating additional lists for learners as these would reach very small coverages. Such results for French did not discourage other researchers to pursue creating corpus-based academic word lists for other languages, however. A Nordic joint-research project resulted in the creation of the academic word lists for Swedish, Norwegian and Danish (Kokkinakis et al. 2012; Jansson et al. 2012; Ribeck et al. 2014; Johannessen et al. 2016). Two more independent lists have also been created for Danish — a word list of general, high-frequency items (2,000 words), as well as a word list of academic vocabulary (402 words) (Jakobsen et al. 2018). An Academic Vocabulary List in Russian has also been compiled recently (Talalakina et al. 2020). The development of all these word lists heavily relied on the English word-list projects presented above.

The word lists mentioned so far include general and non-discipline specific academic word lists. Unlike these, other word lists are much more specialised and these are the focus of this paper. They and the methods used for compiling them will be presented in the following section.

2.1 Corpus-based headword selection procedures for LSP word lists

Realising the importance of the role of the communicative contexts in which certain foreign language learners will typically find themselves (Miller 2014: 305),

teaching LSP began to be strongly differentiated from teaching General Foreign Language in the 1960's. LSP teachers and researchers realised that taking the learners' specific needs into account, particularly their vocabulary needs, led to more effective teaching of the specialised language that they needed. With the rise of the ITC industry, corpus-based discipline-specific word lists, produced with the use of computers and from vast corpora, began to emerge at the turn of this century.

An overview of recent LSP word lists, along with the details of the corpora from which they were derived and the methods used for their creation, is given in the Appendix (while not entirely exhaustive, the table presents most of the word lists which have been described in scholarly papers). As was the case with general and academic word lists, the field of researching and compiling LSP word lists is almost exclusively related to the English language and, consequently, English word lists dominate the literature (as can be seen in the Appendix). Many of these lists follow in the AWL's footsteps given that they rely or build on the criteria used by Coxhead (2000) (see Section 2). Here we will provide a generalised description of the corpora and methods typically used to create LSP word lists.

The texts for LSP corpora are chosen bearing the LSP word list's target users in mind. The corpora from which word lists are produced are typically custom-made, which makes their creation challenging and time-consuming. They also need to be of a relevant size. The corpora from which the LSP word lists were made vary widely in terms of their size — most of the word lists were developed from a specialised corpus of 1–2 million words (Mudraya 2006; Coxhead and Hirsh 2007; Wang, Liang and Ge 2008; Vongpumivitch, Huang and Chang 2009; Khani and Tazik 2013; Yang 2015; Todd 2017; Kwary and Artha 2017; Tongpoon-Patanasorn 2018; Đurović 2021). However, a recent tendency is to use larger corpora — most of the corpora from the last decade featured 4 or more million words (Valipouri and Nassaji 2013; Hsu 2013; Hsu 2014; Lei and Liu 2016; Moini and Islamizadeh 2016; Dang 2018; Khany and Kalantari 2021; Kamrotov et al. 2022). The biggest corpus used is the most recent one — a corpus of almost 30 million words of accounting research articles, which was used to obtain a list of the most frequent 658 accounting words (Khany and Kalantari 2021).

The LSP word-list compilers who intend to apply the word selection criteria of range and dispersion need to think carefully about the make-up of their corpora as they generally need to have equal subsections of texts from various subfields. These corpora thus need to be well-structured and balanced; even though this is a challenging task, some researchers were able to produce significantly large and at the same time well-structured corpora — for instance, such is the English Hard Science Spoken Corpus of 6.5 million words, produced by Dang (2018), which features 12 subsections representing 12 hard science disciplines. This size is all the more impressive bearing in mind that this is a corpus of spoken language.

The sizes of LSP word lists also vary widely — from 92 (Martínez, Beck and Panza 2009) to 1,595 headwords (Dang 2018) and, again, the needs of the end users are taken into account when determining the list's size, as is the case with dictionaries.

The criteria used for the selection of words for various recent LSP word lists can be summarised as follows:

1. frequency (the number is set depending on how large a list is wanted),
2. specialised occurrence (being outside the most frequent 2,000 or 3,000 words, so as to avoid general high-frequency words; additionally, being outside the most frequent academic words (as represented by a chosen academic word list); finally, this also assumes the exclusion of proper nouns, symbols, abbreviations, numbers, non-words, etc.),
3. dispersion (typically, occurrence in at least half the disciplines/subsections which make the corpus, or being below some dispersion value (different methods for determining these are available)),
4. keyness (being found in the specialised corpus more frequently than in a reference corpus),
5. expert opinion (experts use rating scales and assign more points to more technical words),
6. cross-comparison with specialised dictionaries.

The first four are purely corpus-linguistic methods and assume automatic extraction of words based on the word-list compiler's decisions regarding the thresholds applied, while the last two depend on consulting either experts or specialised dictionaries, and are much more time-consuming. The final two steps have been generally avoided in developing most LSP word lists; having applied several corpus-linguistic filters, the word-list compilers found them unnecessary. Experts and dictionaries were consulted in the creation of just four out twenty-four LSP word lists presented in the Appendix (Wang, Liang and Ge 2008; Valipouri and Nassaji 2013; Jin et al. 2013; Tongpoon-Patanasorn 2018).

It should be added that the finalised LSP word lists are also typically validated in one or several independent corpora (following Coxhead 2000) and, if their expected coverages hold in new corpora, such word lists are assumed to be truly representative.

Few studies, typically those early ones or those using a vast corpus, used just one word-selection criterion (typically, frequency or keyness) (Mudraya 2006), while most of the studies employed a combined approach by using several of the methods — most often, following Coxhead's method (2000) (the first three steps above). None of the studies applied all the six methods combined.

As can be seen, the field of producing and investigating word lists developed as part of applied linguistics by Anglo-Saxon scholars, who, despite the fact that there are now many authors in it who are not Anglo-Saxon, still dominate it to a large extent. Most of the word lists are in fact English word lists. The creation of word lists is guided by pragmatic principles and the field remains

atheoretical. So far, in the literature, there have not been any proposals to introduce a theory which would support the field.

3. LSP dictionaries

As Bowker (2010) explains, LSP dictionaries belong to specialised dictionaries, i.e. dictionaries which treat specialised fields. They are also seen as a type of restricted dictionaries (Burkhanov 1998), where the term *restricted* does not imply their smaller size but reflects the fact that they focus on specific and precise vocabulary (Mihindou 2004). LSP dictionaries exist in many fields of knowledge (Landau 2001), while developing the metalexigraphy related to them is in full swing (Fuertes-Olivera and Arribas-Baño 2008).

While the Anglo-Saxon strand in lexicography is mostly atheoretical (as was the case with the field of compiling word lists), the strand influenced by German and Nordic scholars advocates for developing lexicographical theories for guiding dictionary research and compilation (Fuertes-Olivera et al. 2013). As mentioned earlier, what is taken into account in the process of compiling any dictionary, including a specialised one, are the different types of users, user situations and user needs related to them, in line with the theory of lexicographic functions (Bergenholtz and Tarp 1995; Tarp 2008). This is one of the lexicographic theories which is very influential in pedagogical lexicography, including specialised pedagogical lexicography.

As for users, specialised dictionaries have a more limited target audience than general dictionaries. According to Bergenholtz and Tarp (1995), their user type is decided based on user's mother language, level of encyclopedic knowledge, and native- and foreign-language competence. Applying these criteria, the authors identify four major user types for specialised dictionaries: experts with a high level of encyclopedic and foreign language competence, experts with a high level of encyclopedic competence and low level of foreign language competence, laypersons with a low level of encyclopedic competence and foreign language competence, and layperson with a low level of encyclopedic competence and a high level of foreign language competence. Some more types are added by Fuertes-Olivera and Arribas-Baño (2008), who, among these user types, identify the following: experts from the specific field, semi-experts, experts from related or other fields, interested laypeople who would like to read some books or periodicals from the field, LSP students, translators, interpreters, etc.

Tarp (2010) argues that there are many situations in which learners can benefit from specialised dictionaries — cognitive situations include systematic study of the specialised subject field and of problems related to the translation of specialised texts; communicative situations include reception and production of specialised texts in the mother tongue and in a foreign language, as well translation of specialised texts, while practical situations refer to various operative and interpretive situations.

The mentioned user types have different needs in the mentioned different types of situations. These needs can be primary or function-related needs, which are the needs for information necessary to gain knowledge or solve a problem through using a dictionary, or they can be secondary or usage-related needs, which includes the need to know something about a specific dictionary and to know how to use it (Tarp 2008).

There are different classifications of LSP dictionaries but we will briefly mention two which are relevant for our paper. Based on their size, there are two basic types — *maximising* LSP dictionaries, which attempt at covering as much of a field's terminology as possible, and *minimising* LSP dictionaries, in which a portion of the terminology is covered, typically only the most frequent items (Bergenholtz and Tarp 1995). Another possible classification recognises LSP dictionaries containing field-specific *terms* only, as opposed to general words, and *hybrid* LSP dictionaries, which combine both specialist and general words (Campoy Camillo 2002; Bowker 2010).

LSP dictionaries for learners are a subtype of specialised dictionaries which are intended to assist users in learning about the terms and concepts used in a specific field, in one or more languages (Bowker 2010). Their purpose is to serve as auxiliary tools in the process of teaching and learning the language for specific purposes (Fuertes-Olivera and Arribas-Baño 2008). According to the mentioned theory of lexicographic functions, they are utility tools which assist learners in the process of learning LSP.

3.1 Corpus-based headword selection procedures for LSP dictionaries

The process of headword selection is central in learner's lexicography (Xue and Tarp 2018), given that "dictionaries only function if they contain appropriate data," Nielsen (2018: 79). In this process, the three main questions that need to be posed refer to the size of the headword list, criteria and principles guiding their selection, and the empirical basis that their selection relies on (Tarp 2008). Tarp (2008) further suggests that headwords can be selected based on three sources, i.e. by means of introspection, using available descriptions in various publications (dictionaries, textbooks, etc.), and based on corpora. Building corpora as part of the preparatory stage for headword selection for LSP dictionaries is significant (Nkomo 2008: 105). Having compared corpus-based and intuition-based approaches, Verlinde and Selva (2001: 597) argue that it is the corpus-based lexicography that gives the "strong and necessary empirical evidence to the lexicographer's personal intuition", but they also note that intuition still remains helpful in filling in the gaps in cases when corpora are not balanced.

As said earlier, Bowker (2010: 166) argues that the use of corpus linguistic methods has been rather slow to take hold in the creation of specialised dictionaries, on account of the fact that not so many specialised corpora are available. Specialised corpora used for making dictionaries also tend to be relatively small, especially in comparison with the mega-corpora used for producing

general dictionaries. Bowker (2010) cites the example of the specialised dictionary *Dictionnaire d'apprentissage du français des affaires* (DAFA) as a commendable example, given that it was based on a corpus of 25 million words. Taking into account the latest technological developments, recently, the compilation of such, relatively large, corpora has become much less of an issue.

The mentioned theory of lexicographic functions (Bergenholtz and Tarp 1995; Tarp 2008) suggests that headwords should be selected according to user's needs. When selecting headwords based on corpora, this, among other things, practically means that it is the user needs which govern the selection of texts which will enter such corpora. To illustrate how this can work in practice, we will briefly note how headwords for a Spanish accounting dictionary were selected (Fuertes-Olivera et al. 2013). Thus, following the mentioned function theory and the principle of *relevance*, the authors created a list of around 6,000 accounting texts, based on which three experts in accounting and one lexicographer derived a stock of around 3,000 terms. Of the corpus-linguistic methods applied in this processing of the corpus, the authors calculated the word *frequencies* in their corpus, to inform their decisions of which terms to include in their specialised dictionary. They also used the Internet as a corpus and performed Google searches using particular word strings to extract additional 1,000 terms. Finally, 2,000 more terms were added through intensive reading of basic accounting texts. Such a hybrid approach was applied so as to ensure that the principle of relevance is adhered to. The authors argue and add that future updates of the term stock will be done by additionally analysing the log-files related to the online use of this dictionary (Fuertes-Olivera et al. 2013).

Other authors, too, mention applying the principle of frequency as one of the key steps taken in the process of selecting headwords for dictionaries (cf. Campoy Cubillo 2002; Hanks 2012; Rundell and Kilgarriff 2011). This criterion provides "solid empirical evidence for the occurrence of a word in actual language" (Xue and Tarp 2018). At the same time, they also argue that frequency may be misleading in some specialised fields which are updated constantly, such as accounting (Fuertes-Olivera and Nielsen 2011). Rundell and Kilgarriff (2011) rightly mention the fact that frequency is not a good selection criterion for extracting multiword items as candidates for headword lists. Likewise, Nielsen (2018: 81-82) suggests that frequency solely cannot guarantee that all relevant words will be selected, but that it should be used as a basis for the further selection process.

In some LSP dictionary compilation projects, similar to the methodology used in the production of LSP word lists, frequency is combined with additional corpus-linguistic methods — thus, for instance, Khumalo (2015) and Đurović (2021) also use keyness; however, they do not ensure that the corpus contains equal shares of various subdisciplines of the field which it represents and, consequently, they do not apply the range filter. Some LSP dictionary compilers additionally use a more innovative, pattern-based approach (Kruse and Heid 2021).

Frequency and relevance are suggested as two major criteria in Xue and

Tarp too (2018). However, Tarp (2008) warns against the exalted status given to corpora and corpus-linguistic methods by certain lexicographers, arguing that corpora, however large they may be, can still be unrepresentative, and that the criteria of *relevance* and *systematicity* also need to be taken into account.

What may be deduced from these various accounts is that corpora play an important role when selecting headwords for specialised dictionaries, and that word frequencies in a corpus can significantly inform the process of headword selection.

4. Comparison of corpus-based headword selection procedures for LSP word lists and LSP dictionaries

As we have seen, headword selection procedures for both LSP word lists and LSP dictionaries are guided by the needs of their users. The chief users of LSP word lists are LSP learners. LSP word lists are also used by LSP teachers and LSP material developers but, again, to the benefit of their end users — LSP learners. When it comes to the users of LSP dictionaries, as noted earlier, LSP learners make up an important category among them, however, many more categories of users are possible as well (e.g. translators, semi-experts, experts from other fields, etc.). This basic distinction in the types of users of the two products — LSP word lists and LSP dictionaries, has implications for how headwords are selected as part of their compilation procedures.

When comparing corpus-based headword selection procedures for LSP word lists and LSP dictionaries, we can see that the former are compiled using corpus-linguistics methods almost exclusively, whereas a greater complexity of methods is used for the latter. A significant part of this difference may be explained by the respective homogeneity and heterogeneity of the end users of the two products, as explained above.

The corpora from which LSP word lists are derived are rather large and typically well-structured and balanced, as we have seen. The details regarding their make-up are usually presented very precisely and transparently in the scholarly papers on LSP word lists, as well as given central prominence in them. On the other hand, the descriptions of corpora used for developing headword lists for LSP dictionaries are usually not presented in such details and, typically, in the papers describing these projects relatively little space is devoted to the process of term extraction. In addition, equal representation of various subfields is rarely ensured in them. LSP word lists compilers argue that this is a good practice which allows that the frequencies of the terms obtained to reflect all subfields equally, and we tend to agree here. An implication from this comparison is that LSP lexicographers might invest this type of effort into compiling corpora from which they intend to extract terms. Moreover, given that many useful and balanced corpora have already been produced as part of LSP word-list research, some of these could be used for making LSP dictionaries as well.

Both compilers of LSP word lists and compilers LSP dictionaries use frequency as a major criterion for deciding which words should enter their products. In the process of producing LSP word lists, compilers typically either follow the cut-off points used in seminal research (such as Coxhead 2000) or, more frequently nowadays, the cut-off points are governed by the coverage achieved with the obtained word list, a coverage that allows for a certain threshold of reading or listening comprehension to be met.

As for LSP dictionaries, in the literature we have not encountered detailed arguments around the chosen thresholds. The size of LSP dictionaries, in theory, should be governed by the user needs (even though there are always practical and financial constraints to LSP dictionary projects) (Tarp 2008). However, so far, no method of quantifying them has been developed yet (and might not be, given the complexities involved).

Research and projects involving LSP dictionaries frequently mention that frequency cannot be the sole criterion for selecting headwords, usually citing *relevance* as another major criterion to be applied, which, however, is much more difficult to define and employ. Likewise, as we have seen in the LSP word-list research, the criterion of simple frequency is also never applied as the sole criterion. Additional criteria may be applied as well, although these are also based on frequency to some extent. Thus, an important criterion for selecting headwords for LSP word lists is that of *specialised occurrence*, as presented earlier, applied by excluding words which are highly frequent in general, reference corpora (typically 2,000 to 3,000 most frequent words in the case of English). Academic words can also be excluded, to ensure more technicality. Another criterion is that of *range* — applying this filter ensures that a word appears in a sufficient number of a discipline's subfields, so that it is equally valuable across that discipline, and not more valuable for some subspecialisations and less valuable for others. To apply this criterion, however, one needs a corpus with equal subsections from the various subfields, as argued above. If the required structure of the corpus is not achieved, various *dispersion* thresholds can be applied. These criteria for guiding term extraction are rarely used when compiling headword lists for LSP dictionaries.

One more criterion frequently mentioned when compiling LSP word lists is that of *keyness*, which is relatively easy to apply as no special make-up of the corpus is needed for it. As explained earlier, the frequency of the words in a specialised corpus is compared against that featured in a reference corpus and so the words found to be much more frequent in that specialised corpus are identified as terms. As we have seen, this criterion is sometimes used when extracting terms for LSP dictionaries as well.

Very often, the mentioned additional criteria are used in combination when compiling LSP word lists. LSP word list compilers argue that applying them, in addition to simple frequency, ensures that the headwords selected are indeed *relevant*. The notion of *relevance* is more difficult to define for a product such as an LSP dictionary given its rather heterogeneous target audience; how-

ever, applying at least some of the forementioned filters could help facilitate and automate that process.

The mentioned filters used for obtaining LSP word lists have been found deficient, however, when it comes to extracting multi-word units and collocations and, in fact, none of the word lists presented here contain such items. This is a major drawback to LSP word lists in general and a limitation that should be borne in mind if one were to apply some of the said methods for selecting preliminary headword lists for LSP dictionaries. Still, the ease with which most of the presented filters can be applied certainly recommends them for use in combination with other methods.

Once an LSP word list is obtained via corpus linguistic methods, the work of the LSP word list compiler is either completed or almost completed in most cases, whereas much more work remains for a lexicographer compiling a headword list for their LSP dictionary.

The principle of *systematicity* is hardly ever applied to the LSP word lists obtained via corpus-linguistic methods. For instance, the Science List (Coxhead and Hirsh 2007) contains names of some common chemical elements (such as *oxygen*, *potassium*, etc.), while the names of other common elements are not mentioned (such as *sulfur*, for instance); it is debateable whether the word *sulfur* is less useful to a science student learning English than the word *potassium*, for instance. Moreover, the Science List includes the word *chloride*, however, it does not include the name of the chemical element whose negatively charged ionic form it represents — *chlorine*. Thus, in general, word-list makers rely, perhaps too much, on automated procedures and avoid discussing these types of issues. As opposed to that, in LSP dictionary research and projects, systematicity is one of the central principles guiding the creation of headword lists. Observing this principle when developing LSP word lists, we argue, could improve them, as the illogicalities of the types exemplified above typically stem from the imperfections of the corpus (in this case, the over-presence of texts mentioning the names of some particular chemical elements) and ought to be corrected when noticed. We would argue that, however large, well-structured and balanced a corpus may be, it will always suffer from some imperfections and cannot be trusted entirely.

When finalised, LSP word lists are sometimes subjected to validation in additional corpora (not the ones they were derived from), to test how much coverage they would have in new texts. Validation, although effort- and time-consuming, is a commendable step to be taken, in our opinion. The frequency of preliminary, candidate headword lists for LSP dictionaries, could also be checked in additional specialised corpora, so as to, perhaps, rule out some candidate terms which in validation corpora feature significantly lower frequencies as opposed to that from the first corpus.

In developing LSP word lists, experts from the specialist fields are almost never involved, while they are always involved in compiling LSP dictionaries. This step is usually skipped in the making of modern word lists, given that

several automatic corpus-linguistic filters have already been applied. Although this is a demanding step, involving experts in the creation of any LSP product is advisable.

5. Conclusion

In this paper, we presented most modern LSP word lists and commented on how they were created. We also discussed corpus-based headword selection procedures for LSP dictionaries. A number of both similarities and differences were found in the two selection procedures and it was noted that both of them could, in some ways, benefit from being mutually informed.

On the one hand, more effort could be invested in the creation of LSP corpora, in terms of their size, make-up and balance, and also more corpus-linguistic selection procedures could be applied when compiling headword lists for LSP dictionaries than is currently typically the case, to facilitate the process. More transparency and precision when reporting on the corpora used and the corpus-linguistic methods applied for compiling headwords lists for LSP dictionaries is also advised. Lists obtained should also be validated in additional corpora, when possible.

On the other hand, the creation of LSP word lists could be improved by applying additional non-corpus linguistic methods in their compilation, which is necessary to eliminate the illogicalities stemming from imperfectly balanced corpora, as well as to add the necessary multi-word units to them.

Another observation that imposes itself from the comparison made in this paper is that the compilation and study of word lists remain atheoretical, while at least one strand of LSP dictionaries research has strong theoretical foundations. As we conclude this paper, we will ask the reader and ourselves if, perhaps, the moment has arrived that the field of word-list compilation and research be supported by a theory similar to that of the theory of lexicographic functions.

Endnote

1. A word family includes the headword with all its inflected and derived forms (for instance, *suggest, suggests, suggested, suggesting, suggestion, suggestions*).

References

- Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins.
- Bowker, L. 2010. The Contribution of Corpus Linguistics to the Development of Specialised Dictionaries for Learners. Fuertes-Olivera, P. (Ed.). 2010. *Specialised Dictionaries for Learners*: 155-168. Berlin: De Gruyter.

- Brezina, V. and D. Gablasova.** 2013. Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics* 36(1): 1-22.
- Browne, C., B. Culligan and J. Phillips.** 2013a. *New General Service List Project*. <http://www.newgeneralservicelist.org/>. Accessed 1 September 2019.
- Browne, C., B. Culligan and J. Phillips.** 2013b. *New Academic World List*. <http://www.newgeneralservicelist.org/nawl-new-academic-word-list/>. Accessed 1 September 2019.
- Burkhanov, I.Y.** 1998. *Lexicography: A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo. Wyższej Szkoły Pedagogicznej w Rzeszowie.
- Campion, M.E. and W.B. Elley.** 1971. *An Academic Vocabulary List*. Wellington: New Zealand Council for Educational Research.
- Campoy Cubillo, M.C.** 2002. Dictionary Use and Dictionary Needs of ESP Students: An Experimental Approach. *International Journal of Lexicography* 15(3): 206-228.
- Chen, Q. and G. Ge.** 2007. A Corpus-based Lexical Study on Frequency and Distribution of Coxhead's AWL Word Families in Medical Research Articles. *English for Specific Purposes* 26: 502-514.
- Coxhead, A.** 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Coxhead, A.** 2018. *Vocabulary and English for Specific Purposes Research: Quantitative and Qualitative Perspectives*. London: Routledge.
- Coxhead, A. and D. Hirsch.** 2007. A Pilot Science-specific Word List. *Revue Française de Linguistique Appliquée* 12(2): 65-78.
- Dang, T.N.Y.** 2018. A Hard Science Spoken Word List. *ITL — International Journal of Applied Linguistics* 169(1): 44-71.
- Dang, T.N.Y., A. Coxhead and S. Webb.** 2017. The Academic Spoken Word List. *Language Learning* 67(4): 959-997.
- Dang, T.N.Y. and S. Webb.** 2016. Evaluating Lists of High-frequency Words. *ITL — International Journal of Applied Linguistics* 167(2): 132-158.
- Đurović, Z.** 2021. Corpus Linguistics Methods for Building ESP Word Lists, Glossaries and Dictionaries on the Example of a Marine Engineering Word List. *Lexikos* 31: 259-282.
- Esandi-Baztan, M.A. and P.A. Fuertes-Olivera.** 2020. Selecting an Initial Lemma List in Specialized Lexicography: A Case Study in the Field of Graphic Engineering. *Lexikos* 30: 57-89.
- Fraser, S.** 2007. Providing ESP Learners with the Vocabulary They Need: Corpora and the Creation of Specialized Word Lists. *Hiroshima Studies in Language and Language Education* 10: 127-143.
- Fuertes-Olivera, P.A. and A. Arribas-Baño.** 2008. *Pedagogical Specialised Lexicography: The Representation of Meaning in English and Spanish Business Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Fuertes-Olivera, P.A. and S. Nielsen.** 2011. Online Dictionaries for Assisting Translators of LSP Texts: The Accounting Dictionaries. *International Journal of Lexicography* 25(2): 191-215.
- Fuertes-Olivera, P.A., H. Bergenholtz, P. Gordo-Gómez, S. Nielsen, M. Niño-Amo, Á. de los Ríos, A. Sastre-Ruano and M. Velasco-Sacristán.** 2013. From Theory to Practice: The Selection of Spanish Lemmas in the Accounting Dictionaries. *Fachsprache: International Journal of Specialized Communication* 35(1-2): 25-42.
- Gardner, D. and M. Davies.** 2014. A New Academic Vocabulary List. *Applied Linguistics* 35(3): 305-327.
- Ghadessy, M.** 1979. Frequency Counts, Wordlists, and Material Preparation: A New Approach. *English Teaching Forum* 17(1): 24-27.
- Gilner, L.** 2011. A Primer on the General Service List. *Reading in a Foreign Language* 23(1): 65-83.

- Hernandez, M.M.** 2017. *Comparing the AWL and AVL in Textbooks from an Intensive English Program*. M.A. thesis. Provo, USA: Brigham Young University.
- Hsu, W.** 2013. Bridging the Vocabulary Gap for EFL Medical Undergraduates: The Establishment of a Medical Word List. *Language Teaching Research* 17(4): 454-484.
- Hsu, W.** 2014. Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates. *English for Specific Purposes* 33: 54-65.
- Jansson, H., S. Johansson Kokkinakis, J. Ribbeck and E. Sköldbberg.** 2012. A Swedish Academic Word List: Methods and Data. Vatvedt Fjeld, R. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress 7–11 August, 2012, Oslo, Norway*: 955-960. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Jin, N.Y. et al.** 2013. Development of the Engineering Technology Word List for Vocational Schools in Malaysia. *International Education Research* 1(1): 43-49.
- Johannessen, J.M., A. Saidi and K. Hagen.** 2016. Constructing a Norwegian Academic Wordlist. Calzolari, N. et al. (Eds.). 2016. *Proceedings of the Tenth International Conference on Language Resources and Evaluation, May 23–28, 2016, Portorož, Slovenia (LREC'16)*: 1457-1462. Portorož: European Language Resources Association.
- Kamrotov, M., E. Talalakina and D. Stukal.** 2022. Technical Vocabulary in Languages for Special Purposes: The Corpus-based Russian Economics Word List. *Lingua* 273: 103326.
- Khani, R. and K. Tazik.** 2013. Towards the Development of an Academic Word List for Applied Linguistics Research Articles. *RELC Journal* 44(2): 209-232.
- Khany, R. and B. Kalantari.** 2021. Accounting Academic Word List (AAWL): A Corpus-based Study. *Journal of Foreign Language Teaching and Translation Studies* 6(1): 35-58.
- Khumalo, L.** 2015. Semi-automatic Term Extraction for an isiZulu Linguistic Terms Dictionary. *Lexikos* 25: 495-506.
- Kokkinakis, S.J., E. Sköldbberg, B. Henriksen, K. Kinn and J.B. Johannessen.** 2012. Developing Academic Word Lists for Swedish, Norwegian and Danish — A Joint Research Project. Vatvedt Fjeld, R. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress 7–11 August, 2012, Oslo, Norway*: 563-569. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Konstantakis, N.** 2007. Creating a Business Word List for Teaching Business English. *ELIA* 7: 79-102.
- Kruse, T. and U. Heid.** 2021. Lemma Selection and Microstructure: Definitions and Semantic Relations of a Domain-Specific e-Dictionary of the Mathematical Field of Graph Theory. Gavriilidou, Z., M. Mitsiaki and A. Fliatouras (Eds.). 2021. *Proceedings of the XIX Euralex International Congress, Alexandroupolis, Greece, 7–11 September 2021. Volume 1*: 227-233. Komotini, Greece: European Association for Lexicography.
- Kwary, D.A. and A.F. Artha.** 2017. The Academic Article Word List for Social Sciences. *MEXTESOL* 41(4): 1-11.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. New York/Cambridge: Cambridge University Press.
- Lei, L. and D. Liu.** 2016. A New Medical Academic Word List: A Corpus-based Study with Enhanced Methodology. *English for Academic Purposes* 22: 42-53.
- Liu, J. and L. Han.** 2015. A Corpus-based Environmental Academic Word List Building and Its Validity Test. *English for Specific Purposes* 39: 1-11.
- Lynn, R.W.** 1973. Preparing Word Lists: A Suggested Method. *RELC Journal* 4: 25-32.

- Martínez, I.A., S.C. Beck and C.B. Panza.** 2009. Academic Vocabulary in Agriculture Research Articles: A Corpus-based Study. *English for Specific Purposes* 28: 183-198.
- McQuillan, J.** 2020. Harry Potter and the Prisoners of Vocabulary Instruction: Acquiring Academic Language at Hogwarts. *Reading in a Foreign Language* 32(2): 122-142.
- Mihindou, G.-R.** 2004. Some Features of Monolingual LSP Dictionaries. *Lexikos* 14: 118-136.
- Miller, L.** 2014. English for Science and Technology. Bhatia, V. and S. Bremner (Eds.). 2014. *The Routledge Handbook of Language and Professional Communication*: 332-468. Oxon/New York: Routledge.
- Minshall, D.E.** 2013. *A Computer Science Word List*. M.A. thesis. Swansea, Wales: Swansea University.
- Moini, R. and Z. Islamizadeh.** 2016. Do We Need Discipline-specific Academic Word Lists? Linguistics Academic Word List (LAWL). *Journal of Teaching Language Skills* 35(3): 65-90.
- Mudraya, O.** 2006. Engineering English: A Lexical Frequency Instructional Model. *English for Specific Purposes* 25(2): 235-256.
- Nagy, W.E. and R.C. Anderson.** 1984. How Many Words are There in Printed School English? *Reading Research Quarterly*: 304-330.
- Nation, I.S.P.** 2013. *Learning Vocabulary in Another Language*. Second edition. Cambridge: Cambridge University Press.
- Newman, J.A.** 2016. *A Corpus-based Comparison of the Academic Word List and the Academic Vocabulary List*. M.A. thesis. Provo, USA: Brigham Young University.
- Nielsen, S.** 2018. LSP Lexicography and Typology of Specialised Dictionaries. Humbley, J., G. Budin and C. Laurén (Eds.). 2018. *Languages for Special Purposes: An International Handbook*: 71-95. Berlin/Boston: De Gruyter.
- Nkomo, D.** 2008. *Towards a Theoretical Model for LSP Lexicography in Ndebele with Special Reference to a Dictionary of Linguistic and Literary Terms*. M.A. thesis. Stellenbosch: Stellenbosch University.
- Praninskas, J.** 1972. *American University Word List*. London: Longman.
- Ribeck, J., H. Jansson and E. Sköldberg.** 2014. Från aspekt till övergripande — en ordlista över svensk akademisk vokabulär. *Nordiske studier i leksikografi* 12: 370-384.
- Rundell, M. and A. Kilgarriff.** 2011. Automating the Creation of Dictionaries: Where Will it All End? Meunier, F. et al. (Eds.). 2011. *A Taste for Corpora. In Honour of Sylviane Granger*: 257-282. Amsterdam/Philadelphia: John Benjamins.
- Savický, P. and J. Hlaváčová.** 2002. Measures of Word Commonness. *Journal of Quantitative Linguistics* 9(3): 215-231.
- Talalakina, E., D. Stukal and M. Kamrotov.** 2020. Developing and Validating an Academic Vocabulary List in Russian: A Computational Approach. *Modern Language Journal* 104(3): 618-646.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Berlin/New York: Max Niemeyer.
- Tarp, S.** 2010. Functions of Specialised Learners' Dictionaries. Fuertes-Olivera, P. (Ed.). 2010. *Specialised Dictionaries for Learners*: 39-53. Berlin/New York: De Gruyter.
- Todd, R.W.** 2017. An Opaque Engineering Word List: Which Words Should a Teacher Focus on? *English for Specific Purposes* 45: 31-39.
- Tongpoon-Patanasorn, A.** 2018. Developing a Frequent Technical Words List for Finance: A Hybrid Approach. *English for Specific Purposes* 51: 45-54.

- Valipouri, L. and H. Nassaji.** 2013. A Corpus-based Study of Academic Vocabulary in Chemistry Research Articles. *Journal of English for Academic Purposes* 12(4): 248-263.
- Verlinde, S. and T. Selva.** 2001. Corpus-based Versus Intuition-based Lexicography: Defining a Word List for a French Learners' Dictionary. Rayson, P., A. Wilson, T. McEnery, A. Hardie and S. Khoja (Eds.). 2001. *Proceedings of the Corpus Linguistics 2001 Conference*: 594-598. Lancaster: University Centre for Computer Corpus Research on Language, Lancaster University.
- Vongpumivitch, V., J.-Y. Huang and Y.-C. Chang.** 2009. Frequency Analysis of the Words in the Academic Word List (AWL) and Non-AWL Content Words in Applied Linguistics Research Papers. *English for Specific Purposes* 28: 33-41.
- Wang, J., S. Liang and G. Ge.** 2008. Establishment of a Medical Academic Word List. *English for Specific Purposes* 27(4): 442-458.
- Ward, J.** 2009. A Basic Engineering English Word List for Less Proficient Foundation Engineering Undergraduates. *English for Specific Purposes* 28(3): 170-182.
- West, M.** 1953. *A General Service List of English Words*. London: Longman, Green & Co.
- Xue, G. and P. Nation.** 1984. A University Wordlist. *Language Learning and Communication* 3(2): 215-229.
- Xue, M. and S. Tarp.** 2018. Towards Chinese Learner's Dictionaries for Foreigners Living in China: Some Problems Related to Lemma Selection. *Lexikos* 28: 384-404.
- Yang, M.-N.** 2015. A Nursing Academic Word List. *English for Specific Purposes* 37: 27-38.

Appendix

Author	List	Number of words	Corpus/ Coverage in the corpus	Methodology
Mudraya (2006)	Engineering English word list	1,260 word families	2 mill. (13 text-books) / not given	— frequency of the word of at least 100
Coxhead and Hirsh (2007)	Pilot science list	318 word families	1,761,380 words (14 disciplines) / 3.79%	— exclusion of the GSL & the AWL words — occurrence in at least half the disciplines — frequency of at least 50 in the corpus — the dispersion factor of at least 35 — exclusion of proper nouns, symbols and abbreviations
Fraser (2007)	Pharmacology word list (PWI)	601 word families	185,000 (51 research articles (RAs))	— exclusion of the GSL and the AWL words — exclusion of proper names, nationalities, numbers, abbreviations and acronyms — occurrence in at least 6 RAs — frequency of 10 or more and occurrence in at least 2 RAs
Konstantakis (2007)	Business word list for undergraduates (BWL)	560 word families	600,000 (Published Materials Corpus, compiled by Nelson (2000); 33 course books) / 2.79%	— exclusion of the GSL and the AWL words — occurrence in at least 5 course books — exclusion of proper names, numerals, Latin words, nationalities, acronyms, interjections and abbreviations — frequency of at least 10 in the corpus
Wang, Liang and Ge (2008)	Medical academic word list (MAWL)	623 word families	1,093,000 words (288 research articles from 96 journals) / 12.24%	— exclusion of the GSL words — occurrence in at least half the disciplines — frequency of at least 30 in the corpus — two English for Medicine professors consulted for differentiating between technical and academic vocabulary
Ward (2009)	Basic Engineering list (BEL)	299 word types	271,000 words (25 textbooks in engineering) / 16.4%	— exclusion of function words — definition of word as a word type — frequency of at least 5 in each of the 5 subsections
Martínez, Beck and Panza (2009)	A reduced AWL for agriculture	92 word families	826,416 (218 RAs) / coverage not given	— inclusion of the academic words above the mean for academic words

Vongpumi-vitch, Huang and Chang (2009)	Applied linguistics word list	475 AWL word forms and 128 non-AWL content word forms	1.5 mill. (200 RAs from 5 journals) / 8.6% for the 475 AWL words + 2.8% for the non-AWL words	<ul style="list-style-type: none"> — frequency of at least 50 in the corpus — occurrence of at least 5 times in at least 5 journals — exclusion of the GSL, function words and abbreviations
Jin et al. (2013)	Engineering technology word list (ETWL)	313 word lists	124,584 words (2 textbooks) / 8.7%	<ul style="list-style-type: none"> — exclusion of the GSL and the AWL words — to be defined as technical, a word had to appear as an entry in an online engineering dictionary — cross-checked by two experts from the field to make sure the words were technical and semi-technical
Khani and Tazik (2013)	Applied linguistics academic word list	773 word families	1,553,450 words (240 RAs from 12 journals) / 12.48%	<ul style="list-style-type: none"> — exclusion of the GSL words — frequency of at least 50 in the corpus — occurrence of at least 4 times in at least half the journals
Valipouri and Nassaji (2013)	Chemistry Academic word list (CAWL)	1,400 word families	4 mill. (1,185 RAs from 38 journals) / 81.18%	<ul style="list-style-type: none"> — frequency of at least 114 in the corpus — frequency of at least 10 in all subsections — exclusion of abbreviations and function words — three chemistry professors excluded technical words using a rating scale
Hsu (2013)	Medical word list (MWL)	595 word families	15 mill. (155 medical textbooks) / 10.72%	<ul style="list-style-type: none"> — exclusion of the BNC 3000 — occurrence in more than half the subsections — frequency of at least 863 in the corpus
Minshall (2013)	Computer science word list	433 word families	3,661,337 tokens (RAs and conference proceedings from 10 subdisciplines) / 6%	<ul style="list-style-type: none"> — outside the GSL and the AWL — occurrence in at least half the sub-disciplines — occurrence of at least 80 in the corpus
Hsu (2014)	Engineering English word list (EEWL)	729 word families	4.57 mill. (100 engineering textbooks) / 14.3%	<ul style="list-style-type: none"> — exclusion of the BNC 2000 — occurrence in all subsections — occurrence in at least 95 out 100 textbooks — frequency of at least 288 times in the corpus — exclusion of exclamations, interjections and proper names
Liu and Han (2015)	Environmental academic word list (EAWL)	458 word families	862,242 tokens (200 RAs) / 15.43%	<ul style="list-style-type: none"> — exclusion of the GSL words — frequency of at least 30 in the corpus — occurrence in at least 8 out 10 subsections

Yang (2015)	Nursing academic word list	676 word families	1,006,934 words (252 RAs) / 13.64%	<ul style="list-style-type: none"> — exclusion of the GSL words range — occurrence in at least half the subsections — frequency of at least 33 in the corpus
Lei and Liu (2016)	Medical Academic Vocabulary List (MAVL)	819 lemmas	6.2 mill. (760 medical RAs from 38 journals (MAEC corpus + 1 three-volume textbook (MAET corpus)), 19.44% in MAEC and 20.18% in MAET	<ul style="list-style-type: none"> — frequency of 28.57 per 1 mill — ratio of at least 1.5 (at least 50% higher frequency in the academic corpus than in a non-academic corpus) — occurrence of 20% of the expected frequency in at least half the subsections — dispersion of at least 0.5 (Jullian's D) — no lemma should occur more than 3 times the expected frequency in more than any 3 out 21 subsections — special meaning criterion checked via 2 medical dictionaries
Moini and Islamizadeh (2016)	Linguistics academic word list	1,263 word families	4 mill. (700 RAs from 4 subdisciplines) / not given	<ul style="list-style-type: none"> — a frequency of at least 114 in the corpus — occurrence of at least 10 times in each subdiscipline
Todd (2017)	Opaque engineering word list	186 items	1.15 mill. (27 textbooks) / not given	<ul style="list-style-type: none"> — exclusion of non-words, abbreviations, and function words — occurrence in at least 15 out 27 textbooks — similar word types combined under one entry
Kwary and Artha (2017)	Academic article word list for social sciences	350 word families	1,040,259 tokens (122 RAs)	<ul style="list-style-type: none"> — outside the GSL — occurrence in all the 6 subdisciplines — frequency
Dang (2018)	Hard science spoken word list	1,595 word families	6.5 mill. words of spoken language from 12 disciplines	<ul style="list-style-type: none"> — occurrence in at least half the disciplines and both subsections of the corpus — frequency of at least 175 in the corpus — dispersion (DP value below 0.6)
Tongpoon-Patanasorn (2018)	Frequent technical word list for finance	569 word families	2 mill. words (books, journals, websites and newspapers)	<ul style="list-style-type: none"> — keyness — rating scale, completed by experts
Khany and Kalantari (2021)	Accounting academic word list	658 word families	29.5 mill. words (2,098 accounting RAs)	<ul style="list-style-type: none"> — outside the GSL — frequency of at least 839 in the corpus and at least 50 in each journal
Kamrotov, Talalakina and Stukal (2022)	Russian economics word list	507 lemmas	10.5 million words (economics RAs and mass media economics articles)	<ul style="list-style-type: none"> — keyness (1.5 more frequent in the economics corpus than in other corpora) — degree of dispersion over 0.25 — minimum frequency of 10 in the corpus

An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years

Esra M. Abdelzaher, *University of Debrecen, Doctoral School of Linguistics;
and University of Debrecen, Institute of English and American Studies,
Hungary (esra.abdelzaher@gmail.com)*

Abstract: This study explores the role of corpus linguistics in addressing lexicographic challenges over the past ten years. It examines 25 studies published from 2012 to 2021 that employed corpus solutions to solve lexicographic problems. Challenging tasks are either relevant to the macrostructure or the microstructure of dictionaries. In the past decade, lexicographers made extensive use of corpus tools to create dictionaries and improve existing ones. This effort included compiling general and specialized headword lists, identifying idiom variations, detecting collocational patterns, identifying and ordering words senses and differentiating polysemous words and near-synonyms. Frequency, keyword extraction, and word sketches are among the most effective aids for lexicographers. According to the investigated studies, almost all dictionaries may benefit from corpus tools at the macro and microstructure levels.

Keywords: ACCESS STRUCTURE, CORPUS TOOLS, COLLOCATIONAL ANALYSIS, LEXICOGRAPHIC CHALLENGES, MICROSTRUCTURE, MACROSTRUCTURE, SENSE DELINEATION, SPECIALIZED DICTIONARIES, TERM EXTRACTION

Opsomming: 'n Ondersoek na die bydrae wat korpuse oor die afgelope tien jaar tot leksikografiese uitdagings gemaak het. Hierdie studie ondersoek die rol van die korpuslinguistiek in die benadering tot leksikografiese uitdagings oor die afgelope tien jaar. Dit ondersoek 25 studies wat vanaf 2012 tot 2021 gepubliseer is. Uitdagende take is óf relevant tot die makrostruktuur óf tot die mikrostruktuur van woordeboeke. In die afgelope dekade het leksikograwe uitgebreid gebruik gemaak van korpuseredskap om woordeboeke te skep en bestaandes te verbeter. Dit het die samestelling van algemene en gespesialiseerde lemmalyste, die identifisering van idioomvariasies, die bepaling van kollokasionele patrone, die identifisering en ordening van woordbetekenisse en die onderskeiding van polisemiese woorde en ampersinomieme ingesluit. Frekwensie, sleutelwoordonttrekking en woordbeskrywings is van die mees effektiewe hulpmiddels vir leksikograwe. Volgens die studies wat ondersoek is, kan byna alle woordeboeke baat by korpuseredskap op makro- sowel as mikrostruktuurvlakke.

Sleutelwoorde: TOEGANGSTRUKTUUR, KORPUSGEREEDSKAP, KOLLOKASIONELE ANALISE, LEKSIKOGRAFIESE UITDAGINGS, MIKROSTRUKTUUR, MAKROSTRUKTUUR, BETEKENISOMSKRYWING, GESPELISEERDE WOORDEBOEKE, TERM-ONTTREKING

1. Introduction

The field of lexicography has witnessed significant changes over the years. In his description of the significant influences in the field over 2000 years, Hanks (2013) referred to computers and corpora as the second most influential factor, after the invention of printing, in lexicography. The new computational and corpus technology provided both lexicographers and users with innovative tools to compile and consult dictionaries.

The significant contributions of corpus tools in lexicography were first and most salient in the context of Monolingual Learner's Dictionaries (MLDs). As Sinclair (1992) stated, *Collins COBUILD English Language Dictionary* was the first to rely on corpus evidence in the compilation process. The compilers aimed to improve the understandability of the dictionary and provide a representative picture of modern English. Corpus evidence facilitated the inclusion of common senses and the exclusion of obsolete word forms and word senses. Also, the frequency of use was the main criterion in organizing information in the entry. The dictionary replaced the authoritative artificial examples with corpus-based examples of word use. It used complete actual sentences to explain the meaning and the typical use of a word.

The innovative corpus-driven approaches in the multiple editions of *COBUILD* were further discussed by lexicographers. For instance, Heuberger (2018) pointed to presenting the frequency of words as an initiative launched in the second edition of *COBUILD*. This initiative helped learners identify the words they should remember because they are likely to encounter them. Unlike most dictionaries, the online version of *COBUILD* presents frequency information for a large number of words. Moreover, it offers information about the frequency of using the word over the years. Distributional and frequency information is one of the major advantages of corpus methods.

Nowadays, the role of corpus tools is almost conventional in dictionary-making. The continuous update of corpora and their instant accessibility made them unmatched valuable resources for dictionary-makers (Abecassis 2007). Basically, corpus-analysis helped lexicographers take several decisions about the macro- and microstructure of dictionaries. The macrostructure of a dictionary refers to all entries of the lemmas in the wordlist. At the macrostructure level, lexicographers have to decide which words will be included in the dictionary based on the purpose and the target user. They may include or exclude specialized terms, jargon or loan words. Also, they should decide upon the lexical items that will be granted headword status, e.g., lemmas, inflected words or phrases (Atkins 2008). In this regard, corpus-based frequency, user needs and dictionary types have a significant role in the choice of words for lexicographical treatment. Although frequency lists are now essential to compile wordlists for dictionaries, user needs and dictionary types have more significant roles. For instance, a synonymy dictionary would discard a frequent word if it does not have a synonym (Bogaards 2013).

The microstructure level is form and sense-related. It is relevant to the

selection and presentation of information in the entry. Lexicographers may choose to include or discard specific pieces of etymological, external and internal information about a word. Internal word information is relevant to its morphology, orthography, semantics and phonological features. Internal information embraces both facts relevant to the word's form and sense. External information refers to the relation between a word and other words. It includes paradigmatic relations (e.g., POS and synonyms), syntagmatic relations (e.g., collocations), relational links (e.g., cross-references to derivational forms) and usage information (e.g., genre, dialect). Lexicographers need to find senses, split or lump and order them, and find lexicographically relevant information for each sense. They also have to select elaborative examples for each sense (Atkins 2008, Atkins and Rundell 2008).

There are several online and offline tools to conduct corpus analysis. The Sketch Engine software is probably the most effective and widely used by lexicographers. The introduction of word sketches presented lexicographers with sorted lexico-grammatical collocations of the target. Word sketches effectively summarize thousands of concordance lines and provide lexicographers with a report of the word's behavior (Heuberger 2018). This feature was efficiently employed in the compilation of *Macmillan English Dictionary for Advanced Learners* (MEDAL). In most cases, lexicographers could link a specific word sense to a particular collocational pattern (Hanks 2013).

In addition, the software has a random sampling option to extract a sample of the concordance for analysis. It also hosts a large number of corpora in several languages as well as the option of uploading the user's own corpus. The software has recently introduced an option to automatically select examples that better serve the purposes of lexicographers.

Offline tools include ANTConc which can be installed on computer devices and used offline. It processes files stored on the device and offers a variety of content analysis options. In addition to the basic frequency list compilation and n-gram clusters, Key Words In Context (KWIC) can be displayed in different modes through ANTConc. It also has several statistical methods for detecting collocations (Faaß 2018). WordSmith Tools offer, in addition to the standard content analysis options, part of speech concordancing for tagged corpora, manual lemmatizing of wordlists and plotting keywords. Relevantly, Yoshikoder is unique for its dictionary options. This offline software allows the search for nested lists of words collectively or individually. Users can add a dictionary that includes words referring to animals, for instance, subcategorize them into domestic and predators and search for the entire dictionary or the part relevant to domestic animals only. There are no limits on the number of added dictionaries or their sub-categories.

Although corpus tools have revolutionized the field of lexicography, some challenging tasks remain hard to address. Kilgarriff (1998) assessed the most difficult choices lexicographers had to make during the compilation of the *Longman Dictionary of Contemporary English*. In his short report, he considered the most complicated tasks the ones that lack clear rules or guidelines. He categorized

the tasks into analysis, i.e., tasks relevant to the pre-writing stage and the analysis of word behavior in context, and synthesis, i.e., tasks relevant to the content that will be finally presented in the dictionary. The most challenging task during the analysis stage was "splitting; identifying senses of a word" (Kilgarriff 1998: 53). There are no instructions or in-dictionary style guides on how to split or lump senses in an entry. Also, at that time, there was no valuable information on how to deal with this problem in books. Therefore, lexicographers used to, and perhaps still, depend on their experience and intuition. Accordingly, the task ranked second in the list of the most challenging tasks after choosing the words of a definition. The problematic tasks consist of the inclusion of headwords, treatment of Multiple Word Expressions (MWEs), selection or invention of examples, and dealing with grammar and inflected words, among others.

This study is concerned with the lexicographic tasks to which the contribution of corpora and corpus tools is most significant. It offers an investigation of the lexicographic studies that highlighted the role of the corpus in solving macro- and microstructure problems in the last decade. It addresses the following questions. In the past ten years,

1. What are the lexicographic challenges that were successfully addressed by corpus tools?
2. Which corpus options are highly contributing to improving the lexicographic practice?
3. Which types of dictionaries did benefit from corpus tools?

2. Methodology

The literature on the use of corpus tools in lexicography is voluminous. The present investigation used the flow diagram of PRISMA 2020 to search the literature in order to answer the previously mentioned questions. The PRISMA 2020 statement provides detailed guidelines for scholars aiming at investigating the literature to find answers for certain questions. It clarifies that scholars should clearly state the sources of data, the search terms and the date on which the databases were accessed. Scholars are required to precisely report the number of the retrieved data items (reflected in the identification step in the diagram) and should clarify if all of the results were examined to reach the set objective or not (screening step). After that, the selected articles should be examined according to the eligibility criteria (i.e., inclusion and exclusion) that best suit each investigation so that the inclusion of the selected articles can be justified (Page et al. 2021).

2.1 Search Procedure

The following databases were the primary sources of data: PubMed, Science Direct, Web of Science, Scopus, Mendeley, ERIH PLUS, PsycINFO, ProQuest, and Crossref. The searched keywords were "Corpus", AND "lexicograph*", OR "lexis" OR "lexicon*"AND "dictionar*". The search excluded "NLP" AND/OR

"survey" OR "computation*."

The time range was customized from 2012 to 2021. 358 articles were found. Using Mendeley software, the duplicated articles were deleted. Also, non-English results, book chapters and articles published before the specified date were filtered manually. The unique articles were 98. After excluding the author-focused articles, surveys and NLP-oriented articles, the total number of research articles became 73. Screening the titles and abstracts of all articles excluded studies that presented only theoretical argument, mentioned only lexicographic implications and discussed only the role of corpus-based dictionaries in pedagogical contexts. The screened articles amounted to 53. After reading the full manuscripts to be vetted for eligibility, articles that addressed the role of the corpus with no relevance to a lexicographic problem and articles that provided no analytic examples were excluded. The accepted analytical examples had to involve corpora or corpus tools with relevance to a lexicographic task. The articles that described the construction of a corpus tool without defining a particular lexicographic challenge or used audiovisual corpora are considered out of the scope of this study. The total number of the included article became 25 (Figure 1).

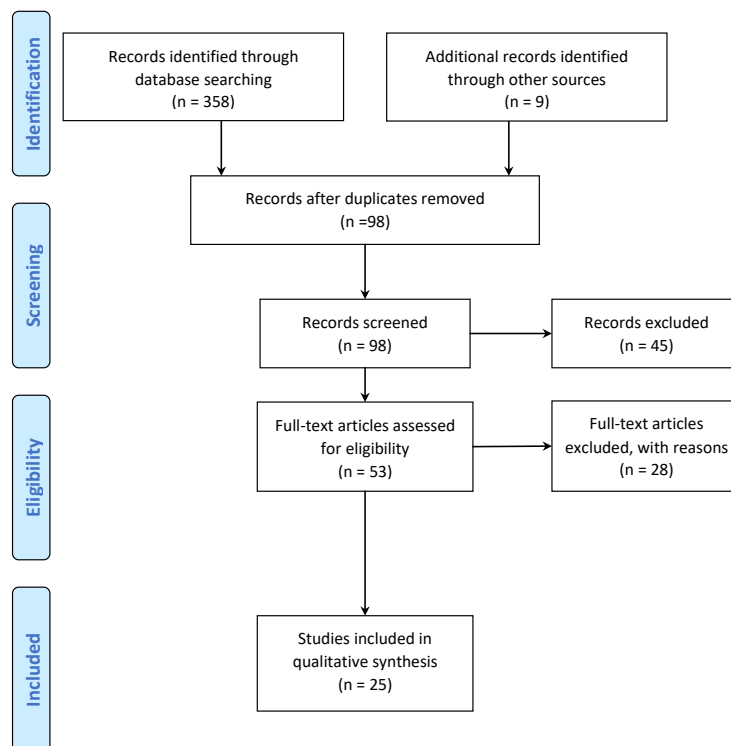


Figure 1: PRISMA flow chart showing the search method

2.2 Eligibility criteria

2.2.1 Criteria of Inclusion

The inclusion criteria for this study were as follows:

1. Original articles that integrated corpus analysis to address a lexicographic task or challenge.
2. The included studies must analyze at least a single case study relevant to a particular lexicographic challenge.
3. All studies must offer a corpus-based solution, either primary or secondary.
4. All studies must use a lexicographic resource, either primary or secondary.

2.2.2 Criteria of Exclusion

The exclusion criteria for the present study were as follows:

1. Book reviews, review articles, book chapters
2. NLP-oriented research, e.g., report on corpus-tools or algorithms
3. Theoretical arguments only
4. Articles that are written in a non-English language
5. Non-textual corpora

3. Results

Only four out of the 25 analyzed studies used corpus tools to address a task at the macrostructure level, whereas 12 studies were mainly concerned with challenges at the microstructure level. In addition, ten studies used corpus tools to take decisions at both macro- and microstructure levels.

At the macrostructure level, lemma-related problems show considerable diversity. The use of corpus in the compilation of a lemma list as the first step towards a representative headword list is central to the macrostructure tasks. This task, in some cases, was further complicated by the type of language the lexicographer needs a lemma list for. For English, the task is less challenging because of the availability of a large tagged corpus and the poor-morphological system of the language. However, for endangered, less-resourced and agglutinative languages, the compilation of a lemma list is problematic. It usually depends on corpus-driven analysis that retrieves frequency wordlists and complements the process with a corpus-based analysis of the frequency and distribution of other words cited from previous dictionaries or suggested by lexicographers.

At the microstructure level, concordances (especially KWIC), word sketches,

collocations and frequency were the corpus options employed by scholars to tackle sense-related tasks. The tools were proved to be effective in detecting and describing collocational patterns, identifying, ordering and describing senses, detecting polysemy and differentiating near-synonyms. Nevertheless, this corpus-based analysis was usually triangulated with qualitative and often theoretical analyses. Theories like Frame Semantics, Conceptual Metaphor Theory, Norms and Exploitations were usually present to motivate this corpus-based analysis in order to obtain the best results. These tools also improved the entries of colloquial and regional words in dictionaries.

Moreover, several studies focused on a linguistic aspect or phenomenon and adopted a corpus approach to detect, include and record it in dictionaries. Therefore, they dealt with challenges at the macro- and microstructure levels. For instance, the identification of specialized terms, through frequency and term extraction, and including them in dictionaries should be relevant only to the macrostructure of a dictionary. However, conducting a collocational analysis and relying on corpus citations of construct entries for specialized terms is associated with the microstructure. The same applies to the selection, inclusion and description of MWEs and phrases.

All of the analyzed studies used at least one lexicographic resource at the analysis stage. In this regard, almost all types of dictionaries were used either as the main project reported in the study (13 studies) or as a reference point for comparison with the corpus-based results (12 studies). The studies focused on monolingual, bilingual and multilingual dictionaries that mostly covered Germanic, Romance and Slavic languages. The typological spread of dictionaries included native, learners, specialized, general language, adult and children dictionaries. In addition to traditional dictionaries, lexical databases and innovative dictionaries such as the *Pattern Dictionary of English Verbs* (PDEV) were used (Table 1 in the Appendix).

4. Discussion

From a psycholinguistic perspective, dictionary makers attempt to reflect the mental lexicon of a native speaker so that learners would benefit the most from dictionary entries, i.e., lemmas and senses (Anshen and Aronoff 1999, Atkins and Rundell 2008, Fillmore and Atkins 1992, Jorgensen 1990). However, compiling a headword list that is comprehensive is an unattainable goal given the varieties and dynamicity of natural languages (Atkins and Rundell 2008, Lew 2013). Therefore, lexicographers aim at collecting the most frequent words in a language. They assume that the most frequent words represent the core vocabulary in the lexicon and speakers are most likely to encounter them in everyday life.

4.1 Frequency lists and compiling wordlists for general language dictionaries

Collins COBUILD pioneered the use of corpus frequency to compile wordlists

for monolingual dictionaries. Sinclair (1992) explained that the dictionary used a star-rated system to indicate the frequency of a headword in spoken and written English (according to corpus frequency). The online version of the dictionary still refers to the frequency of words using a single or five stars for rarest and most frequent words respectively. After this successful attempt, corpora have been widely used to compile dictionary wordlists.

De Schryver and Nabirye (2018a) displayed their successful use of corpora in building a dictionary for Bantu languages and used Lusoga as a case study. The standard case for lexicographers is generating a lemmatized wordlist based on a tagged general reference corpus. For Lusoga and less-resourced languages, tagged corpora and automatic lemmatizers are hardly available. Therefore, the author(s) had to depend only on a corpus-driven frequency list and apply several steps to convert it into a lemmatized one. They used *WordSmith Tools* to retrieve a frequency wordlist from a 1.7-million-word corpus of Lusoga along with their distributions. The list contained 200,000 types that were further processed to obtain a headword list for the dictionary. The author(s) had to rely on frequency again to shorten the wordlist into the most frequent 10,000 words. This step kept all words that had 12 or more instances in the corpus, which are arguably the most common words in Lusoga. After that, the authors used *TLex* to annotate the shortlist according to the POS and the lemma.

Unlike Lusoga, creating a lemma list for Swahili was less problematic given the availability of a larger and tagged corpus. Wójtowicz (2016) described the role corpus played in constructing a learner Swahili–Polish dictionary that was, unlike Swahili dictionaries, not based on previous dictionaries. The lemmas and the dictionary entries were based on the Helsinki Corpus of Swahili, the largest for an African language. The corpus is tagged, and it contains more than 12 million tokens of Standard Swahili. The headword list consists of 10,000 types, including nouns, verbs, adjectives, adverbs and MWEs. It was mainly derived from the corpus and expanded with words from pedagogical Swahili resources.

Similarly, Spence (2021) aimed at using the Hupa corpus as an integral part of compiling the Hupa–English dictionary. However, the task was far more complicated for this endangered and lesser-resourced language. Given its productive polysynthetic morphology, the lexicon of Hupa is more dynamic than most languages. The Hupa lexicon is not a list of content and functional words that are subject to neologism. Instead, it is a set of rules according to which morphological forms combine to form new words. Given this language-specificity, corpus provides authentic examples of the productive processes of word formation and for the created words. However, the available Hupa corpus consists of approximately 55,000 words which is a very small size. Although the corpus consists of the speech of 20 speakers talking about personal experiences, cultural practices and traditional stories, its size is too small to allow effective corpus-driven analysis. On the one hand, it is impossible to include the infinite number of words that can be formed according to the Hupa rules. On the other hand, the Hupa corpus is too small to derive a representa-

tive frequency-based wordlist. Therefore, the project had to rely on a previous print dictionary and complement the wordlist with words appearing in the Hupa corpus. The corpus helped in creating new dictionary entries and enriching existing ones which were originally based on the print dictionary. It was a solution to bridge the wide gap between the infinite possible words and the ones that have already been produced by Hupa speakers.

Prinsloo (2015) discussed how the size of the corpus might affect the coverage of the lemma list, especially for less-resourced languages. It mainly focused on Afrikaans and Sepedi. The study compared the overlap between frequency lists obtained from one-million and ten-million tokens of the same corpora. There was 77.5% and 72.8% overlap between the two lists for Afrikaans and Sepedi, respectively. However, the loss was significant as it missed words that are part of the core vocabulary, e.g., *seyalemoya/radio* and *kamano/relationship*. Also, the coverage of idioms was influenced by the size of the corpus. Although essential idioms were present in the one-million corpus, they were not frequent enough to be detected by a lexicographer.

4.2 Keyword extraction and compiling wordlists for specialized dictionaries

Specialized dictionaries may target a specific language phenomenon such as collocations or idioms or specific terminology such as legal or medical terms. For instance, Pimentel (2013) used a term extractor (which is somehow equivalent to keyword extraction) to identify legal terms from a specific corpus of legal judgments called JuriDiCo. The extractor automatically identified words that are specific to JuriDiCo if compared to a general reference corpus. The term extraction function also enabled Boz et al. (2018) to select terms relevant to Turkish lexicography from their specialized corpus of dissertations, textbooks and research articles relevant to lexicography. In the same vein, Cabezas-García and Faber (2018) created an 8-million-word specialized corpus from environmental and eco texts. Then, they used a term extractor in order to identify words that are specific to the ecological discourse automatically.

Aiming at different target users, Wild, Kilgarriff and Tugwell (2013) explained the role of corpus analysis in compiling children's dictionaries. They clarified the specificity of the discourse directed to children and argued that children dictionaries should not be merely simplified versions of adult dictionaries. They compared the keywords of *Oxford Children's Corpus* (30 million tokens) with the *Oxford English Corpus* to make decisions on the headword list for the children's dictionary. The classification of the key lemmas in the two corpora showed the discrepancy in the themes presented for adults and children. For instance, language related to the physical world was highly present in the children corpus, whereas politics and religion were characteristic of the adult corpus. Therefore, the study advocated the use of specific children's corpus in compiling children's dictionaries in order to refine the headword lists.

Specialized dictionaries can make partial use of corpus tools too. For instance, Gizatova (2018) relied on frequency and collocations to construct a bilingual dictionary of idioms. The list of idioms was not corpus-based; it was based on previous dictionaries of idioms. However, the frequency of the idiom in the corpus was the main criterion for keeping or excluding it from the constructed dictionary (40% of dictionary-based idioms were excluded). Also, collocational analysis of the idioms detected the variant forms of an idiom and its semantic preferences. Moreover, analyzing the occurrences of the idioms in parallel corpus revealed dissimilarities between the supposedly equivalent idioms in English and Russian. It displayed the contexts in which the idioms and their dictionary-recommended translations are not equivalent.

4.3 Collocations/concordance and finding senses and structuring dictionary entries

The challenge of sense delineation is common between lexical semantics and lexicography. Over time, linguists and lexicographers started to adopt conflicting views on meaning. To elaborate, in the traditional view, words are believed to have several types of meaning, such as lexical and contextual meaning. Lexical meaning is the semantic content of the word regardless of the contexts in which it may be used. In contrast, contextual meaning arises when the word is used in real communicative situations. On the one hand, some linguists rejected the possibility of having a word meaning outside the context of use (cf. Kilgarriff 1992; Hanks 2004). On the other hand, linguists such as Louw (1995) rejected the existence of contextual meaning as part of the word meaning. He equated any context-based interpretation of a word with word use and called for recording only the lexical meaning of a word in a dictionary (Bergenholtz and Gouws 2017).

Although corpus tools present numerous authentic word usages, converting corpus citations into an organized list of senses that appeal to dictionary users is a laborious lexicographic task. The process of discovering senses from corpus citations does not follow a conventional method (Lew 2013). For instance, Kilgarriff (2005) proposed a model that aimed at putting corpus into dictionaries (PCID). The model relied on collocate-to-sense mapping but added a grammatical dimension to the collocational relation (word sketch collocations which are now conventionally used in lexicographic practice). In addition, Hanks (2004) revealed through his *Corpus Pattern Analysis* (CPA) how the meanings of a word could be mapped to patterns of usage. He adopted a corpus-driven approach based on the *Theory of Norms and Exploitations* (TNE) in order to examine word meanings in contexts instead of assuming the existence of meaning in isolation from the context. His project required a massive lexicographic effort to process word uses, find and record usage patterns and associate each pattern with a meaning.

In this regard, the discussed studies relied heavily on word sketches and collocational patterns to identify, split, lump and order senses in dictionary

entries. They usually combined corpus-based analysis with theoretical analysis. Dalpanagioti (2018) conducted a frame-based analysis to detect polysemy in a corpus through the co-occurrence patterns. The paper provided the verb *fly* as a case study. Word sketches detected the verb's co-occurrence patterns which were further analyzed (in a sample concordance) according to the principles of the Frame Semantics theory. The analysis of the co-occurrence pattern of *fly* and *flag*, *kite* and *banner* showed that the collocates fill in the theme semantic role (e.g., *the flag flew from the castle's topmost tower*). The meaning of *fly* in this pattern indicates that the theme is fixed at one point while moving in the air. This sense evokes the frame of "Moving in place". When this pattern changes to involve a human, building or a vessel in the agent position (e.g., *the ship was flying a quarantine flag*), another sense is identified. In the second pattern, the meaning of *fly* denotes "raise a flag and make it float in the air". This sense evokes the frame of "Cause motion" (Dalpanagioti 2018: 14). That is to say; polysemy is detected if the word is used in different patterns that represent different arguments (i.e., frame elements which are frame-specific semantic roles). Accordingly, the senses are split into different frames.

Smirnova (2021) used collocational patterns as sense distinguishers and polysemy detectors in another context. The study did not rely on a theoretical linguistic background. It was rather motivated by the literature on psychology which argued that "vastness and accommodation" are typical features of this psychological experience whereas "threat", "beauty", "ability", "virtue" and "supernatural" are non-central to it (Smirnova 2021: 231). Two binary variants (i.e., positive and negative) of *awe* have been the focus of psychological studies. The study analyzed a sample concordance of the noun *awe* cited from a 14-billion-word corpus. The collocations of the target word and the concordance helped the scholar define the multiple senses of *awe* and the evaluative attitude of the experiences expressed by the different uses of the word. Although the results confirmed vastness as a core feature of *awe*, it suggested distinguishing between transcendental *awe* (linguistically reflected in collocates such as *God* or *supernatural*) and mundane *awe* (linguistically reflected in collocates such as *landscape* or *technology*). Moreover, the study argued for classifying the variants of *awe* into transcendental ambivalent, mundane ambivalent, transcendental positive, mundane positive, transcendental negative and mundane negative based on the collocational analysis.

Similarly, De Schryver and Nabirye (2018b) performed manual annotation of a sample concordance of the verb *-v-* in Lusoga. They were able to map the different usage patterns to meaning potential and construct two entries for the verb accordingly. Senses were organized according to their frequency in the analyzed sample.

4.4 Collocations/concordance and accounting for language varieties

From a sociolinguistic perspective, a dictionary should reflect a representative

picture of the language and cover its varieties (Dolezal 2020). Whereas scholars such as Siepmann (2015) argued for allowing a better description of the colloquial variety and reliance on spoken corpus in dictionaries (for French), other scholars such as Xia et al. (2016) called for the inclusion of varieties that are specific to the non-native speakers of the language.

Adopting the perspective of "world Englishes", Xia et al. (2016) argued for the inclusion of "China English" in learner's dictionaries. They considered it part of the core vocabulary of English as it represents a variety spoken by an expanding number of users and has its own characteristics. First, they criticized current English dictionaries for the marginal inclusion of China English words. Although they acknowledge frequency as a criterion for inclusion, they called for using English corpora produced by Chinese speakers so that frequency-based wordlists would reflect the centrality of some China English words. They compared the frequency of a sample of common China English words in the China English Corpus and the British National Corpus (BNC) to support their argument. In addition, they further called for a Chinese-oriented description of the English word senses. They used concordance lines and world knowledge to show how the meaning of the noun *house* is associated with two different concepts for Chinese and English speakers. Accordingly, the definitions targeting Chinese learners should consider such differences.

Notwithstanding, this argument does not sound plausible for various reasons. First, monolingual learner's dictionaries usually target a proficiency level (e.g., beginners, intermediate, advanced). They cannot specify users according to their first language, given the status of English as a lingua franca. Moreover, non-native varieties of English such as Spanish English or Indian English are also expanding and displaying their own distinguishing features. They are admitted in English dictionaries when they are influential enough to be frequently uttered by native speakers and, hence, reflected in native general reference corpora. According to Ooi (2021), there are several cases of admitting Japanese words to the big five dictionaries even without the regional label "in Japan" because they were frequent enough in native general corpora.

5. Conclusion

To conclude, lexicographers made extensive use of corpus tools in the past decade to construct dictionaries and improve existing ones. Compiling headword lists, detecting collocational patterns, identifying word senses and revealing idiom variations have been some of the lexicographic challenges that are successfully met through corpus tools. The frequency either for wordlists or specific lemmas, keyword extraction and word sketches are among the most effective aids for lexicographers. According to the explored studies, almost all dictionaries benefit from corpus tools at the macro and microstructure levels. The need for using these tools became most pressing while compiling specialized dictionaries and landscaping language varieties.

Worthy mentioning, no study can be comprehensive enough to include every work at the intersection between corpora and lexicography. The scope of the present study aimed at covering some of the corpus contributions to lexicography in the past ten years only. Therefore, early significant contributions were not tackled in the study. Moreover, the discussed studies are cited only from the journals that are indexed in PubMed, Science Direct, Web of Science, Scopus, Mendeley, ERIH PLUS, PsycINFO, ProQuest, and Crossref. Accordingly, studies that were published in conference proceedings and journals that are not indexed in the abovementioned databases were not covered in the study regardless of their relevance.

References

- Abecassis, M.** 2007. Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs. *Lexikos* 17: 247-258. <https://doi.org/10.5788/17-0-555>.
- Anshen, F. and M. Aronoff.** 1999. Using Dictionaries to Study the Mental Lexicon. *Brain and Language* 68(1-2): 16-26.
- Atkins, S.** 2008. Theoretical Lexicography and its Relation to Dictionary-making. Fontenelle, T. (Ed.). 2008. *Practical Lexicography: A Reader*: 31-50. Oxford: Oxford University Press.
- Atkins, S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Azkarate, M. and D. Lindemann.** 2018. Basque Lexicography and Purism. *International Journal of Lexicography* 31(2): 132-150. <https://doi.org/10.1093/ijl/icy003>.
- Bergenholtz, H. and R.H. Gouws.** 2017. Polyseme Selection, Lemma Selection and Article Selection. *Lexikos* 27: 107-131. <https://doi.org/10.5788/27-1-1396>.
- Bogaards, P.** 2013. A History of Research in Lexicography. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 19-31. London/New York: Bloomsbury Academic.
- Boz, E., F. Bozkurt and F. Dođru.** 2018. Corpus-based Research on Terminology of Turkish Lexicography (CBRT-TURKLEX). *Lexikos* 28: 428-439. <https://doi.org/10.5788/28-1-1472>.
- Cabezas-García, M. and P. Faber.** 2018. Phraseology in Specialized Resources: An Approach to Complex Nominals. *Lexicography* 5(1): 55-83. <https://doi.org/10.1007/s40607-018-0046-x>.
- Dalpanagioti, T.** 2018. A Frame-semantic Approach to Co-occurrence Patterns: A Lexicographic Study of English and Greek Motion Verbs. *International Journal of Lexicography* 31(4): 420-451. <https://doi.org/10.1093/ijl/icy016>.
- Dalpanagioti, T.** 2019. From Corpus Usages to Cognitively Informed Dictionary Senses: Reconstructing an MLD Entry for the Verb 'float'. *Lexicography* 6(2): 75-104. <https://doi.org/10.1007/s40607-019-00059-5>.
- De Schryver, G.-M. and M. Nabirye.** 2018a. Corpus-driven Bantu Lexicography. Part 2: Lemmatization and Rulers for Lusoga. *Lexikos* 28: 79-111. <https://doi.org/10.5788/28-1-1458>.
- De Schryver, G.-M. and M. Nabirye.** 2018b. Corpus-driven Bantu Lexicography. Part 3: Mapping Meaning onto Use in Lusoga. *Lexikos* 28: 112-151. <https://doi.org/10.5788/28-1-1459>.
- Dolezal, F.T.** 2020. World Englishes and Lexicography. Nelson, C.L. et al. 2019. *The Handbook of World Englishes*: 725-740. Second edition. Hoboken, NJ: Wiley Blackwell.

- Faaf, G.** 2018. Lexicography and Corpus Linguistics. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 123-137. Abingdon, Oxon: Routledge. <https://doi.org/10.4324/9781315104942-9>.
- Frankenberg-Garcia, A., G.P. Rees and R. Lew.** 2021. Slipping through the Cracks of e-Lexicography. *International Journal of Lexicography* 34(2): 206-234. <https://doi.org/10.1093/ijl/ecaa022>.
- Gizatova, G.** 2018. A Corpus-based Approach to Lexicography: A New English–Russian Phraseological Dictionary. *International Journal of English Linguistics* 8(3): 357-363. <https://doi.org/10.5539/ijel.v8n3p357>.
- Hanks, P.** 2004. Corpus Pattern Analysis. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 87-97. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne-Sud.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Allan, K. 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.
- Hanks, P. and S. Može.** 2019. The Way to Analyze "way": A Case Study in Word-specific Local Grammar. *International Journal of Lexicography* 32(3): 247-269. <https://doi.org/10.1093/ijl/ecz005>.
- Heuberger, R.** 2018. Dictionaries to Assist Teaching and Learning. Fuertes-Olivera (Ed.). 2018. *The Routledge Handbook of Lexicography*: 300-316. London: Routledge. <https://doi.org/10.4324/9781315104942-20>.
- Hudeček, L. and M. Mihaljević.** 2020. The Croatian Web Dictionary — mrežnik project — Goals and Achievements. *Rasprave Instituta Za Hrvatski Jezik i Jezikoslovlje* 46(2): 645-667. <https://doi.org/10.20344/AMP.12295>.
- Jorgensen, J.C.** 1990. The Psychological Reality of Word Senses. *Journal of Psycholinguistic Research* 19(3): 167-190.
- Kilgarriff, A.** 1992. Dictionary Word Sense Distinctions: An Enquiry into their Nature. *Computers and the Humanities* 26(5–6): 365-387.
- Kilgarriff, A.** 1998. The Hard Parts of Lexicography. *International Journal of Lexicography* 11(1): 51-54. <https://doi.org/10.1093/ijl/11.1.51>.
- Kilgarriff, A.** 2005. Putting the Corpus into the Dictionary. *Proceedings of the Second MEANING Workshop, Trento, Italy, 3–4 February 2005*.
- Kochová, P.** 2019. Frequency in Corpora as a Signal of Lexicalization (On the Absolute Usage of Comparative and Superlative Adjectives). *Jazykovedný Časopis* 70(2): 148-157. <https://doi.org/10.2478/jazcas-2019-0046>.
- Moon, R.** 2013. Braving Synonymy: From Data to Dictionary. *International Journal of Lexicography* 26(3): 260-278. <https://doi.org/10.1093/ijl/ect022>.
- Nelson, K.** 2020. Informing Lexicographic Choices through Corpus and Perceptual Data. *International Journal of Lexicography* 33(3): 251-268. <https://doi.org/10.1093/ijl/ecz030>.
- Ooi, V.B.Y.** 2021. Issues and Prospects for Incorporating English Use in Japan into the Dictionary. *Asian Englishes* 23(1): 62-78. <https://doi.org/10.1080/13488678.2021.1876952>.
- Page, M.J. et al.** 2021. PRISMA 2020. Explanation and Elaboration: Updated Guidance and Examples for Reporting Systematic Reviews. *BMJ* 372(71).
- Park, M.** 2020. New Loan Words in the Neologismenwörterbuch: Corpus-based Development of Lexicographic Information for an Online Dictionary of Contemporary German. *Lexicography* 7(1–2): 5-23. <https://doi.org/10.1007/s40607-020-00070-1>.

- Pimentel, J.** 2013. Methodological Bases for Assigning Terminological Equivalents. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 19(2): 237-257. <https://doi.org/10.1075/term.19.2.04pim>.
- Prinsloo, D.J.** 2015. Corpus-based Lexicography for Lesser-resourced Languages — Maximizing the Limited Corpus. *Lexikos* 25: 285-300. <https://doi.org/10.5788/25-1-1300>.
- Renau, I. and A. Alonso Campo.** 2016. Systematising Corpus-based Definitions in Second Language Lexicography. *DELTA Documentacao de Estudos Em Linguistica Teorica e Aplicada* 32(4): 953-979. <https://doi.org/10.1590/0102-44500700542375362>.
- Siepmann, D.** 2015. Dictionaries and Spoken Language: A Corpus-based Review of French Dictionaries. *International Journal of Lexicography* 28(2): 139-168. <https://doi.org/10.1093/ijl/ecv006>.
- Sinclair, J.** 1992. *Collins COBUILD English Language Dictionary*. London: HarperCollins.
- Smirnova, A.** 2021. In Awe of God, Nature and Technology: A Lexical Approach to the Differentiation of Emotional Responses. *3L: Language, Linguistics, Literature* 27(4): 230-243. <https://doi.org/10.17576/3L-2021-2704-16>.
- Spence, J.** 2021. A Corpus Too Small: Uses of Text Data in a Hupa–English Bilingual Dictionary. *International Journal of Lexicography* 34(4): 413-436. <https://doi.org/10.1093/ijl/ecab006>.
- Wild, K., A. Kilgarriff and D. Tugwell.** 2013. The Oxford Children's Corpus: Using a Children's Corpus in Lexicography. *International Journal of Lexicography* 26(2): 190-218 <https://doi.org/10.1093/ijl/ecs017>.
- Wójtowicz, B.** 2016. Learner Features in a New Corpus-based Swahili Dictionary. *Lexikos* 26: 402-415. <https://doi.org/10.5788/26-1-1343>.
- Xia, L., Y. Xia, Y. Zhang and H. Nesi.** 2016. The Corpora of China English: Implications for an EFL Dictionary for Chinese Learners of English. *Lexikos* 26: 416-435. <https://doi.org/10.5788/26-1-1342>.

Appendix: Table 1

Author(s)	Lexicographic task	Type of challenge	Relevance to dictionary structure	Relevance to other linguistic fields	Analyzed language	Type of the lexicographic resource	Role of the lexicographic resource	Used corpus	Corpus solution	Type of analysis
Wild et al. (2013)	Compiling lemma lists, defining senses, providing examples	Lexicographic	Macrostructure/ Microstructure	Psycholinguistics [children discourse]	English	Children/school dictionaries	Comparison/ reference point	Oxford Children Corpus; Oxford English Corpus	Keywords; lemma frequency; word sketches; concordance	Corpus-based
Moon (2013)	Differentiating near synonyms	Lexicographic	Microstructure	Lexical semantics	English	Monolingual learner's dictionaries	Comparison/ reference point	English general reference corpus	Frequency; distribution; concordance; collocations	Corpus-based
Parental (2013)	Compiling a specialized lemma list, identifying senses	Lexicographic	Macrostructure/ Microstructure	Terminology	Portuguese-English	Bilingual Portuguese-English language resource	Main project	Comparable specialized corpus	Keyword extraction; Concordance	Corpus-driven; corpus-based; frame-based
Pirrello (2015)	Compiling a lemma list for language-specific	Lexicographic	Macrostructure/ Microstructure	Lexical semantics	Sepedi, Afrikaans; English	Monolingual English dictionary	Comparison/ reference point	Pretoria English Internet Corpus; Media 24; Sepedi corpus	Frequency word-list; Word sketch; Concordance	Corpus-driven; corpus-based
Stepmann (2015)	Describing colloquial words	Lexicographic	Microstructure	Sociolinguistics	French	Bilingual and monolingual French dictionaries	Comparison/ reference point	French general reference corpus	Frequency; collocational analysis	Corpus-based
Renau and Alonso Campo (2016)	Providing systematic definitions	Lexicographic	Microstructure	Lexical semantics	Spanish	Monolingual learner's dictionary	Main project	Spanish corpus	Concordance	Corpus-based
Wojtowicz (2016)	Compiling lemma list, providing examples	Lexicographic	Macrostructure	Psycholinguistics [mental lexicon theories]	Swahili	Bilingual Swahili-Polish dictionary	Main project	Helsinki Corpus of Swahili	Frequency word-list; concordance	Corpus-driven
Xia et al. (2016)	Inclusion and description of non-native varieties	Lexicographic	Macrostructure/ Microstructure	Sociolinguistics	English	Monolingual English dictionaries	Comparison/ reference point	China English Corpus; British English corpus	Frequency; concordance	Corpus-based

Author(s)	Lexicographic task	Type of challenge	Relevance to dictionary structure	Relevance to other linguistic fields	Analyzed language	Type of the lexicographic resource	Role of the lexicographic resource	Used corpus	Corpus solution	Type of analysis
Cabezas-garcía and Faber (2018)	Inclusion and description of specialized phrases	Lexicographic	Macrostructure/ Microstructure	Phraseology	English-Spanish	Terminological knowledge base	Main project	Specialized English-Spanish corpus	Term extraction; word sketches; concordance; frequency	Corpus-driven; corpus-based
Dalpanagioti (2018)	Identifying senses of polysemous words	Lexicographic	Microstructure	Lexical semantics; cognitive semantics	English-Greek	Bilingual lexical database	Main project	General reference corpora in English and Greek	Frequency; concordance; concordance	Corpus-based; frame-based
De Schryver and Nabbye (2018a)	Compiling a lemma list	Lexicographic; language-specific	Macrostructure	Psycholinguistics [mental lexicon theories]	Lusoga	Bantu language dictionary	Main project	Lusoga corpus	Frequency; word list	Corpus-driven
De Schryver and Nabbye (2018b)	Identifying, splitting and lumping senses	Lexicographic	Microstructure	Lexical semantics	Lusoga	Lusoga dictionary	Main project	Bantu language corpus	Concordance; collocations	Corpus-based
Gizatova (2018)	Inclusion of MVEs, recording variations in MVEs, detecting false friends	Lexicographic	Macrostructure/ Microstructure	Applied linguistics [translation equivalence]	English-Russian	Bilingual phraseological dictionary	Main project	English-Russian parallel corpora; general reference corpora in English and Russian	Frequency; concordance; collocation	Corpus-based; contrastive
Dalpanagioti (2019)	Identifying, ordering and linking senses of polysemous lemmas	Lexicographic	Microstructure	Lexical semantics; cognitive semantics	English	Monolingual learners dictionaries	Comparison/reference point	English general reference corpus	Word sketches; concordance	Corpus-based; frame-based
Hanks and Može (2019)	Identifying the phraseology of words	Lexicographic	Microstructure	Lexis-grammar interface	English	Pattern dictionary of English verbs	Comparison/reference point	English general reference corpus	Frequency; word sketches	Corpus Pattern Analysis
Kochová (2019)	Inclusion of inflected forms as lexical units	Lexicographic	Microstructure	Morphology	Czech	Monolingual Czech dictionaries	Comparison/reference point	Czech corpus	Frequency; collocational analysis	Corpus-based
Nelson (2020)	Defining and ordering word senses	Lexicographic	Microstructure	Lexical semantics	English	Monolingual English dictionary	Comparison/reference point	English general reference corpus	Concordance	Corpus-based
Frankenberg-García et al. (2021)	Detection and description of collocations	Lexicographic	Microstructure	Lexicology	English	Lexica database for collocations	Main project	Academic and general reference corpora	Keyword extraction; concordance; word sketches; frequency	Corpus-driven; corpus-based

Author(s)	Lexicographic task	Type of challenge	Relevance to dictionary structure	Relevance to other linguistic fields	Analyzed language	Type of the lexicographic resource	Role of the lexicographic resource	Used corpus	Corpus solution	Type of analysis
Hudeček and Mihajević (2020)	Compiling a lemma list, providing examples	Lexicographic	Macrostructure/ Microstructure	Psycholinguistics [mental lexicon theories]	Croatian	Croatian web dictionary	Main project	Croatian corpus	Frequency, word list, word sketches, concordance	Corpus-based
Park (2020)	Standardizing the description loan words	Lexicographic	Microstructure	Morphology; phonology; orthography	German	Monolingual German general and specialized dictionaries; specialized online dictionary	Comparison/reference point	German reference corpus	Frequency	Corpus-based; diachronic
Ooi (2021)	Inclusion and description of regional words	Lexicographic	Microstructure	Sociolinguistics	Japan English	Monolingual English dictionaries	Comparison/reference point	Japanese English news corpus; GloWbE corpus	Concordance; collocational analysis	Corpus-based; comparative
Srinivasa (2021)	Detection of polysemy	Lexicographic	Microstructure	Lexical semantics	English	Monolingual dictionaries	Comparison/reference point	General reference corpus	Concordance; collocational analysis	Corpus-based
Spence (2021)	Inclusion of words of polysynthetic language; recording inflected forms; providing examples	Language-specific	Macrostructure/ Microstructure	Sociolinguistics	Hupa* - English	Bilingual dictionary	Main project	Hupa text corpus	Frequency, word list, concordance	Corpus-based

Teksgebaseerde tesourusgebruik in 'n Afrikaanse taalonderrigkonteks

Michele F. van der Merwe, *Departement Kurrikulumstudie: Afrikaans,
Fakulteit Opvoedkunde, Universiteit Stellenbosch, Stellenbosch*
(michelevdm@sun.ac.za)

Opsomming: Die onderwerp van die artikel is die ontsluitingsmoontlikhede van tesourusgebruik in 'n Afrikaanse taalonderrigkonteks in die skoolklaskamer. Die tesourus is een van die oudste naslaanwerke in die geskiedenis van leksikografie; tog word dit nie in die aantal navorsingsartikels rakende die onderwerp gereflekteer nie. 'n Tesourus blyk nie 'n algemene fokuspunt van leksikografiese navorsing of opvoedkundige leksikografiese navorsing te wees nie. In 'n poging om die gaping in die literatuur te oorbrug, word daar oor die doel en leksikografiese funksie van 'n tesourus besin, veral in 'n taalonderrigkonteks. Die taalonderrigkonteks word vervolgens belig met verwysing na voorgestelde tesourusgebruik in skoolkurrikuladokumente, aangesien dit rigtinggewend is vir onderrigraamwerke in skole. Teksgebaseerde tesourusgebruik in die taalonderrigkonteks word aanbeveel en gemotiveer teen die agtergrond van 'n ondersoek na 'n tesourus en tematiese woordeboekafdelings. Aanbevelings vir enkele verbeterings tot leksikografiese produkte word vanuit 'n opvoedkundige leksikografiese perspektief gemaak ten einde optimale gebruik van 'n tesourus in 'n taalonderrigkonteks te verseker.

Sleutelwoorde: TESOURUS, TEKSGBASEERD, TAALONDERRIGKONTEKS, WOORDEBOEFUNKSIE, OPVOEDKUNDIGE LEKSIKOGRAFIE, SINONIEM, SEMANTIESE VERHOUDING, WOORDESKATUITBREIDING, WOORDESKATBENOEMING, TAALVARIANT

Abstract: Text-based Use of a Thesaurus in an Afrikaans language and teaching context. The topic of the article articulates and explores possibilities for the use of a thesaurus in the language and teaching context of the school classroom. Although the thesaurus is one of the oldest reference works in the history of lexicography, that fact is not reflected in the number of research articles on the topic. A thesaurus does not seem to be a general focus point of lexicographic research, nor be it of pedagogical lexicography. To address the gap in literature, the purpose and lexicographic functions of a thesaurus, especially in a language education context, are foregrounded. The language education context is illuminated with reference to the prescribed use of the thesaurus in school curricula documents, as they dictate the teaching policy in public schools. Text-based use of a thesaurus is recommended and is illustrated against the background of an investigation of a thesaurus and thematic dictionary sections. Recommendations for improvement of lexicographical products are made from a pedagogical lexicographic perspective to ensure optimal use of a thesaurus in a language education setting.

Keywords: THESAURUS, TEXT-BASED, LANGUAGE EDUCATION CONTEXT, DICTIONARY FUNCTION, PEDAGOGICAL LEXICOGRAPHY, SYNONYM, SEMANTIC RELATION, VOCABULARY DEVELOPMENT, VOCABULARY ACQUISITION, LANGUAGE VARIANT

1. Inleiding

Die potensiële gebruik van 'n tesourus op skool, veral in die laerskool, mag 'n vreemde idee vir leksikograwe en onderwysers wees, aangesien tesourusgebruik gewoonlik met meer gesofistikeerde taalgebruikers geassosieer word. Skoolleerders word nie noodwendig as gesofistikeerde taalgebruikers beskou nie, aangesien hulle oor wisselende taalvaardighede en leksikografiese vaardighede beskik. 'n Tesourus is nie 'n bekende en algemeen gebruikte leksikografiese naslaanbron onder onderwysers nie. Tog verwys Tasovac (2012: 242) na navorsing in opvoedkundige leksikografie wat daarop dui dat die voorsiening van stelle semanties verwante woorde in 'n leksikografiese inskrywing taalaanleerders kan help om betekenis te verstaan en hul memorisering te verbeter (Geckeler 2002, Kremer et al. 2008).

Van Schalkwyk (1996: 286) beveel in 'n resensie van die *Groot Tesourus van Afrikaans* die onderrig van tesourusgebruik aan deur dit as 'n "noodsaaklikheid" te bestempel:

Waarop die voorafgaande afstuur, is die wenslikheid, nee, noodsaaklikheid dat die bestaan en waardevolle gebruike van die tesourus as besondere soort woordeboek dieselfde gebruiksfrekwensie verkry as verklarende en vertalende (tweetalige) woordeboeke. Verder moet die woordeboekgebruiker (kleintyd) geskool word in die woordeboekkultuur en in die gebruik van die tesourus.

Carstens (1995: 205) is van mening dat daar in die moedertaalklas gedurende die primêre onderwysfase met woordeboekpedagogiek begin moet word. Leerlinge en studente sal egter beslis daarby baat as woordeboekvaardighede gedurende ook die intermediêre fases van tweede- en vreemdetaalonderrig geslyp word. Die verwagting word gestel dat leerlinge wat proefondervindelik oortuig geraak het van woordeboeke se probleemoplossende potensiaal selfstandige leerders sal word wat sal aanhou om woordeboeke te gebruik, selfs nadat die tydperk van formele taalonderrig afgesluit is (Carstens 1995: 205). Van der Merwe (2009: 302) bied 'n model vir woordeboekonderrig aan. Volgens die model bestaan woordeboekonderrig uit 'n proses met die volgende komponente: identifisering van uitkomst van die taalkurrikulum, vind van toepaslike woordeboek, metode van onderrig (verwysing na woordeboekteks en tesourusteks), woordeboeknaslaanvaardighede, woordeboektipologie en woordeboekfunksies (Van der Merwe 2009: 302).

Tesourusse, asook tesourusgebruik, figureer nie baie in akademiese opvoedkundige leksikografiese besprekings en tydskrifte nie, allermins in onderrigsituasiebesprekings. Soektogte in EBSCOhost se databasis onder soekwoorde soos "thesaurus" OR "thesaurus use" OR "lexicon" OR "glossary" OR "synonym dictionary" OR "vocabulary list", asook "language learning" OR "language teaching" OR "vocabulary learning" OR "vocabulary" OR "language skills" het tot gebrekkige soekresultate gelei, met slegs ses navorsingsartikels wat gevind is.

Drie navorsingsartikels is gevind wat op die onderrig van semantiese betrekkings in taalonderrig fokus, maar met geen verwysing na tesourusgebruik nie, en word dus buite rekening gelaat in hierdie artikel. Augustosky en Fastman (2021) fokus op tesourusgebruik tydens die skryfproses, Stockdale (2019) op die skep van 'n tesourus vir kollegestudente, en Tamor en Walmsley (1983) op die voordele van die gebruik van 'n tesourus tydens redigeringsstake in taalonderrig. Chainikova, Zatonkiy, Mitiukov en Busygina (2018) skryf oor die gebruik van 'n tesourus in die taalaanleerproses en slegs Mountain (2007) fokus op sinoniemonderrig met behulp van 'n tesourus.

Mountain (2007: 318) beskryf sinoniemonderrig met behulp van 'n tesourus deur na die voordele daarvan vir woordeskatonderrig, soos deur Graves (2006) en Cunningham en Allington (2007) beskryf, te verwys. Sy stel die gebruik van 'n tesourus in die klaskamer bekend deur leerders 'n dinkskrum-aktiwiteit te laat doen om geskikte sinonieme vir 'n semanties oormatig gebruikte woord, soos *sê*, te vind. Sinonieminskrywings in tesourusse word tydens die aktiwiteit verduidelik. Tesourusse is in die klas beskikbaar en leerders maak van ook aanlyn tesourusse gebruik. Soos leerders vertrou raak met tesourusgebruik, word 'n sinoniemspeletjie soos kolletjies-en-kruisies gedemonstreer en daarna ontwerp, aanvanklik deur die onderwyser en later deur die leerders, en leerders speel hul eie weergawes daarvan. Woorde soos *tornado*, *orkaan* en *sikloon* is as voorbeelde in die aanvangspel gebruik, aangesien dit weersomstandighede is wat aan leerders in die nagevorste omgewing bekend is. Met die opstel en speel van die sinoniemspeletjie vind groepsbesprekings plaas en leerders vind gepaste sinonieme deur die semantiese waarde van 'n woord te bespreek. Sodoende word woordbewustheid gekweek. Die aangekweekte woordbewustheid en woordeskatontwikkeling vind volgens Mountain (2007: 322) verder neerslag in die klasgroep se skryfwerk soos die gereelde gebruik van 'n tesourus toeneem.

Met betrekking tot die gebruik van *Tesourus van Afrikaans* (voortaan TA) in die Afrikaanse taalonderrigsituasie is geen inligting gevind nie. Die ooglopende gaping in literatuur is dan ook een van die redes waarom hierdie navorsing onderneem word.

Hierdie artikel word gesitueer binne die konteks van opvoedkundige leksikografie met verwysing na tesourusgebruik in 'n opvoedkundige konteks, asook binne die teorie van leksikografiese funksies, met 'n bespreking van die leksikografiese funksies van tesourusse. Eerstens word die doel en leksikografiese funksies van 'n tesourus bespreek; tweedens word verwysings na tesourusgebruik in Afrikaanse skoolkurrikula ondersoek; derdens word die teksgebaseerde benadering tot die taalonderrigkonteks verduidelik; en in die laaste afdeling word teksgebaseerde tesourusgebruik in 'n skoolonderrigkonteks ondersoek, met verwysing na toepaslike leksikografiese voorbeelde. Aanbevelings word gemaak en moontlike wysigings voorgestel vir meer toepaslike leksikografiese hantering van tesourusartikels.

2. Doel en leksikografiese funksies van 'n tesourus in die taalonderrigkonteks

'n Tesourus verteenwoordig van die oudste leksikografiese werke, veral in die pre-alfabetiese tyd, toe alle woordeboeke tematies gerangskik is. Volgens Kay en Alexander (2016: 366) is 'n tesourus ook 'n leksikografiese naslaanwerk wat leksikale feite binne semantiese domeine weergee. Marelló (1989: 1083) beskryf 'n tesourus as 'n woordeboek waarin die leksikale items gerangskik word met inagneming van hulle verhouding tot bepaalde konsepte of konsepkategorieë, maar ook met inagneming van hul onderlinge betekenisbetrekkinge. Kilgarriff en Yallop (2000: 1373) definieer 'n tesourus as 'n bron waar woorde met soortgelyke betekenis saamgegroeper is. Vier soorte tesourusse word onderskei, naamlik (1) tesourusse in boekformaat om gebruikers met woordkeuse te help; (2) Word-Net-sagteware; (3) tesourusse in inligtingsprosesseringstelsels; en (4) outomatiese tesourusse wat met behulp van korpora saamgestel is om ooreenkomste tussen woorde te identifiseer (Kilgarriff en Yallop 2000: 1373). Dit is in teenstelling met drie soorte tesourusse wat deur Marelló (1989: 1083) onderskei word, naamlik (1) kumulatiewe, (2) verklarende en (3) vertalende tesourusse, en verwys duidelik na die ontwikkeling van tesourusse, ook in die elektroniese omgewing. In 'n kumulatiewe tesourus word leksikale items onder 'n bepaalde konsep gerangskik sonder om vir hulle betekenisverklarings aan te bied., soos byvoorbeeld TA. 'n Verklarende tesourus bied betekenisverklarings vir woorde en uitdrukkinge aan en 'n vertalende tesourus bied vertaalekwivalente vir woorde aan (De Stadler en De Stadler 2021: voorwoord).

In 'n tesourus word die struktuur van die woordeskat en die besondere verhoudings wat bestaan tussen die leksikale items in daardie woordeskat op die voorgrond geplaas. 'n Woordeskat vertoon 'n bepaalde struktuur waarbinne die verskillende leksikale items op grond van veral hulle betekenis met mekaar saamhang. Veral vier betrekkinge kan binne hierdie struktuur onderskei word: sinonimie, hiponimie, meronimie en antonimie.

De Stadler en De Stadler (2021: voorwoord) beskryf die funksies van 'n tesourus volledig, naamlik (1) om die regte woord vir die regte konteks te suggereer; (2) hulp aan die skrywer om die stilisties lastige herhaling van 'n bepaalde woord te vermy deur die aandag te vestig op ander woorde wat min of meer dieselfde betekenis oordra; (3) die uitbreiding van die woordeskat van 'n taalgebruiker; en (4) om terminologieë vir bepaalde onderwerpe te verskaf.

In ooreenstemming met die bogenoemde uiteensetting beskou Kay en Alexander (2016: 375) ook die algemeenste funksie van 'n tesourus as 'n hulpmiddel tot enkoderende take, naamlik die produksie van 'n teks. 'n Tesourus kan skrywers ondersteun om 'n alternatiewe woord te vind vir een wat reeds gebruik is, of om 'n meer gepaste woord vir 'n konsep te vind.

Met behulp van die leksikografiese funksieteorie vir gebruikersgerigte situasies kan daar aangetoon word dat 'n tesourus oor meer funksies beskik, veral as dit in 'n skoolonderrigkonteks ondersoek word.

Binne die raamwerk van die leksikografiese funksieteorie onderskei Tarp (2000: 196) tussen verskillende leksikografiese funksies van 'n woordeboek, waar 'n leksikografiese funksie gedefinieer word as "the endeavour and ability of the dictionary to cover the complex needs that arise in the user in a particular user situation". Hy onderskei twee hoof funksies met betrekking tot gebruikersituasies, naamlik kommunikatief en kognitief. Die belangrikste kommunikatiefgerigte funksies sluit onder meer hulp met die resepsie en produksie van tekste in die huistaal en addisionele taal, asook vertalings van tekste uit die huistaal in 'n addisionele taal, in. Wanneer die kommunikatiefgerigte funksies van 'n tesourus ter sprake is, is dit van toepassing op die resepsie en produksie van 'n huistaal en 'n addisionele taal. Die fokus van hierdie artikel is op tesourusgebruik in die skoolkonteks. 'n Tesourus kan kommunikatiefgerigte funksies binne 'n onderrigkonteks vervul tydens die resepsie en produksie van beide taalvlakke.

Vervolgens verwys Tarp (2000: 197) na kognitiefgerigte funksies, wat onder meer inligting rakende taal, algemene kulturele en ensiklopediese inligting, asook spesifieke inligting omtrent 'n vakgebied insluit. Binne die onderrigkonteks kan die genoemde funksies ter sprake kom tydens tesourusgebruik. Inligting aangaande taal is ter sprake in die onderrigkonteks, veral tydens woordeskatontwikkeling, woordeskatuitbreiding en die identifisering van semantiese verhoudings tussen leksikale items. Betreffende kulturele en ensiklopediese inligting verskaf 'n tesourus tematiesgebaseerde inligting omtrent die organisasie van kennis in die wêreld. Kulturele en ensiklopediese inligting kan baie relevant tot die onderrigkonteks wees.

Kennis van die leksikografiese doel en funksies van 'n tesourus in 'n taalonderrigkonteks kan lei tot die gebruik van 'n tesourus in die skoolkonteks, maar die beskouing van 'n tesourus in rigtinggewende kurrikulumdokumente vir taalonderrig is deurslaggewend. Onderwysers se onderrig word deur die kurrikulum gerig en indien tesourusgebruik in die kurrikulum genoem word, kan dit lei tot tesourusgebruik in die klaskamer. Die tersaaklike kurrikulumdokumente word in die volgende afdeling ondersoek om vas te stel of die gebruik van 'n tesourus genoem word.

3. Verwysing na tesourusgebruik in Afrikaans-skoolkurrikula

Onderwysers in Suid-Afrikaanse openbare skole word gelei deur kurrikulumbeleidsdokumente soos die Kurrikulum- en Assesseringsbeleidsverklaring (voortaan KABV) en in 'n mindere mate departementeel goedgekeurde skoolhandboeke. Tydens die oorweging van (woordeboekgebruik en) tesourusgebruik in onderrigverband sal dit rigtinggewend wees om kurrikulumbeleidsdokumente te ondersoek. Volgens Nkomo (2015: 76) verwys KABV-dokumente vir die algehele skoolstelsel herhaalde male na die belangrikheid van woordeboeke en woordeboekvaardighede. Nkomo (2015: 76) beskou kurrikulumdokumente as verteenwoordigend van Suid-Afrika se opvoedkundige beleid aangaande woor-

deboekgebruik en woordeboekkultuur. In 'n soektog na die woord *woordeboek* of *woordeboeke* is gevind dat die woord in 11 dokumente ten opsigte van Engels as vak voorkom. Die dokumente verteenwoordig Engels op verskillende vlakke, naamlik as huistaal en eerste addisionele taal in die Grondslagfase, as huistaal, eerste addisionele taal en tweede addisionele taal vanaf die Intermediêre Fase tot die Verdere Onderwys- en Opleidingsfase (Nkomo 2015: 76). Kurrikulumdokumente vir ander tale, byvoorbeeld Afrikaans, kan as blote vertalings van die Engels beskou word. Nkomo (2015: 77) beskou die verwysing na woordeboekgebruik in taalvakke in die Suid-Afrikaanse opvoedingstelsel as 'n merkbare aanduiding dat woordeboeke as taalaanleerinstrument beskou word. Leerders word vanaf die Grondslagfase aangemoedig om woordeboeke te gebruik om hul woordeskat uit te brei en om betekenis- en spelinligting te soek (Nkomo 2015: 77). Ouer leerders word aangemoedig om woordeboeke te gebruik vir inligtingstipes soos uitspraak, woordsoorte, sinonieme, antonieme en homonieme, as 'n erkenning van die belangrikheid van taal as kommunikasie-medium in leersituasies en die feit dat woordeboeke kommunikatiewe sowel as kognitiewe aktiwiteite kan ondersteun (Nkomo 2015: 77).

'n Elektroniese soektog na die soekwoord *tesourus* in kurrikulumdokumente vir Afrikaans Huistaal in die Verdere Onderwys- en Opleidingsfase (voortaan VOO-fase) lewer insiggewende resultate op, naamlik 'n voorkoms van ses keer. Die voorkoms van die soekwoord *tesourus* is insiggewend aangesien die navorser nie voorsien het dat tesourusgebruik in skoolverband wel genoem sal word nie. Tesourusgebruik word gewoonlik met gesofistikeerde taalgebruikers geassosieer. Vervolgens word die voorkoms van die woord *tesourus* in die KABV in konteks (kursiefdruk) verskaf, gevolg deur 'n kort kommentaar deur die navorser in terme van leerdoelwitte, onderwysersrol en leerderaktiwiteite:

- *Beplanning, navorsing en organiserings. Gebruik bronne en verwysingsmateriaal soos woordeboeke en tesourusse om effektiewe en presiese woordeskat te kies en aanbiedings te lewer deur gebruik te maak van notas en rekwisiete, oudio- en/of visuele hulpmiddels en grafieke om die aanslag en akkuraatheid van aanbiedings te verhoog* (DBO 2011c: 18). Leerdoelwitte word ten opsigte van twee taalvaardighede gespesifiseer. Die onderwysersrol word gestipuleer, maar verwagtinge ten opsigte van leerders word nie verskaf nie.
- *Woordeboeke/tesourus vir luisterbegrip* (DBO 2011c: 22). 'n Vae, nie-spesifieke leerdoelwit ten opsigte van een van die vier taalvaardighede word aangebied. Geen onderwysersrol word gestipuleer nie en verwagtinge ten opsigte van leerders word nie verskaf nie.
- *Gebruik van woordeboeke, tesourusse en ander naslaanboeke om die betekenis, spelling, uitspraak, lettergreepverdeling en woordsoortlikheid van onbekende en komplekse woorde te bepaal* (DBO 2011c: 24). Duidelike en spesifieke leerdoelwitte ten opsigte van taalstrukture en -konvensies word verskaf. Geen onderwysersrol word gestipuleer nie, maar verwagtinge ten opsigte van leerders word verskaf.

- *Tekste vir gebruik tydens die geïntegreerde onderrig van taalvaardighede. Naslaan-tekste: Ensiklopedieë, handboeke, roosters, skedules, telefoongidse, tesourus, TV-gidse, woordeboeke* (DBO 2011c: 29). 'n Tesourus word as 'n naslaanbron erken en kan dus as teks vir integrasie van taalstrukture en -konvensies gebruik word, maar leerdoelwitte word nie gespesifiseer nie. Geen onderwysersrol word gestipuleer nie en verwagtinge ten opsigte van leerders word ongelukkig nie verskaf nie.
- *Woordkeuse. Gebruik bronne en navorsingsmateriaal soos woordeboeke en tesourusse om effektiewe en gepaste woorde te kies* (DBO 2011c: 32). Die waarde van 'n tesourus as naslaanbron word erken en 'n spesifieke leerdoelwit word daarmee verbind, alhoewel dit vaag is en die konteks nie verskaf word nie. 'n Gepaste konteks kon moontlik kreatiewe skryf wees. Geen onderwysersrol word gestipuleer nie en verwagtinge ten opsigte van leerders word in 'n mindere mate verskaf.
- *Woordelys. tesourus* — 'n woordeboek wat volgens saakverbande georden is, saakwoordeboek (DBO 2011c: 101). In die kurrikulumdokumente word die konsep van 'n tesourus verklaar. Geen bronverwysing word aangedui nie, maar by nadere ondersoek is die bron opgespoor, naamlik die *Pharos Verklarende Woordeboek*. Dit is belangrik dat die woord *tesourus* in die woorde-lys opgeneem word en van 'n definisie voorsien word (die geslaagdheid van die betekenisoordrag van die definisie kan bevraagteken word). 'n Moontlike motivering vir die opname van *tesourus* is dat dit 'n minder bekende konsep as 'n woordeboek verteenwoordig. 'n Meer gepaste betekenisverklaring vir die gebruik deur onderwysers sou die verklaring deur De Stadler en De Stadler (2021: voorwoord) (sien Afdeling 2 in hierdie verband) wees. 'n Gepaste verklaring vir hoërskoolleerders kan uit die *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT) verkry word, naamlik: "woordeboek waarin woorde sistematies volgens begrippe of konsepte gegroepeer is". 'n Meer gepaste verklaring vir die gebruik deur laerskoolleerders sou die *Pharos Junior Tweetalige Skoolwoordeboek — Afr./Eng.* se verklaring wees:

tesourus (te·sou·rus) [tesourusse; tesourussie] **selfstandige naamwoord** (soortnaam)
thesaurus ♦ *Jy kan sinonieme in 'n tesourus opsoek.* You can look up synonyms in a **thesaurus**.

Die elektroniese soektog na die soekwoord *tesourus* in kurrikulumdokumente vir Afrikaans Eerste Addisionale Taal in die VOO-fase lewer eweneens positiewe resultate op, naamlik 'n voorkoms van nege keer. Dit is min of meer 'n herhaling van resultate vir die Huistaalkurrikulum, met die volgende bykomende inligting:

- *Gebruik woordeboeke en tesourusse om woordeskat uit te brei* (DBO 2011e: 37). 'n Leerdoelwit, alhoewel ietwat vaag, word ten opsigte van taalvaardig-

hede gespesifiseer. Die onderwysersrol word gestipuleer, maar verwagtinge ten opsigte van leerders word nie verskaf nie.

- *Gebruik woordeboeke en 'n tesourus om nuwe/onbekende woorde na te slaan* (DBO 2011e: 62). 'n Leerdoelwit, alhoewel ietwat algemeen en nie-spesifiek, word ten opsigte van taalvaardighede verskaf. Die onderwysersrol word nie gestipuleer nie, maar verwagtinge ten opsigte van leerders word verskaf.
- *Woordeboeke en 'n tesourus vir woordbetekenisse* (DBO 2011e: 71). 'n Leerdoelwit word ten opsigte van taalvaardighede verskaf, maar dit is nie-spesifiek, aangesien dit net na woordbetekenis verwys. Die onderwysersrol word nie gestipuleer nie en verwagtinge ten opsigte van leerders word nie verskaf nie.

Die elektroniese soektog na die soekwoord *tesourus* in kurrikulumdokumente vir Afrikaans Tweede Addisionele Taal in die VOO-fase (DBO 2011f) lewer ook positiewe resultate op, naamlik 'n voorkoms van drie keer. In die kurrikulumverklaring word 'n tesourus aanbeveel as hulpmiddel vir die onderwyser en 'n tesourus kan ook as 'n teks gebruik word vir leerders. Te min inligting aangaande die gebruik van 'n tesourus as teks word aangebied om 'n onderwysersrol te stipuleer of om verwagtinge ten opsigte van leerders te konkretiseer. Aangesien die kurrikulum onderrig vir taalaanleerders beskryf, kan volledige inligting omtrent die gebruik van 'n tesourusteks vir die aanleerderkonteks waardevol vir die taalonderwyser wees.

Dit is opvallend dat tesourusgebruik in die skool aanbeveel word, op alle vlakke van taalonderrig, insluitend huistaalvlak, addisionele-taal-vlak en tweede-addisionele-taal-vlak. Aangesien 'n tesourus op 'n gesofistikeerde gebruiker gerig is, naamlik 'n gebruiker wat die tematiese werking van 'n tesourus verstaan en onder die knie het, kan dit nie aangeneem word dat 'n taalaanleerder oor die nodige vaardighede beskik om 'n tesourus suksesvol te gebruik nie. Taalaanleerders sal voldoende leiding van 'n onderwyser moet ontvang om 'n tesourus suksesvol te kan gebruik.

In sekere opsigte is kurrikulumvoorskrifte dus baie progressief en kan dit 'n groot bydrae maak tot doeltreffende tesourusgebruik in skoolonderrig. Gouws en Prinsloo (2005: 43) wys op die gebrek aan 'n woordeboekkultuur in Suid-Afrika en dit kan betwyfel word of die meerderheid van taalonderwysers bewus is van die voorstel in kurrikulumdokumente, asook van die onderrigpotensiaal van tesourusse. 'n Tesourus kan as 'n minder bekende leksikografiese produk, veral in skoolverband, beskou word en aktiwiteite vir tesourusgebruik figureer volgens die navorser se ervaring nie in voorgeskrewe taalhandboekreekse nie. Indien tesourusgebruik wel in handboekreekse sou figureer en meer onderwysers bewus word van die potensiële gebruik van tesourusse, kan die woordeboekkultuur in Suid-Afrika op 'n besliste wyse versterk word.

4. Teksgebaseerde benadering tot die taalonderrigkonteks

Tesourusgebruik in taalonderrigklasse, asook die onderrigbenadering vir onderwysers om te volg, word, soos in die vorige afdeling beskryf, in kurrikulumdokumente genoem. Volgens die KABV (DBO 2011c: 11) vir Afrikaans Huistaal moet die benadering tot taalonderrig "teksgebaseerd, kommunikatief, geïntegreerd" asook "proses-georiënteerd" wees. Volgens Lawrence et al. (2014: 258) stel 'n teksgebaseerde benadering leerders in staat om vaardige, vrymoedige en kritiese lesers, skrywers en kykers van tekste te word. Dit sluit die luister en kyk na en lees en ontleding van tekste in om te verstaan hoe dit saamgestel is en watter effek dit het. Die hoofbron van inhoud en konteks vir die kommunikatiewe, geïntegreerde leer en onderrig van taal is outentieke tekste (Lawrence et al. 2014: 258).

Die kommunikatiewe benadering vereis dat 'n leerder baie en ryk blootstelling aan die teikentaal moet kry. Om dit te bereik, moet leerders genoegsame geleentheid gebied word om taal te gebruik en te oefen.

Breë riglyne word in die KABV voorsien, soos leerderdeelname aan 'n debat waarvan die inhoud gebruik word vir die skryf van 'n argumenterende of beredenerende opstel, waarin grammatikale strukture soos sinonieme, antonieme, ontkenningvorme, voegwoorde, ens. gebruik moet word (DBO 2011c: 11). Daarby word die gebruik van outentieke tekste (soos dialoë, onderhoude, ens.) aangemoedig ten einde leerders aan realistiese taal, waaruit grammatika onderrig kan word, bloot te stel. Die gebruik van outentieke tekste skep ook die geleentheid om leerders bewus te maak van die grammatikale strukture wat met spesifieke tekstipes geassosieer word, byvoorbeeld die gebruik van die verlede tyd in verhalende opstelle. Daarenbove moet grammatika gekoppel word aan funksionele gebruike, onder andere die uitdrukking van gedagtes of gevoelens, die rig van versoeke, en die vergelyking en kontrastering van dinge (DBO 2011b: 10).

Goed-ingeligte taalonderwysers wat die voordele van tesourusse en woordeboeke in die klas insien op grond van die gebruikswaarde daarvan en die voordele aan woordeboekgebruik verbonde, sal van 'n tesourus gebruik kan maak. 'n Tesourus, vanuit die spesifieke aard daarvan, kan byvoorbeeld vir taalondersteuning en die uitbreiding van woordeskat gebruik word. Richards (1976: 78) se beskrywing van wat dit beteken "om 'n woord te ken" is volgens hom nie net beperk tot kennis van die woord se betekenis en vorm nie, maar ook afleidings, wat kennis van ook sintaktiese optrede en die netwerk van assosiasies met ander woorde insluit. Meara (1996) beskryf leksikale kennis deur aspekte van woordeskatgrootte, die getal leksikale eenhede wat die persoon ken en woordeskatorganisasie, naamlik hoe woorde 'n netwerk met mekaar vorm, te beklemtoon. Volgens Meara (1996) is daar 'n beduidende korrelasie tussen 'n groot woordeskat en 'n netwerk van betrekkinge en assosiasies tussen leksikale items.

Die bogenoemde netwerk van semantiese betrekkings tussen leksikale items word in 'n tesourus verskaf en daarom kan 'n tesourusteks dus as 'n uifers gepaste teks beskou word vir die onderrig van semantiese betrekkinge. Vergelyk De Stadler en De Stadler (2021: voorwoord) in hierdie verband:

'n Woordeskat (ook leksikon genoem) — en die Afrikaanse woordeskat is hier geen uitsondering nie — vertoon naamlik 'n bepaalde struktuur waarbinne die verskillende leksikale items op grond van veral hulle betekenis met mekaar saamhang.

In 'n tesourus word die struktuur van die woordeskat en die besondere verhoudings wat bestaan tussen die leksikale items in daardie woordeskat egter wel op die voorgrond geplaas en moet die woordeboekgebruiker dus verwag om inligting te kry oor die sinonieme, teenoorgesteldes, hiponieme of superordinate van 'n bepaalde woord.

Die onderrig van 'n netwerk van semantiese betrekkinge van die leksikon kan direk met sekere leerdoelwitte van die KABV geskakel word, naamlik:

- Onderskeiding tussen woorde wat algemeen verwar word: homofone, homonieme, sinonieme (bv. *erken/herken*, *bedank/afdanke*, *verliefde/geliefde*). Skep van lyste homonieme, sinonieme en antonieme. (DBO 2011c: 24, 25)

Teksgebaseerde onderrig kan met behulp van 'n tesourus plaasvind wanneer die onderwyser die tesourus as teks gebruik en taalkonvensies en -strukture daarmee onderrig. Die onderwyser kan artikels uit die tesourus as teks tydens die onderrigproses gebruik. In die volgende afdeling word voorbeelde van die teksgebaseerde kommunikatiewe onderrig van semantiese betrekkinge met behulp van 'n tesourus bespreek, met spesifieke verwysing na tesourusartikels wat as teks gebruik word.

5. Teksgebaseerde tesourusgebruik in 'n skoolonderrigkonteks

Vervolgens word kurrikulumgerigte taalleerdoelwitte in die skoolonderrigkonteks geskakel met voorbeelde van teksgebaseerde tesourusgebruik, met verwysing na die onderskeie onderrigfasies. Voorbeelde van tesourusartikels as tekste in kommunikatiewe onderrigaktiwiteite word bespreek. Vier tesourusverwante onderrigdoelwitte, naamlik woordeskatonderrig, woordeskatuitbreiding, onderrig van semantiese verhoudings en onderrig van taalvariante, word aan die hand van toepaslike tesourustekste geïllustreer.

Woordeskatonderrig vind in veral die Grondslagfase in die huistaal, maar ook addisionele taal, plaas, waar leerders oor ontluikende geletterdheidsvaardighede beskik. Woordeboeke gerig op die Grondslagfase word gewoonlik tematies of gedeeltelik tematies gestruktureer met verwante konsepte bymekaar gegroepeer, en verteenwoordig dus 'n tesourusafdeling in 'n woordeboek. Alhoewel aktiwiteite uit 'n formele tesourus nie in die Grondslagfase aanbeveel word nie, kan tematiese afdelings vir woordeskatuitbreiding gebruik word.

Tesourusafdelings in woordeboeke gerig op die Grondslagfase verwys gewoonlik na kurrikulumgerigte temas wat gereeld in die Grondslagfase gebruik word. Sien 'n illustrasie uit die *Oxford First Bilingual Dictionary — English + Afrikaans* as voorbeeld.



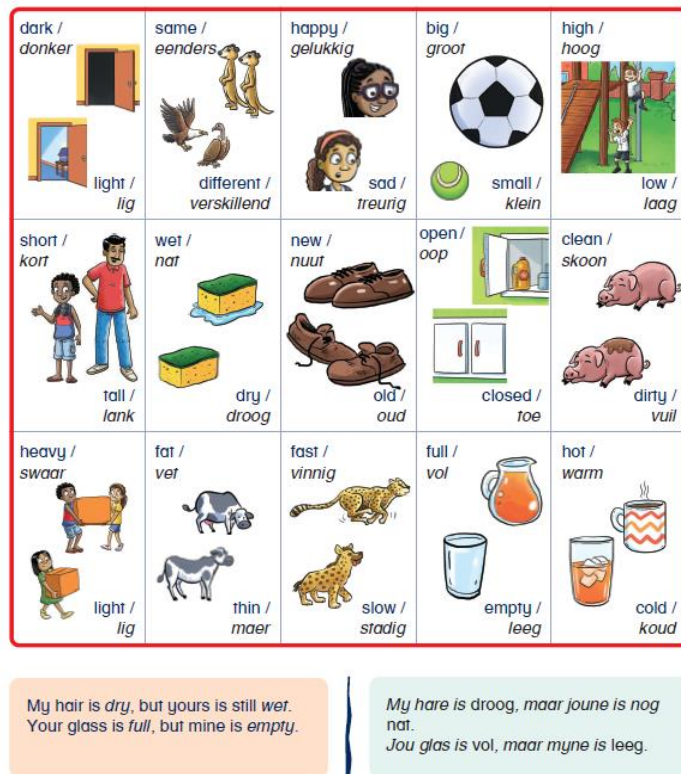
Figuur 1: At the aquarium/By die akwarium, uit *Oxford First Bilingual Dictionary — English + Afrikaans* (2021: 38)

'n Groot, kleurvolle beeldende illustrasie van 'n akwarium word verskaf. Aan weerskante van die bladsy word komponente wat in die beeldende illustrasie voorkom in twee tale benoem en van 'n prentjie voorsien. Woordeskatontwikkeling kan plaasvind met die verskaf van konsepte wat tematies verband hou met die akwarium, byvoorbeeld *dolfyn*, *paling*, *vin*, *kieu* en *seekat*. Dit kan impetus gee vir die ontwikkeling van woordeskatvaardighede met behulp die van inoefening van praat- en skryfvaardighede.

'n Sogenaamde raaiselvraag word telkens gevra, wat verseker dat leerders by die prent betrokke raak. In die betrokke toneel is die raaiselvraag: "Water soort seedier sou jy graag wou wees?" So 'n vraag laat leerders dink en 'n mening oor 'n saak lug, wat denkvaardighede, praatvaardighede en fonemiese bewustheid aanmoedig.

'n Bladsy word gewy aan konsepte wat in 'n antonimiese verhouding tot mekaar staan, byvoorbeeld *donker* teenoor *lig*, *eenders* teenoor *verskillend*, en *gelukkig* teenoor *treurig*. Konsepte word met behulp van verduidelikende en verhelderende illustrasies duidelik aan leerders oorgedra. Aan die onderkant van die bladsy word voorbeeldsinne in beide tale verskaf waar die gebruik van antonieme gedemonstreer word. Die onderwyser kan die bladsy op 'n teksgebaseerde wyse in die klaskamer gebruik om antonimie te demonstreer en te oefen.

Opposites / Teenoorgesteldes



45

Figuur 2: Opposites/Teenoorgesteldes, uit *Oxford First Bilingual Dictionary — English + Afrikaans* (2021: 45)

Volgens Antonacci en O'Callaghan (2011: 10) ontwikkel leerders se woordeskat vinnig in die Grondslagfase, maar vanaf graad 4, oftewel die Intermediêre Fase, begin leerders met 'n swakker woordeskat struikelblokke met leesbegrip en vakwoordeskat ervaar. Woordeskatuitbreiding speel dus 'n deurslaggewende rol in huistaalonderrig en addisionele taalonderrig. Tesourusgebruik word nie in die laerskool in kurrikulumdokumente (DBO 2011a, DBO 2011d) aanbeveel nie, maar soms is daar wel gepaste leksikografiese produkte om te gebruik in die onderrigkonteks. 'n Woordeboek soos die *HAT Afrikaanse Skoolwoordeboek*, waar innoverende leksikografiese praktyke soos die invoeg van tesouruskasies gebruik word, kan vir Afrikaans Huistaal vanaf graad 4 of Afrikaans Addisionele Taal vanaf graad 6 gebruik word. Vergelyk byvoorbeeld die lemma *kieskeurig* in hierdie verband:

kieskeurig *b.nw., bw.* [**kieskeuriger, die kieskeurigste**] **kieskeurig** wees moeilik wees om tevrede te stel ten opsigte van kos/klere. ens.: *Nkola is so kieskeurig; ek weet nie wat om hom vir sy verjaardag te gee nie. Ek is baie kieskeurig as dit by vis kom; ek maak altyd seker dis vars.*

TESOURUS

lekkerbekkig *My lekkerbekkige kat eet net vis.*
neusoptrekkerig *Die neusoptrekkerige gaste wou niks van Ma se afoal weet nie.*
puntene(u)rig *Dis moeilik om puntenerige eters tevrede te stel.*
uitsoekerig *Die uitsoekerige kleuter hou nie van groente nie.*
vol fiemies *Selfs die hond is vol fiemies.*

Figuur 3: Uittreksel uit *HAT Afrikaanse Skoolwoordeboek* (2009: 261)

Die gebruiker van die woordeboek vind 'n verklaring van die woord, maar bykomend ook 'n ingevoegde teks, die sogenaamde tesouruskassie, waar sinonieme vir die lemma, asook voorbeelde van die gebruik daarvan, verskaf word. Dit is tot voordeel van die gebruiker, veral die ontwikkelende woordeboekgebruiker in die laerskool, wanneer addisionele inligting rakende sinoniemgebruik verskaf word. So 'n hibriede karakter van 'n skoolwoordeboek, met alfabetiese sowel as tematiese bewerking, getuig van gebruikersvriendelikheid, waar die gebruiker se taalvaardigheidsvlak in ag geneem word.

Met teksgebaseerde tesourusgebruik kan die onderwyser dus in hierdie geval die onderrig van die betekenisverklarings van 'n woord, asook die sinonieme van 'n woord, as onderrigdoelwit verwesenlik.

Die onderwyser in die laerskool kan ook van 'n tesourus gebruik maak, maar moontlik met beperkte tesourustekste. Vergelyk byvoorbeeld die uittreksel onder die inskrywing *kossoort*, uit die aanlyn weergawe van TA:

426. Kossoort, dis

s.nw.

gereg, dis, kos, spys, nammies (*kindertaal*), nams (*kindertaal*), godespys, gourmetkos, fynproewerskos, fynproewersdis, huiskos, restaurantkos, restourantkos, boerekos, trooskos, tradisionele kos, tradisionele gereg, hoofdis, hoofskottel, voorgereg, nagereg, gang, hors d'oeuvre, stoofgereg, stowegereg, roostergereg, oondgereg, smoorkos, smoor, buffetete, smorgasbord, kruiekos, veldkos, wegneemete, wegneemkos, padkos, wegvattertjie, barakat (*lekties*), kitskos, gemorskos, lekkerny, peuselkos, snoepgereg, happie, proeseltjie, versnapering, smulhappie, verversing, snoephappie, southappie, soutigheid, kanapee, canapé, aartappelskyfie; konkoksie, brousel

Figuur 4: kossoort, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe)

Met teksgebaseerde tesourusgebruik kan die onderwyser woordeskatontwikkeling tydens kreatiewe skryfwerk as onderrigdoelwit verwesenlik deur leerders verskillende woorde uit die bogenoemde teks te laat gebruik. Voordat die verskillende woorde gebruik word, kan 'n klasbespreking oor die verskillende gebruike en betekenis van die tematies verwante woorde plaasvind.

Die volgende uittreksel uit die lemma *natuur* uit TA kan moontlik gebruik word vir woordeskatontwikkeling tydens kreatiewe skryfwerk in die hoërskool, aangesien dit uiteenlopende sake onder die spesifieke onderwerp in bestek het:

255. Natuur

s.nw.

natuur, natuurryk, diereryk, fauna 357, planteryk, flora 318, natuurlewe, natuurskoon, natuurlike hulpbronne, biosfeer, ekologie, ekosfeer, ekosisteem, biogenese, biogenesis, biodiversiteit, biodinamika, biomassa, bioom, mikrobioom, habitat, natuurlike habitat, natuurverskynsel, natuurwonder, natuurwet, natuurkrag, natuurmag, natuurfrats, natuurbeskouing, natuurbeskrywing; vrye natuur, buitelewe, buitelug

natuurmens, natuurvriend, natuurliefhebber, natuuraanbidder, natuurkind, buitemens; stroper, wildstroper

natuurbewaring, bewaring, omgewingsbewaring, omgewingsbewustheid, wildbewaring, waterbewaring, natuurbeskerming, ekologiebewaring, ekologiese balans, omgewingsimpak, omgewingsimpakstudie, natuurliefde, natuurresewaat, resewaat, natuurbewaringsgebied, wildtuin, wildpark, oorgrenspark, natuurtuin, botaniese tuin, biobank, natuurfonds, groen ekonomie, ekologiese voetspoor, koolstofvoetspoor, besoedelingsbeheer, rookbeheer, volhoubare ontwikkeling, ekologiese ontwikkeling, volhoubaarheid; herwinning, herwinningsaanleg, herwinningsveldtog, rehabilitasie, grondherwinning, grondrehabilitasie, plastiekerwinning, papierherwinning, glasherwinning, afvalherwinning, afvalverwerking, vullisverwerking; natuurbewaringsowerheid, natuurbewaarder, wildbewaarder, ekoloog; natuurskool, veldskool

natuurbewaringsaktiwisme, bewaringsaktiwisme, omgewingsaktiwisme, klimaatsaktiwisme; natuurbewaringsaktiwis, bewaringsaktiwis, omgewingsaktiwis, klimaatsaktiwis, groene, omgewingsdrukgroep, omgewingsvriend

natuurramp, ekologiese ramp, ekoverset, ekoprotes, ekovrees, omgewingsangs, omgewingstres, klimaatsverandering, klimaatsverskuiwing, klimaatsnoodtoestand, stropery, wildstropery, stroping, wildstroping, renosterstropery, renosterstroping, ..., droogte, waternood, watersnood, watertekort, oorstroming, vloedramp, aardverwarming, aardverhitting, koolstofvoetspoor, koolstofemissie, koolstofbelasting, ontbossing, oorbevolking; bedreigde spesie, rooidatalys

besoedeling 628, grondbesoedeling, lugbesoedeling, waterbesoedeling, oseaanbesoedeling, seebesoedeling, kusbesoedeling; geraasbesoedeling, koolstofbesoedeling, ligbesoedeling, oliebesoedeling

Figuur 5: natuur, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

Semantiese verhoudings soos sinonimie, antonimie, hiponimie en meronimie vorm deel van die wesenlike aard van 'n tesourus en die onderrig daarvan kan suksesvol met 'n tesourus bereik word aangesien dit woordbewustheid kan kweek. Soos reeds genoem, val die onderrig van semantiese verhoudings tussen woorde binne die bestek van Afrikaanse taalkurrikula. Die verskillende soorte semantiese betrekkinge tussen leksikale items word in die voorwoord van TA deeglik uiteengesit, in 'n verduideliking waarby die taalonderwyser, maar ook die leerder en ander moontlike gebruikers van die tesourus, sal baat. Volgens De Stadler en De Stadler (2021: voorwoord) kan verskillende soorte verhoudings binne hierdie struktuur onderskei word, waaronder veral die volgende vier:

- a. Sinonimie, die verhouding tussen twee woorde waarvan die betekenis min of meer dieselfde is, soos in die geval van *gierigheid*, *behaagsug* en *hebsug*.
- b. Teenoorgesteldheid, die verhouding wat bestaan tussen twee woorde waarvan die betekenis in opposisie tot mekaar staan, soos in die geval van *liefde* en *haat*.
- c. Hiponimie, die verhouding tussen 'n superordinate term (oorkoepelende term) en daardie terme wat as subordinate (of hiponime) dien. Dit is die verhouding wat bestaan tussen 'n woord soos *dier* en sy subordinate of hiponime *bees*, *skaap*, *koedoe*, *springbok*, *leeu*, *renoster*, *olifant*, ens.
- d. Die deel-geheel-verhouding (meronimie), 'n verhouding wat bestaan tussen woorde soos *voertuig* en *stuurwiel*.

Die bogenoemde teks rakende *natuur* kan as 'n voorbeeld dien vir teksgebaseerde onderrig van bovermelde semantiese verhoudinge. Indien die teks vanaf die begin bestudeer word, kan daar opgemerk word dat die woorde *diereryk* en *fauna* in 'n sinonimiese verhouding met mekaar staan, en hulle is dan ook met behulp van 'n komma geskei. Dieselfde geld vir die woorde *planteryk* en *flora*. Tussen *natuurryk* en *diereryk* bestaan 'n hiponimiese verhouding, maar dit word nie so vir die gebruiker aangedui nie. 'n Gebruiksetiket sou hier waardevol wees, veral vir die hoëskoolgebruiker.

Onder *natuurmens* word *stroper* en *wildstroper* as teenstelling tot *buitemens* en ander sinonieme verskaf. Dit word heeltemal korrek met 'n kommapunt aangedui en is 'n konvensie wat die gebruiker kan aanleer. Dit sou egter meer gebruikersvriendelik wees indien 'n etiket soos "antoniem" of "teenstelling" voorsien kon word.

natuurmens, natuurvriend, natuurliefhebber, natuuraanbidder, natuurkind, buitemens; stroper, wildstroper

Figuur 6: natuurmens, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

Tesourustekse leen hulle dus by uitstek tot die onderrig van semantiese betrekkinge tussen leksikale items, wat gekombineer kan word met verskeie lees- en skryfvaardighede, soos soeklees en kreatiewe skryfwerk.

Tesourustekste uit TA kan effektief gebruik word tydens die onderrig van taalvariante ook, alhoewel die verskaffing van taalvariante nie noodwendig as deel van die wesensaard van 'n tesourus gereken word nie. Dit wil voorkom asof die etiket "lekties" gebruik word om na taalvariante te verwys, maar die gebruik en bestek van die etiket word nie in TA verklaar nie. Met 'n elektroniese soektog is vasgestel dat die etiket "lekties" in die meeste gevalle na Kaapse Afrikaans verwys. Die etiket word 149 maal in TA gebruik. Die artikel *liefde en vriendskap* dien as voorbeeld waar talle woorde van etikette voorsien is, onder andere "lekties":

vriend, vriendin, boesemvriend, boesemvriendin, dik vriende (*meervoud*), maat, maatjie, kornuit, amigo (*informeel*), amice (*ongewoon*), pel (*informeel*), pël (*informeel*), pellie (*informeel*), pëllie (*informeel*), tjom (*informeel*), tjommie (*informeel*), tjomma (*informeel*), china (*lekties, informeel*), gabba (*lekties, informeel*), bra (*lekties, informeel*), brasse (*lekties, meervoud, informeel*), sielsvriend, geesgenoot, sielsgenoot, kameraad, bloedbroer, familievriend, gesinsvriend, huisvriend, jeugvriend, kindervriend, makker, lotsgenoot, vertroude, vertroueling, vertrouensman, dissipel, metgesel, gesel, alter ego, gunsteling, held, heldin, idool, allemansvriend; vriendekring, vriendskapskring, tweemanskap, driemanskap, ..., vriendskapsband, vriendskapsbetrekking, aanhang
gunsteling, oogappel, witbroodjie, pa se kind, ma se kind, masekin (*lekties*)

Figuur 7: vriendin, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe)

Met die bestudering van hierdie artikel kan interessante klasgesprekke rakende taalvariante in Afrikaans ontgin word, byvoorbeeld oor taalidentiteit, taalnorme en taalpersepsies. Ongelukkig gaan onderwysers en leerders eerstens nie weet dat taalvariante deur die tesourus aangedui word nie en tweedens heel moontlik nie daarvoor soek nie. Hier kort die gebruiker dus leiding deur 'n verklaring van die verskynsel en etiket, hoe dit in die tesourus gebruik word en hoe om daarvoor te soek.

Dit kan vir 'n onderwyser nuttig wees om 'n soektog na verskeie taalvariante, byvoorbeeld Overbergse Afrikaans, Noordwestelike Afrikaans, asook Swartlandse Afrikaans (wat nie onder taalvariante gelys word nie) te kan loods. Aanlyn soektogte na ander taalvariante as Kaapse Afrikaans het geen resultate opgelewer nie. Bykomende waardevolle soektogte kan ook na voorbeeldwoorde met etikette soos *sleng*, *figuurlike taalgebruik* of *verouderde taalgebruik* uitgevoer word, aangesien dit binne die onderrigraamwerk van die Afrikaanse vakkurrikulum val. Dit blyk tans nie moontlik te wees nie.

'n Etiket wat die samestellers moontlik kan oorweeg om by te voeg, is dié van idiomatiese uitdrukkings, aangesien dit 'n belangrike komponent van die onderrig van taalstrukture en -konvensies behoort uit te maak. Idiomatiese uit-

drukkings kan ook 'n rol speel in die woordeskatontwikkeling van taalaanleerders. Indien idiomatiese uitdrukkings duidelik gemerk is, kan die gebruiker (onderwyser en leerder) dit maklik identifiseer en taalkundig korrek daarmee omgaan.

Die gebruik van 'n tesourus in die taalklas kan met die gebruik van verklarende en vertalende woordeboeke in die huistaal- en addisionele taalklas gepaardgaan. Kundige onderwysers en gebruikersvriendelike woordeboeke kan tot voordeel van die gebruiker wees en sodoende bydra tot die skep van 'n woordeboek- (en tesourus-) kultuur in Suid-Afrika.

6. Gevolgtrekking

Ondersoek is ingestel na die moontlike gebruik van 'n tesourus in die taalonderrigkonteks van verskillende taalvlakke in Suid-Afrika. Die onontginde potensiaal van die tesourus in 'n taalonderrigkonteks op skool is in die artikel beklemtoon, met verwysing na die ontsluiting van moontlikhede vir onderrig.

Die doel en leksikografiese funksies van 'n tesourus is beskryf en teen die agtergrond van 'n taalonderrigkonteks gesitueer in 'n poging om die gaping in die literatuur te oorbrug. Konteks vir taalonderrig is verskaf, met verwysing na voorgeskrewe tesourusgebruik in skoolkurrikuladokumente, aangesien dit rigtinggewend en beleidmakend vir onderrigraamwerke in skole is. Teksgebaseerde tesourusgebruik word voorgestel en dit word geïllustreer teen die agtergrond van 'n ondersoek na 'n tesourus en tematiese woordeboekafdelings in woordeboeke. Met behulp van teksgebaseerde tesourusgebruik kan die onderwyser 'n tesourus as teks gebruik in die onderrig van taalstrukture en -konvensies.

Dit word sterk aanbeveel dat naslaanwerke soos woordeboeke en tesourusse as verpligte materiaal vir skole aangekoop word. Dit kan verseker dat onderrigdoelwitte, soos in die taalkurrikula gestipuleer, verwesenlik word indien onderwysers en leerders voldoende toegang tot woordeboeke en tesourusse het. Samewerking tussen die Departement van Basiese Onderwys, woordeboekuitgewers en nasionale woordeboekeenhede kan lei tot verhoogde en verbeterde gebruik van leksikografiese produkte. Dit kan moontlik lei tot 'n verbetering in die geletterdheidsvlak in Suid-Afrika.

Aanbevelings van enkele verbeterings tot leksikografiese produkte word vanuit 'n opvoedkundige leksikografiese perspektief gefundeer ten einde optimale gebruik in 'n taalonderrigkonteks te verseker. Die volgende aanbevelings vir *Tesourus van Afrikaans* geld indien die skoolkonteks ook as gebruikerskonteks verreken word:

- (1) Oorweeg 'n meer gebruikersvriendelike skooluitgawe gemik op leerders en onderwysers.
- (2) Ontwerp aanlyn kurrikulumgerigte aktiwiteite op alle taalvlakke gerig op tesourusgebruik.

- (3) Ontwerp aanlyn onderwysershandleidings gerig op tesourusgebruik.
- (4) Maak onderwysers, ouers en leerders op die sosiale media bewus van tesourusgebruik.
- (5) Verlewendig die bestaande webwerf met daaglikse bywerkings, byvoorbeeld "sinonieme van die dag".

Daar word aanbeveel dat soektogte in die aanlyn weergawe van TA verbeter word. Die gebruiker kry 'n aantal opsies om 'n soektog uit te voer, maar slegs een opsie lewer resultate, aangesien die gebruiker nie 'n antwoord kry wanneer die soekbevel "soek vir enige van hierdie woorde" gegee word nie, maar slegs as "presiese frase" as soekbevel gebruik word. Dit is inligting wat nie aan die gebruiker in die inleiding ("Die aard en gebruik van 'n tesourus") verskaf word nie. Indien 'n gebruiker dit nie toevallig agterkom nie, is daar geen waarborg dat 'n suksesvolle soektog uitgevoer kan word nie, want heel moontlik gee die gebruiker moed op na die eerste mislukte soektog.

Aangesien kurrikulumdokumente vir die onderrig van al die taalvakke op skool dieselfde beleid ten opsigte van taalonderrig in Suid-Afrika volg, kan die samestelling van 'n tesourus asook die bogenoemde aanbevelings vir nasionale woordeboekeenhede van toepassing wees op al die inheemse tale van Suid-Afrika.

Bibliografie

Woordeboeke

- De Stadler, L.G. en A. de Stadler.** 2021. *Tesourus van Afrikaans*. Kaapstad: Pharos (aanlyn).
- Luther, J.** 2009. *HAT Afrikaanse Skoolwoordeboek*. Kaapstad: Pearson Education.
- Odendal, F.F. en R.H. Gouws.** 2005. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson Education.
- Paizee, D., L. Roos en M. Janse van Rensburg.** 2021. *Oxford First Bilingual Dictionary — English + Afrikaans*. Kaapstad: Oxford University Press Southern Africa.

Ander bronne

- Antonacci, P.A. en C.M. O'Callaghan.** 2011. *Developing Content Area Literacy*. Los Angeles, CA: Sage.
- Augustosky, T. en B. Fastman.** 2021. Tools to Help You Write Clear Environmental Health Messages. *Journal of Environmental Health* 84(5): 30-31.
- Carstens, A.** 1995. Language Teaching and Dictionary Use: An Overview. *Lexikos* 5: 105-116.
- Chainikova, G.R., A.V. Zatonkiy, N.W. Mitiukov en H.L. Busygina.** 2018. Development of Foreign Language Lexical Competence on the Basis of a Learner's Terminological Thesaurus and Dictionary. *European Journal of Contemporary Education* 7(1): 51-59.

- Cruse, D.A., F. Hundsnurscher, M. Job en P.R. Lutzeier (Reds.).** 2002. *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. Handbücher zur Sprach- und Kommunikationswissenschaft 21(1). Berlin/New York: Walter de Gruyter.
- Cunningham, P.M. en R.L. Allington.** 2007. *Classrooms that Work: They Can All Read and Write*. Boston: Allyn & Bacon.
- Departement van Basiese Onderwys (DBO).** 2011a. *Kurrikulum- en assesseringsbeleidsverklaring. Afrikaans Huistaal: Graad 4–6*. Finale hersiening. Pretoria: Staatsdrukker.
- Departement van Basiese Onderwys (DBO).** 2011b. *Kurrikulum- en assesseringsbeleidsverklaring. Afrikaans Huistaal: Graad 7–9*. Finale hersiening. Pretoria: Staatsdrukker.
- Departement van Basiese Onderwys (DBO).** 2011c. *Kurrikulum- en assesseringsbeleidsverklaring. Afrikaans Huistaal: Graad 10–12*. Finale hersiening. Pretoria: Staatsdrukker.
- Departement van Basiese Onderwys (DBO).** 2011d. *Kurrikulum- en assesseringsbeleidsverklaring. Afrikaans Eerste Addisionele Taal: Graad 4–6*. Finale hersiening. Pretoria: Staatsdrukker.
- Departement van Basiese Onderwys (DBO).** 2011e. *Kurrikulum- en assesseringsbeleidsverklaring. Afrikaans Eerste Addisionele Taal: Graad 7–9*. Finale hersiening. Pretoria: Staatsdrukker.
- Departement van Basiese Onderwys (DBO).** 2011f. *Kurrikulum- en assesseringsbeleidsverklaring. Afrikaans Eerste Addisionele Taal: Graad 10–12*. Finale hersiening. Pretoria: Staatsdrukker.
- Durkin, P. (Red.).** 2016. *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press.
- Geckeler, H.** 2002. Anfänge und Ausbau des Wortfeldgedankens. Cruse, D.A., F. Hundsnurscher, M. Job en P.R. Lutzeier (Reds.). 2002: 713-728.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: African Sun Media.
- Granger, S. en M. Paquot (Reds.).** 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Graves, M.F.** 2006. *The Vocabulary Book: Learning & Instruction*. New York: Teachers College Press.
- Hausmann, F.J. et al. (Reds.).** 1989–1991. *Wörterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography*. Berlin: Walter de Gruyter.
- KABV.** Sien Departement van Basiese Onderwys (DBO).
- Kay, C. en M. Alexander.** 2016. Diachronic and Synchronic Thesauruses. Durkin, P. (Red.). 2016: 367-392.
- Kilgarriff, A. en C. Yallop.** 2000. What's in a Thesaurus? *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00), May 2000, Athens, Greece*. European Language Resources Association (ELRA): 1371-1379.
Besikbaar: <https://aclanthology.org/L00-1134/> [26 April 2022].
- Kremer, G., A. Abel en M. Baroni.** 2008. Cognitively Salient Relations for Multilingual Lexicography. Zock, M. en C.-R. Huang (Reds.). 2008: 94-101.
- Lawrence, D., M. le Cordeur, L. van der Merwe, C. van der Vyver en R. van Oort.** 2014. *Afrikaans-metodiek deur 'n nuwe bril*. Kaapstad: Oxford University Press.
- Marelo, C.** 1989. The Thesaurus. Hausmann, F.J. et al. (Reds.). 1989–1991: 1083-1094.
- Meara, I.S.P.** 1996. The Dimensions of Lexical Competence. Brown, G., K. Malmkjaer en J. Williams (Reds.). 1996. *Performance and Competence in Second Language Acquisition*: 35-53. Cambridge: Cambridge University Press.
- Mountain, L.** 2007. Synonym Success: Thanks to the Thesaurus. *Journal of Adolescent & Adult Literacy* 51(4): 318-324.
- Nkomo, D.** 2015. Developing a Dictionary Culture Through Integrated Dictionary Pedagogy in the Outer Texts of South African School Dictionaries. *Lexicography* 2(1): 71-99.

- Richards, J.C.** 1976. The Role of Vocabulary Teaching. *TESOL Quarterly* 10(1): 77-89.
- Stockdale, J.G.** 2019. *A Dictionary and Thesaurus of Contemporary Figurative Language and Metaphor*. Aanlyn voorlegging.
Beskikbaar: <https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED594968&site=ehost-live&scope=site> [30 Julie 2022].
- Tamor, L. en S. Walmsley.** 1983. Using a Thesaurus in the Elementary School: Some Cautions. *Language Arts* (60)5: 554-57.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tasovac, T.** 2012. Potentials and Challenges of WordNet-based Pedagogical Lexicography: The Transpoetika Dictionary. Granger, S. en M. Paquot (Reds.). 2012: 237-258.
- Van der Merwe, M.F.** 2009. Die opvoedkundige waarde van woordeboeke: Voorstelle vir woordeboekonderrig in Suid-Afrika. *Lexikos* 19: 297-313.
- Van Schalkwyk, C.** 1996. Resensie: De Stadler, L.G. en Amanda de Stadler. *Groot Tesourus van Afrikaans*. *Lexikos* 6: 283-293.
- Zock, M. en C.-R. Huang (Reds.).** 2008. *COLING 2008: Proceedings of the Workshop on Cognitive Aspects of the Lexicon (COGALEX 2008), 24 August 2008, Manchester, United Kingdom*. Stroudsburg, PA, USA: Association for Computational Linguistics [Intyds].
Beskikbaar: <https://dl.acm.org/action/showBook?doi=10.5555/1598848>.

Die Behandlung der Wortarten in Wörterbüchern zur Lexikographie

Monika Bielińska, *Uniwersytet Śląski w Katowicach, Polen,*
(monika.bielinska@us.edu.pl)

Zusammenfassung: Ziel des Beitrags ist es, zu beschreiben und zu analysieren, wie drei Wörterbücher zur Lexikographie, nämlich *Lexicography. A Dictionary of Basic Terminology* (1998) von Igor Burkhanov, *Dictionary of Lexicography* (2001) von Reinhard R.K. Hartmann und Gregory James und *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* (2010–2020) von Herbert Ernst Wiegand et al., Wortarten behandeln. Es wird untersucht, ob die entsprechenden Einträge nur grammatische oder auch lexikographische Informationen enthalten, und ob sie mit anderen Artikeln mit lexikographiebezogenem Inhalt vernetzt sind. Wortart-Artikel mit Informationen zur Lexikographie werden anschließend unter qualitativem Aspekt erörtert. Der Artikel schließt mit einigen allgemeinen Bemerkungen zur Beschreibung von lexikographie-externen Begriffen in Wörterbüchern zur Lexikographie. Der Autor postuliert, dass sowohl die Wortart-Artikel als auch alle anderen Einträge zu Termini von außerhalb der Lexikographie, z.B. aus dem Bereich der Linguistik, lexikographische Informationen enthalten. Es ist nämlich nicht die Funktion dieser Artikel, analoge Artikel aus Fachwörterbüchern zu betreffenden Disziplinen zu ersetzen und den gleichen Inhalt zu bieten. In Wörterbüchern zur Lexikographie sollte stets die Lexikographie im Mittelpunkt stehen und die Beschreibungsperspektive prägen.

Schlüsselwörter: LEXIKOGRAPHIE, LINGUISTIK, LINGUISTISCHE TERMINOLOGIE, LEXIKOGRAPHISCHE TERMINOLOGIE, GRAMMATIK, WORTARTEN, FACHWÖRTERBUCH, WÖRTERBUCH ZUR LEXIKOGRAPHIE, WÖRTERBUCH-EVALUATION

Abstract: The Treatment of Parts of Speech in Dictionaries of Lexicography.

The aim of the article is to present and analyse how three dictionaries of lexicography, i.e. *Lexicography. A Dictionary of Basic Terminology* (1998) by Igor Burkhanov, *Dictionary of Lexicography* (2001) by Reinhard R.K. Hartmann and Gregory James and *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* (2010–2020) by Herbert Ernst Wiegand et al., treat parts of speech. It is examined whether the entries contain only grammatical or also lexicography-related information. Then, the interconnectedness of the articles on parts of speech and other lexicographic content is scrutinised. Subsequently, entries with lexicography-related information are examined qualitatively. The article concludes with some general remarks on the description of lexicography-external terms in dictionaries of lexicography. The author postulates that both the part-of-speech articles and all other entries on terms from outside lexicography, e.g. from the field of linguistics, contain lexicographic information. It is not the function of these articles to replace analogous articles from specialised dictionaries on the disciplines concerned, and to offer the same

content. In dictionaries on lexicography, lexicography should always be the focus and shape the descriptive perspective.

Keywords: LEXICOGRAPHY, LINGUISTICS, LINGUISTIC TERMINOLOGY, LEXICOGRAPHICAL TERMINOLOGY, GRAMMAR, PART OF SPEECH, SPECIALISED DICTIONARY, DICTIONARY OF LEXICOGRAPHY, DICTIONARY EVALUATION

1. Einleitendes

Die Lexikographie ist eine Geber- und Nehmerdisziplin, die auf vielerlei Art und Weise mit anderen vernetzt ist. Einen höchst interdisziplinären Charakter weisen sowohl die lexikographische Praxis — in Wörterbüchern und Enzyklopädiën werden zahlreiche Wissensbereiche und Betätigungsfelder des Menschen dargestellt — als auch die Metalexikographie auf. Die Lexikographie ist nicht nur von Disziplinen abhängig, die ihren Gegenstandsbereich bilden, wie z.B. Sprachwissenschaft, Biologie, Geschichte oder Physik, sondern auch von Wissenschaftszweigen, deren Forschungsmethoden und -ergebnisse sie nutzt, wie etwa Informatik und Soziologie. Da Sprachwörterbücher einen zentralen Platz in der Lexikographie einnehmen, verbindet die Lexikographie mit der Sprachwissenschaft ein besonders enges Verhältnis (vgl. einen Überblick in Bielińska 2019: 115-126). Lexikographische Beschreibung in Sprachwörterbüchern basiert auf linguistischem Wissen, in Wörterbüchern und metalexikographischen Arbeiten wird oft linguistische Terminologie verwendet.

Der hier geschilderte Umstand hat zur Folge, dass Spezialwörterbücher zur Lexikographie außer lexikographiespezifischen Termini solche verzeichnen können, die aus anderen Disziplinen stammen und dass dabei linguistische Terminologie einen beträchtlichen Teil der Stichwörter ausmachen kann. Die Entscheidung über die Aufnahme linguistischer Termini in den Lemmabestand der Wörterbücher zur Lexikographie, die keinesfalls als selbstverständlich gilt (es können auch "nur rein lexikographische Termini" (Bergenholtz 1994: 323) als Lemmata angesetzt werden), zieht weitere konzeptionelle Entscheidungen nach sich. Was die Selektion von linguistischen Termini betrifft, nennt Bergenholtz (1994) im Zusammenhang mit der Erarbeitung von "Nordisk leksikografisk ordbok" folgende nach diversen Kriterien zusammengestellte Gruppen von Termini als Lemmakandidaten:

- (1) "lexikographische Termini und andere linguistische Termini, die in der Lexikographie anders als in der Linguistik sonst üblich gebraucht werden"
- (2) "lexikographische Termini und andere linguistische Termini, die besonders wichtig für die Lexikographie sind"
- (3) "Wörter, von denen anzunehmen ist, daß sie besonders regelmäßig in lexikographischen Texten vorkommen" und
- (4) "alle Wörter oder Wörter, die mit einer gegebenen Mindestfrequenz in lexikographischen Texten auftreten" (Bergenholtz 1994: 323).

Diesbezügliche Entscheidungen sowie alle weiteren Entscheidungen im Hinblick auf die Behandlung von linguistischen Begriffen sind mit anderen Grundelementen der Konzeption des jeweiligen Nachschlagewerks verbunden, d.h. mit dem geplanten Umfang, der Beschreibungstiefe, den Funktionen des Wörterbuches und den Situationen, in denen es benutzt wird, der Wörterbuchbasis sowie mit dem Profil der anvisierten Benutzer, also mit deren Wissen, Kompetenzen und Bedürfnissen.

Eine dem Komplexitätsgrad und dem Stellenwert des Problems gebührende Erforschung der Behandlung von sprachwissenschaftlichen Phänomenen in Wörterbüchern zur Lexikographie ist noch ein Desiderat. Sie kann jedoch nur im Rahmen eines größer angelegten Vorhabens erfolgen. Für die Zwecke dieses Beitrags muss eine Eingrenzung vorgenommen werden. Zum Untersuchungsgegenstand wird die lexikographische Behandlung eines Teils der Grammatik gewählt, und zwar der Wortarten. Die Einteilung des Wortschatzes in Wortarten spielt in der Lexikographie eine wichtige Rolle, weil die Wortartzugehörigkeit der Lexeme deren Beschreibung in Wörterbüchern determiniert. Im Folgenden werden also ausgewählte Nachschlagewerke unter dem genannten Aspekt analysiert, wobei der Schwerpunkt auf lexikographischen¹ Informationen liegt und der grammatische bzw. linguistische Teil der Beschreibung von Wortarten weitgehend ausgeklammert bleibt.

2. Allgemeines zur Linguistik/Grammatik in ausgewählten Wörterbüchern zur Lexikographie

2.1 Wörterbücher zur Lexikographie bilden eine relativ kleine Gruppe. Außer einigen Glossaren und einem bezogen auf den Umfang sehr bescheidenen schwedischen Wörterbuch „En liten ordbok om ordböcker“ (1990) von Lillemor Swedborg sind hier folgende Nachschlagewerke zu nennen: „Cishuxue cidian“ (1992) von Zuxi Yang und Qingkai Xu, „Diccionario de lexicografía práctica“ (1995) von José Martínez de Sousa, „Nordisk leksikografisk ordbok“ (1997) von Henning Bergenholtz et al., „Dictionary of Lexicography“ (1998, 2001) von Reinhard R.K. Hartmann und Gregory James, „Lexicography. A Dictionary of Basic Terminology“ (1998, 2010) von Igor Burkhanov, „Leksykografia. Słownik specjalistyczny“ (2020) hg. von Monika Bielińska und „Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research“ (2010–2020) von Herbert Ernst Wiegand et al.

Aus sprachlichen Gründen werden in der Untersuchung nur Wörterbücher mit Einträgen auf Englisch berücksichtigt, d.h. „Lexicography. A Dictionary of Basic Terminology“ (1998) (weiter: LDBT), „Dictionary of Lexicography“ (2001) (DoL), und das mehrsprachige „Wörterbuch zur Lexikographie ...“ (WLWF), das — abgesehen von Äquivalenten in sieben bzw. acht Sprachen² — deutsch- und englischsprachige Lemma- und Definiensangaben enthält³.

2.2 Die gewählten Wörterbücher unterscheiden sich voneinander in so gut wie allen Aspekten der lexikographischen Konzeption, darunter sehr beträchtlich in ihrem Umfang, was sie jedoch in Hinsicht auf die bereits erwähnte Zielsetzung nicht unvergleichbar macht. Das LDBT umfasst 400 autonome Artikel und 103 Verweisartikel, das DoL — entsprechend 1225 und 709 (Angaben nach Mann 2013: 446) und das fünfbändige WLWF — über 5500 bzw. etwa 7000. Alle drei Lexika verzeichnen linguistischen Fachwortschatz, wobei auch diesbezüglich die zahlenmäßigen Differenzen zwischen ihnen sehr gravierend sind. Selbst wenn man nur die grammatische Terminologie in Betracht zieht, lassen sich erhebliche Unterschiede feststellen. Einer vorläufigen Schätzung zufolge werden in den analysierten Wörterbüchern insgesamt über 120 grammatische Termini als Lemmata angesetzt. Sie sind allerdings nicht gleichmäßig auf die drei Wörterbücher verteilt. Nur über 20 Termini treten gleichzeitig in allen drei Werken auf, ebenfalls über 20 Termini kommen lediglich in zwei Lexika vor, und über 80 Termini, also etwa zwei Drittel, werden nur in einem der Wörterbücher lexikographisch erfasst. Dabei ist allerdings zu beachten, dass sich die genannten Zahlen nur auf die selbständigen Artikel beziehen und wegen unterschiedlicher Datenverteilung keine zuverlässigen Rückschlüsse auf das in den einzelnen Wörterbüchern dargestellte Fachwissen erlauben. So werden beispielsweise das Präfix und das Suffix in zwei Wörterbüchern in separaten Einträgen beschrieben und im DoL zusammen im Artikel zu *affix*.

Anders verhält es sich, wenn es um die lexikographische Behandlung der Wortarten geht. Zwar werden in Grammatiken abhängig vom Modell verschiedene Klassifikationen erstellt, in den meisten Sprachwörterbüchern für ein breites Publikum wird aber eine traditionelle Einteilung samt traditionellen Bezeichnungen für Wortarten bevorzugt, was sich in den Wörterverzeichnissen der Wörterbücher zur Lexikographie widerspiegelt. Trotz beachtlicher Unterschiede in der Lemmaanzahl sind das LDBT, das DoL und das WLWF in Bezug auf die Aufnahme dieses Teils der grammatischen Terminologie ähnlich. Obwohl der gemeinsame Teil der registrierten linguistischen und auch nur der grammatischen Terminologie prozentual gering ist, decken sich die Listen der aufgenommenen Wortartbezeichnungen in hohem Grade. Darüber hinaus wird jede Wortart in einem separaten Eintrag behandelt, sodass hier — abgesehen von den Einträgen zu den einzelnen Partikeltypen — keine durch unterschiedliche Parzellierung der Daten verursachten Probleme auftreten. Die genannten Faktoren bilden einen geeigneten Ausgangspunkt für die geplante Analyse.

2.3 Die Metatexte im LDBT, im DoL und im WLWF vermitteln einige Informationen, die im Hinblick auf die Analyse von Belang sind. Zum Adressatenkreis und zur Benutzung des Wörterbuchs steht im LDBT Folgendes:

This dictionary is primarily intended for undergraduate and postgraduate students of Linguistics Departments at Philological Faculties and provides an overview of lexicography and its correlation with relevant linguistic disciplines. It

may be particularly useful in the seminars and special courses on lexicography, lexicology and linguistic semantics⁴ (S. 9).

Etwas verwunderlich ist dabei, dass von Adressaten, d.h. von Linguistikstudenten, kein linguistisches (Vor)wissen erwartet wird („No prior knowledge of linguistics or lexicography is required.“ (ebd.)). Im Vorwort werden auch Funktionen des Wörterbuchs genannt. Eine davon ist „to specify the concepts which have been borrowed from other linguistic and non linguistic academic and applied disciplines and note the modifications which these notions have undergone within the framework of lexicographic description“⁵ (S. 8). Der Verfasser geht explizit auf die Zusammenhänge zwischen der Lexikographie und der Sprachbeschreibung ein und begründet damit die Darstellung linguistischer Begriffe im Wörterbuch. Er betont jedoch, dass diese vom Standpunkt der lexicographischen Theorie und Praxis aus betrachtet werden (vgl. S. 8f.).

Weniger informativ in Bezug auf die Zielsetzung des Beitrags sind die Metatexte des DoL. In den Benutzungshinweisen werden beiläufig Wörterbuchadressaten genannt: „a practising lexicographer, a dictionary researcher, an expert in a neighbouring discipline or a student“⁶ (S. xvii). Einer Passage zur äußeren Selektion ist zu entnehmen, welche Schritte bei der Auswahl unternommen, darunter auch nach welchen Kriterien bestimmte Termini ausgeklammert worden sind. Nicht berücksichtigt werden demnach u.a. „terms with no immediate direct relevance to the general theme, for example from descriptive linguistics (e.g. *aphesis*, *double-headed nominal group*) or printing (e.g. *bullet*, *font*), or names of languages and regionalisms“⁷ (S. xv). Die Behandlung aufgenommener linguistischer Termini wird nicht thematisiert.

Das WLWF teilt seine Adressaten in zwei Gruppen auf. Zur ersten zählt es Lexikographen, Wörterbuchforscher und Hersteller lexicographischer Produkte aller Typen und zur zweiten alle, die das Wissen und Kompetenzen der Mitglieder der ersten Gruppe erlernen (möchten) oder lehren sowie Übersetzer metalexicographischer Texte (WLWF, Bd. 1, S. IX-X). Der Vorspann des ersten Bandes enthält ein „alphabetisches Verzeichnis der Lemmazeichen aus anderen Fachgebieten“ (S. LXV-LXIX). Es zeigt sehr deutlich, dass die Lexikographie mit vielen anderen Disziplinen — die Linguistik nimmt hier einen wichtigen Platz ein — verbunden ist, und informiert zugleich die Benutzer, in welchem Ausmaß das WLWF diesen Verbindungen Rechnung trägt.

3. Analyse der Behandlung von Wortarten in Wörterbüchern LDBT, DoL und WLWF

3.1 Das zu analysierende Material wurde direkt den Wörterverzeichnissen entnommen. Im Falle des mehrbändigen, monumentalen WLWF lag es nahe, sich auf den Umtext „Alphabetisches Verzeichnis der Lemmazeichen aus anderen Fachgebieten“ zu verlassen, dies hat sich aber als unmöglich erwiesen. Einer-

seits enthält das Verzeichnis Termini, zu denen es im Artikelteil keine Einträge gibt, andererseits fehlen im Verzeichnis mehrere im Artikelteil behandelte Termini. Unter den übergangenen Termini aus dem Bereich der Linguistik kann man zum Beispiel folgende nennen: *Allomorph*, *Norm*, *Phonem*, *Phonologie*, *Semiotik*, *Signifikat*, *Stilistik*, *synchronisch*, *Tempus*, *Text* sowie zwei Wortartbezeichnungen: *Adverb* und *Interjektion*.

3.2 Den Gegenstand der Analyse bildet die lexikographische Beschreibung von dreizehn Wortarten: Adjektiv, Adverb, Artikel, Artikelwort, Interjektion, Konjunktion, Numerale, Partikel, Postposition, Präposition, Pronomen, Substantiv und Verb sowie die Einträge zum Lemma *Wortart/part of speech*.

Der Status der genannten Klassen von Wörtern ist allerdings unterschiedlich. Neben den Substantiven und den Verben, die in den Grammatiken relativ einheitlich als Wortarten aufgefasst werden, gibt es hier z.B. Pronomina, die wegen ihrer unterschiedlichen syntaktischen Eigenschaften verschieden zugeordnet und beschrieben werden, oder Präpositionen und Interjektionen, die entweder separat oder als Subklasse der Partikeln behandelt werden. Noch anders verhält es sich z.B. im Falle der Numeralia und Artikelwörter. Numeralia gelten traditionell als eine Wortart, aber sie stellen im morphosyntaktischen Sinne keine Wortart dar, was sich auch in ihrer lexikographischen Beschreibung widerspiegelt. Neben lexikalischen Einheiten, die mit der Wortartangabe *Zahlwort* versehen sind, kommen als Lemmata in Sprachwörterbüchern viele Numeralia vor, die den Adjektiven, Adverbien, Substantiven oder Pronomina (Indefinitpronomina) zugeordnet werden. Zu den Artikelwörtern dagegen gehören die Artikel im engeren Sinne, d.h. lexikalische Einheiten, die traditionell zur Wortart Artikel zusammengefasst werden, sowie verschiedene Pronomina. Da aber in Sprachwörterbüchern auch diese Sammelbezeichnung als Wortartangabe verwendet wird⁸, werden in der folgenden Untersuchung ebenfalls die Einträge zu *Artikelwort/determiner*⁹ berücksichtigt.

Die unten stehende Tabelle enthält Termini, die in den zu analysierenden Lexika als Lemmata angesetzt werden. Das Fehlen eines Terminus (sowie des entsprechenden Artikels) in einem Wörterbuch wird mit einem Minuszeichen markiert. Termini, die als Verweislemmata fungieren, z.B. *Nomen* (→ *Substantiv*), *Zahlwort* (→ *Numerale*), *Zeitwort* (→ *Verb*) (WLWF); *word class* (→ *part of speech*) (DoL), werden in die Tabelle nicht aufgenommen. Als Verweislemma wird auch *grammatical class* (Bdt. 1) (LDBT) betrachtet, obwohl es im Artikel außer der Information, dass die Bezeichnung als synonymisch zu *part of speech* gilt, wenige zusätzliche Angaben gibt. Ausgeklammert bleiben auch Bezeichnungen für Verbtypen: *Hilfsverb/auxiliary verb*, *Modalverb/modal verb*, *phrasal verb* und *Partikelverb/particle verb* sowie für (nur im WLWF genannte) Partikeltypen: *Gradpartikel/quantifying particle*, *Modalpartikel/modal particle* und *Gesprächswort/conversation word* (darauf wird vom Artikel zu *Gesprächspartikel* verwiesen).

Lemmata [englisch/deutsch]	LDBT	DoL	WLWF
1. adjective/Adjektiv	x	x	x
2. adverb/Adverb	x	x	x
3. article/Artikel	x	x	x
4. conjunction/Konjunktion	x	x	x
5. determiner/Artikelwort	-	x	x
6. interjection/Interjektion	x	-	x
7. noun/Substantiv	x	x	x
8. numeral/Numerale	x	-	x
9. part of speech/Wortart	x	x	x
10. particle/Partikel	x	-	x
11. postposition ¹⁰	-	x	-
12. preposition/Präposition	x	x	x
13. pronoun/Pronomen	x	x	x
14. verb/Verb	x	x	x

Die Wortart allgemein sowie acht einzelne Wortarten werden in allen drei Wörterbüchern beschrieben. Vier weiteren Wortarten werden nur in zwei Nachschlagewerken eigene Einträge gewidmet, und eine Wortart — die Postposition — wird lediglich in einem der Lexika behandelt.

In Bezug auf die untersuchten Wörterbücher gestaltet sich dies folgendermaßen: Im WLWF fehlt nur ein Eintrag zu *Postposition/postposition*. Im LDBT fehlt außer einem Eintrag zu *postposition* einer zu *determiner*, was angesichts des bescheidenen Umfangs des Wörterbuchs, eines etwas anderen Status dieser Klasse sowie der Aufnahme des Terminus *article* nachvollziehbar ist. Im DoL fehlen drei Einträge, und zwar zu *interjection*, *numeral* und *particle*. Das ist schwer zu begründen, denn das Wörterbuch enthält dreimal so viele autonome Artikel wie das LDBT. Der übergangene Terminus *particle* wird selbst im DoL verwendet (vgl. den Eintrag zu *phrasal verb*). Hinzu kommt, dass das DoL als einziges von den untersuchten Wörterbüchern einen Eintrag zu *postposition* enthält, also einer viel selteneren Wortartbezeichnung.

3.3 Artikel, die Wortarten beschreiben, fallen hinsichtlich ihres Umfangs in keinem der untersuchten Wörterbücher auf. Im LDBT sind zwar viele Einträge zu lexikographischen Begriffen länger als Wortart-Artikel, aber es kommen auch kürzere vor, z.B. die Einträge zu *bilingual lexicography*, *dictionary criticism* und *dictionary definition*. Im DoL, das viel kürzere Einträge als die übrigen Wörterbücher¹¹ enthält, gehören die uns interessierenden Artikel zu den kürzesten. Das ist aber nichts Spezifisches für diese Artikelgruppe. Unter den Artikeln zu

lexikographischen Begriffen finden sich nämlich auch viele mit vergleichbarem Umfang (hier jeweils ohne bibliographische Angaben):

verb — „A part of speech which serves to express an action or state.“

noun — „A part of speech which serves to name or designate entities.“

adverb — „A part of speech which serves primarily to modify a verb, adjective or other adverb.“

numerical information — „An information category based on figures and statistics.“

place-names dictionary — „A reference work which provides information on toponyms.“

valency dictionary — „A specialised grammar dictionary devoted to information on valency relations.“¹²

Auch im WLWF unterscheiden sich die Wortart-Artikel nicht in signifikanter Weise von den übrigen. Ähnlich wie im Falle der Artikel aus anderen lexikographisch erfassten Bereichen variiert der Umfang der untersuchten Einträge beachtlich. So umfasst z.B. der Eintrag zu *Verb/verb* etwa eine halbe Spalte¹³, Einträge zu *Adjektiv/adjectiv*, *Artikel/article* und *Interjektion/interjection* — ca. eine Spalte, Einträge zu *Artikelwort/determiner*, *Konjunktion/conjunction*, *Präposition/preposition* und *Substantiv/noun* — ca. 1,5 Spalten und der Artikel zu *Wortart/part of speech* — fast drei Spalten.

Zusammenfassend lässt sich also feststellen, dass die Wortart-Artikel im Hinblick auf den Umfang den übrigen Artikeln gleichrangig sind. Obwohl sie nicht zu den lexikographische Begriffe beschreibenden Artikeln gehören, heben sie sich von diesen umfangsmäßig nicht ab. Ihr Umfang deutet nicht darauf hin, dass sie in den genannten Lexika eine marginale Rolle spielen und lediglich eine ergänzende (Hilfs-)Funktion haben.

3.4 Nun gilt es also in weiteren Analyseschritten zu untersuchen, ob die Wortarten nur aus linguistischer oder auch — und falls ja, in welchem Grade — aus lexikographischer Sicht beschrieben werden. Ferner wird auch herauszufinden sein, ob die Artikel mediostrukturell in einen lexikographischen Kontext eingebettet sind, d.h. ob sie durch Verweise mit lexikographischem Inhalt in anderen Artikeln verbunden sind. Von besonderer Relevanz ist das im Falle von Wortart-Artikeln mit ausschließlich linguistischer Beschreibung, weil sie nur dank ihrer Vernetzung einen teilweise lexikographischen Charakter annehmen können, was wiederum ihre Funktion im Wörterbuch beeinflusst.

Unter zwölf Artikeln im LDBT findet sich nur einer, der einen grammatischen Begriff auch von lexikographischer Seite beleuchtet. Im ca. anderthalb Seiten langen Eintrag zu *part of speech* wird etwa die Hälfte Informationen zur lexikographischen Behandlung der Wortarten gewidmet. Ähnlich wird im DoL verfahren. Der einzige Eintrag mit lexikographischen Informationen ist ebenfalls der zu *part of speech*. Allerdings fällt der betreffende Teil dürftig aus, denn

der gesamte DoL-Artikel ist sehr kurz. Die Konzeption des WLWF ist das Gegenteil des bereits Geschilderten. Alle Wortart-Artikel enthalten Informationen zur Lexikographie. Dabei variiert ihr Anteil am gesamten Artikeltext erheblich. In vier Einträgen (*Adverb/adverb*, *Artikel¹/article*, *Artikelwort/determiner*, *Partikel/particle*) nehmen die lexikographischen Informationen einen Bruchteil des gesamten Textes ein, in vier weiteren (*Adjektiv/adjective*, *Konjunktion/conjunction*, *Substantiv/noun*, *Pronomen/pronoun*) — etwa je ein Drittel. Zwei Einträge (*Numerale/numeral*, *Wortart/part of speech*) bestehen zu über 50% aus lexikographischen Informationen und in drei (*Interjektion/interjection*, *Präposition/preposition*, *Verb/verb*) machen diese Informationen den überwiegenden Teil des Artikels aus.

Bei der Analyse der mediostrukturellen Vernetzung der Wortart-Artikel werden nicht nur Verweise auf Artikel zu lexikographischen Begriffen berücksichtigt, sondern auch auf solche zu linguistischen Begriffen, die aber u.a. lexikographische Probleme ansprechen.

Unter den untersuchten Artikeln im LDBT gibt es nur einen, der mit lexikographischem Inhalt stark vernetzt ist. Der Eintrag zu *part of speech* enthält Verweise auf Artikel zu *part of speech label*, *lexicographic indicator*, *lemmata*, *word list*, *lexicography*, *general dictionary*, *specialized dictionary*, *dictionary user*, *lexicographer*, *intended user* sowie auf einen Artikel zu *grammatical description*, der auch Informationen zur Lexikographie enthält. In den sonstigen Einträgen gibt es nur Verweise auf die Artikel zu *part of speech label* und zu *part of speech*.

Auch im DoL ist der Artikel zu *part of speech* derjenige mit den meisten Verweisen auf Artikel mit lexikographischem Inhalt. Es wird hier auf Einträge zu *general dictionary*, *grammar dictionary*, *usage guide*, *grammatical code*, *grammatical information* und *grammatical label* verwiesen. Artikel zu *adjective*, *adverb*, *conjunction*, *postposition*, *preposition* und *pronoun* enthalten Verweise auf die Artikel zu *part of speech* und zu *grammatical information*. Die Einträge zu *verb* und zu *noun* enthalten darüber hinaus Verweise auf (entsprechend) *conjugation* und *declension*. Vom Eintrag zu *article* wird nur auf *grammatical information* verwiesen, und der Eintrag zu *determiner* enthält keine Verweise auf Artikel, die lexikographische Informationen enthalten.

Die Wortart-Artikel des WLWF sind wesentlich besser sowohl mit Artikeln zu lexikographischen Begriffen vernetzt (s. unter 1 im folgenden Verzeichnis) als auch mit solchen, die aus lexikographischer Perspektive linguistische Begriffe beschreiben (s. unter 2):

Adjektiv/adjective — 1. Adjektivartikel, Adjektivlemma, Adjektivvalenzwörterbuch 2. Wortart

Adverb/adverb — 1. Adverbartikel 2. Adjektiv, Partikel, Wortart

Artikel¹/article — 1. Artikelangabe, Morphologieangabe bei Substantiven¹⁴
2. Wortart

Artikelwort/determiner — 1. Artikel¹, Wortartangabe 2. Wortart

Interjektion/interjection — 1. Onomatopöienwörterbuch 2. Gesprächswort, Onomatopöie, Partikel

Konjunktion/conjunction — 1. — 2. Partikel, Wortart

Numerale/numeral — 1. — 2. Wortart

Partikel/particle — 1. Partikelwörterbuch 2. Gesprächswort, Gradpartikel, Konjunktion, Modalpartikel, Präposition, Pronomen, Synsemantikon, Wortart

Präposition/preposition — 1. Präpositionsangabe, Präpositionswörterbuch, Präposition-Teilartikel 2. Partikel, Rektion

Pronomen/pronoun — 1. Pronomenangabe, Pronomenklassenangabe, Pronomen-Teilartikel 2. Partikel, Wortart

Substantiv/noun — 1. Angabe der Substantivklasse, framebasierter Wörterbuchartikel 2. Artikel¹, Deklination, Flexion, Genus, Kasus, Kompositum, Numerus, Pluraletantum, Singularetantum, Substantivvalenz, usueller Benennungskontext, Wortart, Wortbildung

Verb/verb — 1. Angabe der Aktionsart, Angabe des Aspekts 2. Aktionsart, Hilfsverb, Infinitiv, Konjugation, Modalverb, Partikelverb, Perfekt, Präteritum, Verbvalenz

Wortart/part of speech — 1. Schulwörterbuch, Wörterbucheinleitung 2. Gesprächswort, Gradpartikel, Interjektion, Konjunktion, Modalpartikel, Modalverb, Numerale, Partikel, Präposition, Pronomen, Substantiv

In den Wortart-Artikeln und folglich in der obigen Zusammenstellung fehlen Angaben zu vielen Artikeln, auf die verwiesen werden könnte. Zum Beispiel im Artikel zu *Adjektiv/adjective* könnten Verweise auf Einträge zu solchen Termini wie *erweiterter Adjektiv-Adverb-Artikel*, *erweiterter Adjektiv-Präposition-Artikel* und *erweiterter Adjektiv-Pronomen-Artikel* platziert werden. Es leuchtet ein, dass es im Falle eines so umfangreichen Wörterbuchs nicht immer möglich ist, in der Verweisposition eines Artikels alle mit ihm zusammenhängenden Artikel zu nennen. Dennoch sind manche Unterschiede¹⁵ zwischen den Verweisangaben einzelner Artikel nicht nachvollziehbar und als fehlende Kohärenz in der mediostrukturellen Vernetzung zu beanstanden. Dies ändert jedoch nichts an der Tatsache, dass die Wortart-Artikel im WLWF sehr stark im lexikographischen Kontext verankert sind.

Nach dieser Etappe der Analyse lässt sich Folgendes festhalten: In den Wörterbüchern LDBT und DoL wird nur der Begriff der Wortart aus lexikographischer Perspektive dargestellt, wohingegen im WLWF alle untersuchten Artikel lexikographische Informationen enthalten. Mehrere Verweise auf Einträge zu lexikographischen Begriffen sowie auf Einträge, die linguistische Begriffe auch unter lexikographischem Aspekt beschreiben, gibt es im WLWF fast in jedem Artikel (nur die Artikel zu *Konjunktion* und *Numerale* sind nicht so gut vernetzt wie die anderen) und im Falle der zwei übrigen Lexika lediglich in den Einträgen zu *part of speech*. Sonst verweisen alle Wortart-Artikel im LDBT nur auf *part of speech* und *part of speech label*, und die Mehrheit der DoL-Wortart-Artikel verweist ausschließlich auf *part of speech* und *grammatical information*.

3.5 Im Folgenden wird denjenigen Artikeln Aufmerksamkeit geschenkt, in denen lexikographische Informationen zu finden sind, d.h. den Artikeln *Wortart /*

part of speech in allen drei Lexika und den Artikeln zu den einzelnen Wortarten im WLWF.

Wenn es um die lexikographische Behandlung der Wortarten geht, ist der DoL-Eintrag zu *part of speech* lakonisch: „Minimal part-of-speech information is given in the general dictionary, usually in abbreviated or coded form. More detail is provided in specialised dictionaries such as grammar dictionaries and usage guides“. Dem umfangreicheren Eintrag im LDBT lässt sich dagegen entnehmen, dass die Artikel der meisten europäischen Sprachwörterbücher Wortartangaben enthalten, dass und warum allgemeine Wörterbücher auf einer traditionellen Wortartklassifikation basieren (diesem Problem wird ein ganzer Absatz gewidmet) und welche Wortarten in Sprachwörterbüchern unterschieden werden. Ähnliche Fragen werden auch im WLWF besprochen, hier allerdings ausführlicher, mit entsprechender Exemplifizierung und kritischer Komponente. Darüber hinaus wird im WLWF am Beispiel ausgewählter Wörterbücher des Deutschen dargestellt, wie das Problem der Wortartangaben in Wörterbucheinleitungen behandelt wird.

Die Gestaltung der Wortart-Artikel des WLWF folgt keiner stabilen Vorlage, wenn es sich um die Auswahl und Reihenfolge lexikographischer Informationen handelt. Nicht in jedem Wortart-Artikel kommen Informationen derselben Typen vor, was einerseits an der Spezifik der jeweiligen Wortart liegt, andererseits aber wahrscheinlich auf das Fehlen eines einheitlichen Musters zurückzuführen ist. Bei einer pauschalen Betrachtung der Wortart-Artikel lassen sich vier Informationstypen unterscheiden:

- (1) Es wird angegeben, welche Eigenschaften von Einheiten, die zur jeweiligen Wortart gehören, in Wörterbüchern dargestellt werden (sollen); vgl. z.B. die Einträge zu *Adjektiv, Substantiv, Verb*.
- (2) Es werden — manchmal sehr allgemein — Probleme der Identifizierung, Zuordnung und Beschreibung der betreffenden Einheiten besprochen; vgl. z.B. die Einträge zu *Artikelwort, Konjunktion, Numerale, Partikel, Präposition*.
- (3) Es werden Wörterbücher bzw. Wörterbuchttypen genannt, die sich vorwiegend oder ausschließlich mit Wörtern der betreffenden Wortklasse befassen; vgl. z.B. die Einträge zu *Adjektiv, Präposition*.
- (4) Einige Artikel enthalten allgemeine metalexikographische Kommentare über die lexikographische Behandlung von Einheiten der betreffenden Wortklasse, die Rolle der metalexikographischen und/oder linguistischen Forschung, über deren Einfluss auf die lexikographische Praxis etc.; vgl. z.B. die Einträge zu *Adverb, Interjektion, Pronomen, Substantiv*.

Anders als die übrigen Lexika weist das WLWF eine sehr weitgehende Parzellierung des vermittelten Fachwissens auf. Deren Folge ist u.a., dass lexikographische Informationen über die einzelnen Wortarten hier nicht nur in den untersuchten Wortart-Artikeln untergebracht sind, sondern auch in Einträgen zu den einzelnen Angabe-, Artikel- und Lemmatypen. Um festzustellen, wie Wortarten in diesem Wörterbuch behandelt werden, ist es daher notwendig, auch

Artikel zu *Adjektivangabe*, *Adjektivartikel*, *Adjektivlemma*, *Verbangabe*, *Verbartikel*, *Verblemma* etc. in die Analyse mit einzubeziehen. Es ist dabei anzumerken, dass es Einträge dieser Art nicht zu allen Wortarten gibt und dass manche der vorhandenen Einträge keine Angaben außer einer Definition enthalten. Darüber hinaus ist die Verteilung der Informationen auf die Einträge zu Lemmata vom Typ „-angabe“, „-artikel“ und „-lemma“ nicht einheitlich, es kommen Überlappungen und Wiederholungen vor. So ist z.B. sowohl im Eintrag zu *Adjektivartikel*, als auch in denen zu *Adjektivangabe* und *Adjektivlemma* zu lesen, dass als Lemma die unflektierte Form des Adjektivs fungiert. Der Artikel zu *Numeralenangabe* enthält dagegen Informationen, die nicht die Angabe als solche, sondern allgemein die Behandlung der Numeralia im Wörterbuch betreffen und daher im Artikel zu *Numerale* zu platzieren wären.

Unabhängig von der Distribution der Daten innerhalb dieser Artikelgruppe lassen sich hier einige Haupttypen der Informationen unterscheiden:

- (1) Im Falle von flektierbaren Wörtern wird die morphologische Form angegeben, die als Lemma angesetzt wird; vgl. z.B. Einträge zu *Adjektivlemma*, *Substantivlemma*, *Verblemma* (s. Abb. 1 im Anhang).
- (2) Es werden Wörterbuchartikel oder deren Teile als Beispiele angeführt und manchmal kurz kommentiert; vgl. z.B. Einträge zu *Konjunktionsangabe*, *Substantivangabe*, *Numeralenartikel* und *Präpositionsartikel* (s. Abb. 2 im Anhang).
- (3) Es werden Wörterbuchtypen genannt, die betreffende Artikel bzw. Lemmata aufweisen; vgl. z.B. Einträge zu *Artikkellemma*, *Numeralenartikel*, *Partikelartikel*, *Pronomenartikel* und *Verbartikel* (vgl. auch Pkt. 3 weiter oben; s. Abb. 3 im Anhang).
- (4) Eine Gruppe von Artikeln enthält einen terminologischen Hinweis zum Gebrauch der Termini *-lemma* und *lemmatische -angabe*, die sich auf dieselbe Artikelkonstituente beziehen, aber entweder den makrostrukturellen (z.B. *Verblemma*, *Substantivlemma*, *Pronomenlemma*) oder den mikrostrukturellen Aspekt (z.B. *lemmatische Verbangabe*, *lemmatische Substantivangabe*, *lemmatische Pronomenangabe*) betonen (s. Abb. 1 im Anhang).

Das WLWF ist das einzige Lexikon unter den untersuchten, das Wortarten auch aus lexikographischer Perspektive darstellt. Trotz einiger Inkonsequenzen und Mängel in der Beschreibung lassen sich diesem Nachschlagewerk die wichtigsten Informationen zur lexikographischen Behandlung von Wörtern der einzelnen Wortarten entnehmen. Die Konsultierung der Wortart-Artikel soll dabei immer durch das Nachschlagen in den entsprechenden *-angabe-*, *-artikel-* und *-lemma-*Artikeln ergänzt werden.

3.6 Die Erkenntnisse aus der obigen Analyse lassen sich zusammenfassend wie folgt darstellen. Umfangsmäßig unterscheiden sich die untersuchten Einträge von den übrigen Artikeln der genannten Lexika nicht.

Die Wortart-Artikel des LDBT enthalten nur grammatische Informationen,

obwohl im Vorwort eine lexikographische Perspektive angekündigt wird (vgl. 2.3). Auch die Wortart-Artikel im DoL bieten keine lexikographischen Informationen. Sie vermitteln grammatisches Basiswissen über die einzelnen Wortarten und können m. E. nur jemandem behilflich sein, der nicht in der Lage ist, Wortarten als solche zu identifizieren. Man fragt sich allerdings, ob solche Personen überhaupt zum Benutzerkreis der Wörterbücher zur Lexikographie gehören (zum Adressatenkreis des LDBT¹⁶ und des DoL vgl. 2.3, s. dazu auch 4) bzw. warum sie nach einfachen grammatischen Erklärungen zu Wortarten in Wörterbüchern dieses Typs statt in einem Handbuch oder Lexikon zur Grammatik/Linguistik suchen sollten. Die einzigen der untersuchten Einträge im LDBT und im DoL, in denen lexikographische Perspektive eingenommen wird und die durch Verweise besser mit anderem lexikographischen Inhalt dieser Lexika verbunden sind, sind die Einträge zu *part of speech*.

Alle analysierten Einträge im WLWF beschreiben die Wortarten auch aus lexikographischer Perspektive, wobei der quantitative Anteil von lexikographischen Informationen an den gesamten Artikeln unterschiedlich ist. Zur Beschreibung der Wortarten gehört in diesem Wörterbuch auch eine Gruppe von Einträgen zu *-angabe*, *-artikel* und *-lemma*, die es daher in der Untersuchung zu berücksichtigen gilt. Wegen des mangelnden einheitlichen Beschreibungsschemas variiert und überlappt sich manchmal das Informationsangebot in den Artikeln, es lassen sich jedoch einige Haupttypen von lexikographischen Informationen über die Wortarten erfassen, die in den meisten der untersuchten Einträge repräsentiert werden (vgl. 3.5). Am Rande ist hier anzumerken, obwohl das nicht zum Untersuchungsgegenstand gehört und daher im Vorangehenden nicht thematisiert wurde, dass die WLWF-Artikel viel detailliertere linguistische Informationen über Wortarten bieten als die anderen zwei Lexika. Somit kann das WLWF seinen (grammatisch gut ausgebildeten) Benutzern auch in dieser Hinsicht als Nachschlagewerk dienen. Die Wortart-Artikel des WLWF sind dank zahlreicher Verweise besser als die entsprechenden Artikel im LDBT und im DoL im lexikographischen Kontext eingebettet, was hauptsächlich durch den Umfang dieses Wörterbuchs bedingt ist.

4. Abschließende Bemerkungen

Die Analyse hat einige Ähnlichkeiten und Unterschiede in der Behandlung der Wortarten in den drei Lexika aufgezeigt. Es fällt auf, dass sich gerade im Hinblick auf die grundlegende Entscheidung, wie die lexikographisch erfassten Wortarten zu beschreiben sind, zwei gegensätzliche Konzeptionen erkennen lassen. Im LDBT und im DoL werden die einzelnen Wortarten — von den Einträgen zu *part of speech* hier mal abgesehen — ausschließlich unter grammatischem Aspekt dargestellt, während das WLWF jede Wortart auch aus lexikographischem Blickwinkel beleuchtet.

Die letztgenannte Verfahrensweise ist unseres Erachtens viel höher zu bewerten als die Beschreibung der Verben, Substantive etc. ohne Berücksichti-

gung des lexikographischen Kontextes. Es ist anzunehmen, dass die Benutzer der Wörterbücher zur Lexikographie nicht an grammatischen Grundinformationen, sondern vor allem oder gar ausschließlich an lexikographischen Informationen zu Wortarten interessiert sind und dass sie erwarten, in diesen Wörterbüchern eben solche Informationen zu finden und nicht die gleichen wie in Lexika zur Linguistik. Wortart-Artikel, wie sie im LDBT und im DoL vorkommen, verfehlen daher ihren Zweck, entsprechen nicht der Funktion des gesamten Fachwörterbuchs und beeinträchtigen dessen Nutzwert. Die Aufnahme der Wortartbezeichnungen in die Lemmaliste scheint bei einem solchen Inhalt der Einträge nicht ganz begründet zu sein. Die Zusammenhänge zwischen den beschriebenen Wortarten und dem Wörterbuchgegenstand, d.h. Lexikographie, sind nicht erkennbar, und die dürftigen Verweise helfen dem nur in geringem Grade ab.

Da in Wörterbüchern zur Lexikographie nicht nur das Wortartensystem, sondern auch andere Aspekte der Linguistik sowie anderer für die Lexikographie relevanter Disziplinen berücksichtigt werden können, lässt sich die obige Einschätzung verallgemeinern und ihr Geltungsbereich ausweiten. Alle Einträge zu Termini von außerhalb der Lexikographie sollten in Wörterbüchern zur Lexikographie auch lexikographische Informationen enthalten, sonst kann ihr Vorkommen in diesen Nachschlagewerken unberechtigt erscheinen. Sie sollten dem Profil des Fachwörterbuchs entsprechen und Bedürfnissen von lexikographisch interessierten Benutzern entgegenkommen. Es ist nicht die Funktion dieser Artikel, analoge Artikel aus Fachwörterbüchern zu betreffenden Disziplinen zu ersetzen und den gleichen Inhalt zu bieten. In Wörterbüchern zur Lexikographie sollte stets, d.h. auch in den Einträgen zu lexikographieexternen Termini, die Lexikographie im Mittelpunkt stehen und die Beschreibungsperspektive prägen.

Endnoten

1. Der Ausdruck „lexikographisch“ wird im Beitrag in zwei Bedeutungsnuancen verwendet: zum einen in der allgemeinen, gängigen Bedeutung, wie z.B. in „lexikographische Beschreibung“ (Beschreibung in einem lexikographischen Nachschlagewerk) und „lexikographische Konzeption“ (Konzeption eines lexikographischen Nachschlagewerks), zum anderen auch als Gegenteil zu „linguistisch“ bzw. „grammatisch“. So ist z.B. die lexikographische Perspektive der Beschreibung von Wortarten nicht als eine für Wörterbücher typische Beschreibungsperspektive zu verstehen, sondern als eine nicht grammatische/nicht linguistische Perspektive. Das bedeutet, dass nicht die linguistischen Eigenschaften der betreffenden Wortart behandelt werden, sondern Probleme, die mit der Beschreibung der Wortart in Sprachwörterbüchern zusammenhängen, und manchmal auch metalexikographische Arbeiten über die Wortart. Unter lexikographischem Inhalt oder lexikographischen Informationen sollten hier nicht alle Informationen aus einem Nachschlagewerk, hier: einem Wörterbuch zur Lexikographie, verstanden werden, sondern nur Informationen über Lexikographie, also über das lexikographische Verhalten der Wortarten und nicht über ihre linguistischen Eigenschaften. Die realisierte Bedeutungsvariante von „lexikographisch“ geht immer aus dem Kontext hervor.

2. Zu den Äquivalentsprachen gehören Afrikaans, Bulgarisch, Französisch, Italienisch, Portugiesisch, Spanisch, Ungarisch und Russisch (nicht in allen Artikeln).
3. Die gewählten Wörterbücher sind Gegenstand einiger Besprechungen, vgl. etwa Bielińska (2011), Heliel (2000), Nied Curcio (2012), Nielsen (1999), Pálfi/Kristiansen (2010) und Urdang (2000). Ihre Autoren befassen sich mit diversen Aspekten der rezensierten Nachschlagewerke, aber das hier thematisierte Problem der Wortartenbeschreibung wird von ihnen nicht behandelt. Daher wird im Folgenden auf diese Rezensionen nicht eingegangen.
4. „Dieses Wörterbuch richtet sich in erster Linie an Studenten und Doktoranden der sprachwissenschaftlichen Abteilungen philologischer Fakultäten und bietet einen Überblick über die Lexikographie und ihre Beziehung zu den einschlägigen linguistischen Disziplinen. Es kann insbesondere in den Seminaren und Spezialkursen über Lexikographie, Lexikologie und linguistische Semantik nützlich sein.“
5. „die Begriffe zu spezifizieren, die aus anderen linguistischen und nicht-linguistischen akademischen und angewandten Disziplinen entlehnt wurden, und die Modifikationen festzustellen, die diese Begriffe im Rahmen der lexikographischen Beschreibung erfahren haben“
6. „ein praktizierender Lexikograph, ein Wörterbuchforscher, ein Experte in einer benachbarten Disziplin oder ein Student“
7. „Begriffe ohne unmittelbaren Bezug zum allgemeinen Thema, z. B. aus der deskriptiven Linguistik (z. B. *aphesis*, *double-headed nominal group*) oder der Drucktechnik (z. B. *bullet*, *font*), oder Namen von Sprachen und Regionalismen“
8. Vgl. den Gebrauch der Wortartangabe *determiner* in den Artikeln zu *this*, *my*, *few*, *which*, *both* etc. in macmillandictionary.com, dictionary.cambridge.org oder oxfordlearnersdictionaries.com.
9. Im WLWF ist der Eintrag zu *Artikelwort/determiner* im Nachtrag zu finden.
10. Der deutschsprachige Terminus *Postposition* wird hier nicht angegeben, weil der entsprechende Eintrag im WLWF fehlt.
11. Vgl. die Angaben zur durchschnittlichen Artikellänge im DoL, im LDBT und in Bd. 1 des WLWF in Mann (2013: 447, 455).
12. Die Typographie der im Aufsatz angeführten Einträge entspricht nicht der typographischen Gestaltung der Einträge in Lexika.
13. Eine Spalte zählt ungefähr 3000 Zeichen.
14. Eintrag zu *Morphologieangabe bei Substantiven* ist nur ein Verweisartikel.
15. Beispielsweise wird von den Artikeln zu *Adjektiv/adjective* und *Partikel/particle* auf die zu *Adjektivvalenzwörterbuch* und *Partikelwörterbuch* verwiesen. In den Artikeln zu *Substantiv* und *Verb* fehlen jedoch Verweise auf *Substantivvalenzwörterbuch*, *Verbwörterbuch* und *Verbvalenzwörterbuch*. Ähnliches trifft bei Verweisen auf *-angabe*, *-artikel*, *-lemma* und *-Teilartikel* im Falle mancher anderen Wortarten zu. So gibt es z.B. im Eintrag zu *Adjektiv* Verweise auf *Adjektivartikel* und *Adjektivlemma*, aber in den Artikeln zu *Substantiv*, *Verb* oder *Partikel* fehlen entsprechende Verweise. Vom Artikel zu *Präposition* wird auf *Präpositionsangabe* und *Präpositions-Teilartikel* verwiesen, von den Artikeln zu *Substantiv*, *Numerale* und *Verb* wird dagegen auf analoge Artikel nicht verwiesen.
16. Von den Adressaten des LDBT wird zwar kein linguistisches Wissen erwartet, dies bezieht sich aber wohl kaum auf die Kenntnis der Wortarten. Es ist anzunehmen, dass jeder Student, also Wörterbuchbenutzer nach dem Abitur, das in der Grundschule zu erwerbende grammatische Wissen bereits besitzt. Vgl. dazu eine Stelle aus dem WLWF-Artikel zu *Wortart*: „Darüber

hinaus gilt gemeinhin die Unfähigkeit, sprachliche Einheiten einer bestimmten Wortart zuzuordnen zu können, als Ausdruck mangelnder Bildung — berechtigterweise zumindest, was die zentralen Wortarten angeht [...].“

Wörterbücher zur Lexikographie

- Bergenholtz, H., I. Cantell, R. Vatvedt Fjeld, D. Gundersen, J.H. Jónsson und B. Svensén.** 1997. *Nordisk Leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Bielińska, M. (Hrsg.).** 2020. *Leksykografia. Słownik specjalistyczny*. Kraków: Universitas.
- Burkhanov, I.** 1998. *Lexicography. A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- De Sousa, J.M.** 1995. *Diccionario de lexicografía práctica*. Barcelona: Bibliograf.
- Hartmann, R.R.K. und G. James.** 2001. *Dictionary of Lexicography*. London/New York: Routledge.
- Swedenborg, L.** 1990. *En liten ordbok om ordböcker*. Stockholm: Esselte.
- Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, M. Mann, A. Storrer und W. Wolski.** 2010–2020. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research, Bd. 1-5*. Berlin/Boston: Walter de Gruyter.
- Yang, Z. und Q. Xu.** 1992. *Cishuxue cidian*. Shanghai: Xuelin Press.

Sekundärliteratur

- Bergenholtz, H.** 1994. Plan eines nordischen Wörterbuches der Lexikographie. Ein Werkstattbericht. Hyldgaard-Jensen, K. und V.H. Pedersen (Hrsg.). 1994. *Symposium on Lexicography VI. Proceedings of the Sixth International Symposium on Lexicography May 7–9, 1992 at the University of Copenhagen*: 315-329. Tübingen: Niemeyer.
- Bielińska, M.** 2011. Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer, Werner Wolski: *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. t. 1. Berlin/New York: Walter de Gruyter 2010. *Biuletyn Polskiego Towarzystwa Językoznawczego* LXVII: 343-350.
- Bielińska, M.** 2019. Status i istota leksykografii. Stanowiska i kontrowersje. *Polonica* 39: 111-148.
- Heliel, M.H.** 2000. Igor Burkhanov. *Lexicography. A Dictionary of Basic Terminology*. *Lexikos* 10: 320-325.
- Mann, M.** 2013. New Developments in Lexicography for Special Purposes II. An Overview of Dictionaries of Lexicography. 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*: 442-460. Berlin/Boston: Walter de Gruyter.
- Nied Curcio, M.** 2012. H.E. Wiegand, M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer, W. Wolski (Hrsg.). *Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research*. *Lexicographica* 28: 399-407.
- Nielsen, S.** 1999. R.R.K. Hartmann and Gregory James: *Dictionary of Lexicography*. *Lexicographica* 15: 271-316.

- Pálfi, L.-L. und B. Kristiansen.** 2010. Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer, Werner Wolski (Herausgeber und Bearbeiter / Editors and Compilers). *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Band 1 / Volume 1: Systematische Einführung / Systematic Introduction + A–C. International Journal of Lexicography* 23(4): 491-502.
- Urdang, L.** 2000. R.R.K. Hartmann and Gregory James. *Dictionary of Lexicography. International Journal of Lexicography* 13(1): 35-42.

Anhang

Verblemma *das*; -ta

Lemma, zu dem als Lemmazeichen eine Verbform gehört.

verb lemma : lemma to which as lemma sign a verb form belongs.

● In der Lexikographie des neueren Deutsch besteht z. B. die Konvention, dass bei Verben der Infinitiv als Verblemma angesetzt wird. In der lateinischen Lexikographie wird dagegen die 1. Person Singular Präsens Indikativ Aktiv als Verblemma angesetzt. – Der Terminus *Verblemma* betont den makrostrukturellen, der Terminus *lemmatische Verbangabe* den mikrostrukturellen Aspekt bei der gleichen Artikelkonstituente.

▷ *Lemma · Verb · Verbartikel*

▣ WIEGAND 2005

[HEW]

● **af** *werkwoordlemma* · **bg** *лема на глагол* · **es** *lema de verbo* · **fr** *lemme de verbe* · **hu** *igei lemma* · **it** *lemma di verbo, lemma verbale* · **pt** *lema de verbo*

Abb. 1: Eintrag zu *Verblemma* / *verb lemma* aus WLWF, Bd. 4

Konjunktionsangabe *die*; -n

elementare Angabe, die das Lemmazeichen als Konjunktion klassifiziert.

item giving a conjunction : elementary item that classifies the lemma sign as a conjunction.

• In:

(1) **oder** *Konjunktion*;

(³LGWDAF)

ist „*Konjunktion*“ die Konjunktionsangabe. – In:

(2) **als**¹ (*Konj.*)

(WAHRIG-DW)

ist „*Konj.*“, die Abkürzung für *Konjunktion*, die verdichtete Konjunktionsangabe. – In

(3) **im·me·di·ate·ly** ⇨ /ɪ'mi:diətli/ *adv., conj.*

■ *adv.* 1..

■ **conj.** (*especially BrE*) as soon as : *Immediately she'd gone, I remembered her name.*

(⁸OALD 2010)

sind „*conj.*“ und „**conj.**“ (beides die Abkürzung für engl. *conjunction*) jeweils die verdichtete Konjunktionsangabe.

Symbol für die Angabeklasse: KonjkA

▷ § 14 | *Angabe · elementare Angabe · Konjunktion*

■ WIEGAND 2005 [MK]

© **af** *voegwoorandaander* · **bg** *сведение за съюз* ·

es *indicación de conjunción* · **fr** *indication de conjonction* · **hu** *kötőszó adata* · **it** *indicazione di congiunzione* · **pt** *indicação de conjunção*

Abb. 2: Eintrag zu *Konjunktionsangabe/item giving a conjunction* aus WLWF, Bd. 3

Partikelartikel *der*; -Ø

Wörterbuchartikel, mit dessen Lemma eine Partikel genannt wird.

particle article : dictionary article of which the lemma indicates a particle.

● Partikelartikel finden sich u. a. in allgemeinen ein- und zweisprachigen Wörterbüchern, in Lernerwörterbüchern, in historischen Wörterbüchern und in speziellen Partikelwörterbüchern, wie z. B. HELBIG 1988 oder MÉTRICH/FAUCHER 2009.

▷ *Partikel*

▣ WIEGAND 2003a · WOLSKI 1986 · 1986a · 1989c · 1991 [HEW]

© **af** *deeltjie-artikel* · **bg** *статия за частуца* · **es** *artículo de partícula* · **fr** *article de particule* · **hu** *partikula szócikke* · **it** *articolo di particella* · **pt** *artigo de partícula*

Abb. 3: Eintrag zu *Partikelartikel* / *particle article* aus WLWF, Bd. 3

After the Digital Revolution: Dictionary Preferences of English Majors at a European University

Bartosz Ptasznik, *University of Warmia and Mazury, Olsztyn, Poland*
(bartosz.ptasznik@uwm.edu.pl)

Abstract: The present contribution reports on a recent survey of dictionary-consulting habits and preferences of university students majoring in English at a European university. Amongst the information sought is the choice between digital online dictionaries and traditional print dictionaries, as well as a number of other issues of interest to lexicographers and educators. It was found that online dictionaries have all but replaced the traditional paper dictionary, suggesting that the digital revolution in lexicography hailed by Lew and De Schryver (2014) is complete. A positive finding is that students prefer dictionaries that are highly rated by experts. The paper concludes with a number of pedagogical suggestions.

Keywords: ENGLISH, LEXICAL INFORMATION, LEXICOGRAPHY, DICTIONARY USE, DIGITAL LITERACY, ONLINE DICTIONARY

Opsomming: Ná die digitale revolusie: Woordeboekvoorkeure van studente met Engels as hoofvak aan 'n Europese universiteit. In hierdie bydrae word verslag gelewer oor 'n onlangse opname van woordeboekraadplegingsgewoontes en -voorkeure van universiteitstudente met Engels as hoofvak aan 'n Europese universiteit. Inligting wat o.a. ondersoek is, is die keuse tussen digitale aanlyn woordeboeke en tradisioneel gedrukte woordeboeke, asook 'n aantal ander kwessies van belang vir leksikograwe en opvoeders. Daar is bevind dat aanlyn woordeboeke byna die tradisionele papierwoordeboek vervang het, wat daarop dui dat die digitale revolusie in die leksikografie, soos voorspel deur Lew en De Schryver (2014), voltooi is. 'n Positiewe bevinding wat gemaak is, dui daarop dat studente woordeboeke verkies wat hoog aangeskrewe word deur kundiges. Hierdie artikel word afgesluit met 'n aantal pedagogiese voorstelle.

Sleutelwoorde: ENGELS, LEKSIKALE INLIGTING, LEKSIKOGRAFIE, WOORDEBOEK-
GEBRUIK, DIGITALE GELETTERDHEID, AANLYN WOORDEBOEK

1. Introduction

The study of dictionary use is a continuously evolving area of research driven by ongoing advances in methodology as well as technology. A variety of research techniques have been employed in the field of dictionary use over the past years, such as participant and non-participant observations, self-accounts,

think-aloud protocols, or videotaping, to mention only a few (Welker 2013; Lew 2015; Schierholz 2015). Also, test-based and experimental research has played a major role in the context of dictionary use by contributing to higher objectivity and reliability of the collected data (Nesi 2000), aided by refinements in statistical inference. A notable development was the pioneering of eye-tracking technology in the study of dictionary consultation behavior (Simonsen 2009, 2011; Tono 2011; Lew et al. 2013, 2018), as was the study of dictionary log-files (De Schryver et al. 2019). Notwithstanding the importance of all the aforementioned methodological approaches, survey-based research remains valid when it comes to gauging user habits, preferences, and attitudes (Kosem et al. 2019). To attain the principal research objectives, this is the technique adopted for the present study which deals with the dictionary preferences of English majors at a European university.

2. Benefits and drawbacks of questionnaires

The questionnaire is a powerful instrument in the sense that survey-based research facilitates the collection of substantial data within a reasonable time-frame. However, there are concerns about the extent to which this approach yields accurate and reliable data. For example, the fact that some respondents could be unwilling to provide honest answers in the questionnaire should be taken seriously (Humb le 2001: 44). Sudman and Bradburn (1982) identify this problem by arriving at the conclusion that either under-reporting or over-reporting can occur when respondents are asked to give detailed information with regards to sensitive issues (family life, private life, income, etc.). As a result, valuable information is withheld from the researcher, which hinders the process of accurate data collection. Another potential problem is specifically related to the responses provided by the respondents. In particular, Hatherall (1984: 184) speculates that some respondents can unintentionally supply researchers with inaccurate data, by posing the much-quoted question: "Are subjects saying here what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three"? Given the fact that numerous surveys are presently administered online, there is no denying that in various cases it would be difficult for the researcher to get in touch with the respondents to obtain any additional information necessary for reasons of clarification, unless of course such question items are included in the questionnaire from the outset. Tono (2001: 67) shares Hatherall's point of view by acknowledging that in surveys "the findings are still "indirect" in the sense that they only record what the users think happens, as opposed to what actually happens". Interestingly, Harris (2014: 2-3) expresses his opinion by contending that "[w]riting questionnaires is arguably one of the most challenging forms of writing. It is, in a sense, a conversation between you and hundreds, if not thousands, of diverse respondents. The conversation may take fifteen, thirty, or even forty-five minutes". Last but not least, as Debois (2019) points out, among other problematic issues with questionnaires are: skipped questions, problems related to the

interpretation of specific questions by the respondents, the inability of the researcher to analyze the respondents' emotional state of mind, problems with interpreting open-ended¹ question responses, respondent bias, lack of personalization, unconscientious answers, questionnaire fatigue.

The reason for which the researcher sets out to prepare a questionnaire is obviously not because this research instrument has its shortcomings in survey-based research. The questionnaire is a tool which allows the researcher to conduct a survey. According to Brown's (2001: 6) definition, questionnaires are "written instruments that present respondents with a series of questions or statements to which they are to react either by writing out their answers or selecting them among existing answers". In other words, a questionnaire constitutes a set of questions prepared by the researcher, whereas a survey is a broader term which could simply be defined as the process of data collection and analysis of the responses provided by the respondents, through the use of research instruments such as the interview or questionnaire. Yet another definition of a survey would be that it is simply "a system for collecting information" (Sue and Ritter 2012: 3). There are countless benefits to drawing up questionnaires, or conducting survey-based research. Debois (2019) lists the following advantages of questionnaires: inexpensive, practical, fast results, scalability, comparability, easy analysis, validity and reliability, standardized, no pressure, respondent anonymity. Mackey and Gass (2005: 94-96) contend that questionnaires hold a serious advantage over individual interviews as they are more economical and practical. In addition, Mackey and Gass mention flexibility as an added advantage, since "questionnaires can provide both qualitative insights and quantifiable data, and (...) are (...) used in a range of research". Furthermore, Jackson and Furnham (2000: 5), who elaborate on surveys and questionnaires for health professionals and administrators, note that questionnaires allow one to "identify opportunities for growth, change and improvement", a remark that is easily transferable to the dictionary user research context. As far as Internet surveys are concerned, Blair et al. (2014: 57) contend that "[t]he two great advantages of Internet surveys are the low cost and the speed of data collection". Blaxter et al. (2006: 79) point to the fact that surveys "may be repeated in the future or in different settings to allow comparisons to be made". In brief, the advantages of surveys for the researcher are plentiful.

All things considered, questionnaires have their advantages and disadvantages. For the sake of clarity, it is the author's belief that the benefits of survey-based research in dictionary user studies outweigh the drawbacks. Questionnaires are an effective tool which provide researchers with comments and reflections of respondents that can be an invaluable source of information. By way of example, bearing in mind that contemporary dictionaries are designed for their users, it must be admitted that depriving oneself of the mere opportunity to acquire inside knowledge or fresh insights from the target user groups, who regularly use newly updated dictionaries, could be perceived as an irrational decision, even if this would mean obtaining only a minuscule amount of information from the collected data. Furthermore, a questionnaire can espe-

cially prove to be a useful instrument for metalexigraphers wishing to enhance their research design or consolidate their findings by supplying additional data. Also, in the author's view, questionnaires can serve as a starting point for the main research by revealing aspects worthy of further study. Last but not least, in the context of dictionary use studies, questionnaires can have a didactic purpose as "[u]niversity leaders [can] use results from surveys of students to revise their undergraduate and graduate education programs" (Dillman et al. 2014: 1), and can influence the didactic approach of teaching students how to use dictionaries, what types of dictionaries to use, etc. Importantly, this will be one of the main aims of the present survey. To sum up, a questionnaire-based survey is justified if it is appropriate to the researcher's aims and objectives, and on condition that the appropriate research design is adopted.

The following sections will elaborate on the present survey and its main findings.

3. Aims of the present survey

The aim of the present survey was threefold. First and foremost, the author set out to analyze how English majors at a European university use dictionaries and what types of dictionaries they consult when searching for pertinent information. In particular, the aim was to see to what extent the *digital revolution* in lexicography (Lew and De Schryver 2014) has completed its cycle among college-level students with an interest in language.

Second, the survey had a didactic purpose. The aim was to collect and analyze the data provided by the participants, and make inferences about how to rethink the approach to teaching English majors how to use dictionaries, what in dictionaries should students pay special attention to, etc.

Finally, the third aim of the survey was to re-establish the usefulness of survey-based research in dictionary user studies. Given the aims of the present survey, an attempt was made to find answers to the following research questions:

1. What types of dictionaries are used by English majors?
2. How do English majors at a European university use dictionaries?
3. Do the findings of the current survey carry any pedagogical implications for dictionary use by English majors?

4. Method, participants and procedure

To meet the objectives of the survey and given the COVID-19 pandemic and restrictions, as well as the fact that students had been attending online classes at the university (academic year 2020/2021), an online questionnaire was created using Microsoft Forms. Microsoft Forms is part of Office 365 and it is an easy-to-use online survey creator. The questionnaire consisted of 20 questions

phrased in Polish, the respondents' native language, and included 17 closed-ended questions and 3 open-ended questions which required lengthier responses (see Appendix B). The aim of presenting the questionnaire in the respondents' first language was to ensure the students' best possible comprehension of the survey questions. A mixed question format (i.e., open-ended along closed-ended items) was deemed as the most appropriate in the present context as the aim was not only to collect information that could be expressed numerically (quantitative analysis) but also to obtain information about the attitudes or habits (qualitative analysis) of the students. As far as the multiple-choice questions are concerned (Cohen et al. 2007: 323), the respondents were asked to select either only one answer (a *single answer* mode), or the students were asked to select more than one answer (*multiple answer* mode). One dichotomous question (answer options: *yes, no*) was also included in the survey. A Likert-type rating scale was adopted for selected questions. Using technical language in the questionnaire was avoided, which might have caused confusion (see Lew 2002, 2004). The questionnaire was distributed to the students via Microsoft Teams and e-mail. In order to fill out the questionnaire, the students had to log into their student accounts and complete the questionnaire via Microsoft Forms. A maximum of twenty minutes was allowed for the completion of the task, taking into consideration the fact that surveys should be kept sufficiently short. In this way it was possible to ensure that the participants would give the task their full attention when completing the questionnaire. Published research (e.g. Macer and Wilson 2013, as cited in Brace 2018: 49-50) suggests that the optimum duration of online surveys should be approximately 15 minutes. Given the format of the present survey (inclusion of both closed-ended and open-ended questions), an additional 5 minutes was given to the respondents.

To reiterate, the aim of the present survey was to analyze how English majors at a European university used dictionaries, and also to see what types of dictionaries students consulted when dealing with specific tasks (for example, translation tasks). Given the serious disruptions related to the COVID-19 pandemic, the sample was limited to one university (University of Warmia and Mazury in Olsztyn, Poland). This sample, however, was reasonably representative of European institutions of higher learning, based on the fact that Poland appeared to be quite a typical country in a recent pan-European survey (Kosem et al. 2019), and the university itself was a mid-sized public university.

The respondents who participated in the survey were males and females between 19 and 25 years old, and they were native speakers of Polish. They were first-year, second-year, third-year, fourth-year, and fifth-year full-time students majoring in English at the Faculty of Humanities. Their English proficiency level ranged from upper-intermediate to advanced (level B2 and C1 by the Common European Framework of Reference for Languages standards). Due to the COVID-19 pandemic and related restrictions, as well as the fact that classes had been taught online for the whole academic year (2020/2021), data were collected from the majority of the students, however, not all of the stu-

dents of English at the University of Warmia and Mazury in Olsztyn could participate in the survey: data were gathered from approximately 66.7% of the students (168 students of English, full-time studies). Consequently, taking into account this limitation, the data were not analyzed by the students' year of study and instead the survey data were summarized altogether as a whole, given the fact that the reasons for self-selection might vary with the year of study.

5. Results and findings

The vast majority of the respondents (99%) said that they usually resorted to online dictionaries for the purpose of dictionary consultation, while only 1% of the students admitted to using paper dictionaries. Nowadays this finding is not at all surprising as online dictionaries hold a real advantage over their printed counterpart, such as easy access, faster entry consultation time, up-to-date content or multimedia features, to name just a few. Also, it is more convenient for the average dictionary user to search for pertinent information in dictionaries which are available online, given that dictionary use very frequently occurs when other computer-based activities are performed (see Kosem et al. 2019).

Table 1: Responses to Question 1 (What dictionaries do you usually use?)

Type of dictionary	No. of responses	Percentage
Online dictionaries	167	99%
Paper dictionaries	1	1%

Notwithstanding the obvious differences between the online and paper dictionary formats and the apparent superiority of the former, in the present survey the respondents were also asked to describe the situations in which they would favor a book-form dictionary over an online dictionary (*Open-ended question: Explain in what situations you would prefer to use a paper dictionary rather than an online dictionary.*). Interestingly, some of the respondents made it clear that they would rather use a paper dictionary when working on their writing assignments (essays, paragraphs, etc.) or during the process of writing longer theses, such as their BA or MA papers. In addition, another student explained that he would rather resort to a paper dictionary during the process of writing which would not involve using a computer. In other words, this finding implies that there are students who use online dictionaries for computer-based tasks, whereas those students who focus solely on paper-based tasks resort to print dictionaries. Also, it appears that there is a minority of students who believe that paper dictionaries have more reliable content than their counterpart. A few students reported that in situations of doubt they would rather

reach for a print dictionary. Further, there were students who said that they would rather use a paper dictionary if they had more time, or when they study or learn for pleasure. A few students reported that paper dictionaries enhanced their ability to memorize words better. Two students said that they consulted book-form dictionaries only when the information that they were looking for in an online dictionary could not be found. Last but not least, one student claimed that he used print dictionaries when giving private lessons.

As for the specific online dictionaries that the respondents consult regularly when dealing with Polish–English or English–Polish translation tasks, the majority of the respondents (71%) chose *Diki.pl*, 67% of the respondents opted for *Bab.la*, whereas 54% of the students admitted to the fact that they also tended to rely on *Google Translate*. For this question, the respondents were allowed to select more than one answer. *PONS*, *Ling.pl* and *Edict.pl* turned out to be the students' least favorite dictionaries, achieving scores of 10%, 4% and 1% respectively. Interestingly, 7% of the respondents said that they used other dictionaries or parallel corpus query tools (for example, *Reverso Context*, *Linguee.pl* or *Cambridge Dictionary*) when translating words, texts, etc.

Table 2: Responses to Question 2 (Which of the following online dictionaries do you use for Polish–English or English–Polish translation?)

Answer option	No. of responses	Percentage
<i>Diki.pl</i>	120	71%
<i>Bab.la</i>	113	67%
<i>Google Translate</i>	91	54%
<i>PONS</i>	42	10%
<i>Ling.pl</i>	18	4%
I use other dictionaries	12	7%
<i>Edict.pl</i>	2	1%

Students who are more fluent and proficient in the English language also endeavor to find relevant information in monolingual learners' dictionaries when dealing with a specific task. Such dictionaries are rich in, for example, collocational information, example sentences, etc. Hence, in the following question the respondents were asked to select the British English monolingual learners' dictionaries that they used (the students were allowed to select more than one option for this particular question). As many as 93% of the students expressed their satisfaction with the *Cambridge Dictionary*, while the *Macmillan Dictionary* was rated by the students as the second most frequently used monolingual learner's dictionary (59%). Interestingly, less than half of the students

admitted to using either the *Oxford Advanced Learner's Dictionary* (43%), *Collins Online Dictionary* (22%) or *Longman Dictionary of Contemporary English* (13%), which all appeared to be less useful dictionaries to the participants.

Table 3: Responses to Question 3 (Which of the following British English monolingual learners' dictionaries do you use?)

Dictionary	No. of responses	Percentage
<i>Cambridge Dictionary</i>	155	93%
<i>Macmillan Dictionary</i>	98	59%
<i>Oxford Advanced Learner's Dictionary</i>	71	43%
<i>Collins Online Dictionary</i>	36	22%
<i>Longman Dictionary of Contemporary English</i>	21	13%

The aim of the following question in the survey was to find out how often students engaged in dictionary consultation. Most respondents declared that they used dictionaries *almost every day* (40%), 31% of them said that they consulted dictionaries *a few times a week*, whereas 23% of the students admitted to searching for information in dictionaries *every day*. 4% of the respondents said they used dictionaries *once a week*, while 2% admitted to using a dictionary *twice a week*.

Table 4: Frequency of dictionary use (Question 4: How often do you use a dictionary?)

Frequency of dictionary use	No. of responses	Percentage
Once a week	7	4%
Twice a week	3	2%
A few times a week	52	31%
Almost every day	68	40%
Every day	38	23%

The respondents were also asked how many dictionaries they usually used when looking for the meaning of a difficult word. The main purpose of this question was to see whether students tend to fully engage themselves in the process of dictionary use and refuse to abandon their attempt to find the correct meaning of a word when dealing with more sophisticated items. 18% of the

students declared that in such situations they resorted to only one dictionary, 52% admitted to being more inquisitive and as a result using two dictionaries, whereas 30% of the students said that they had a tendency to aid their search for meaning through the use of more than two dictionaries. Using two or more dictionaries for more sophisticated terms seems to be a positive trend, whereas the use of one dictionary should not necessarily be frowned upon provided that students manage to find the correct answer in a single source.

Table 5: Responses to Question 5 (How many dictionaries do you usually use when you are looking for the meaning of a difficult word in English?)

Answer option	No. of responses	Percentage
One dictionary	30	18%
Two dictionaries	87	52%
More than two dictionaries	51	30%

By and large, research has shown that dictionaries play a positive role in writing. For example, Lew (2016) and Tall and Hurman (2002) have demonstrated that dictionaries can significantly improve the quality of writing, whereas other studies (Bishop 2000; East 2008) have indicated that the mere availability of dictionaries in the process of writing brings students psychological reassurance during tests and examinations. In addition, Chon's study (2009) on digital dictionaries showed that Korean students were able to use more sophisticated words in their writing tasks once they had discovered those words when using a dictionary. In light of the importance of the presence of dictionaries in the process of writing, the aim of the present survey was to discover whether English majors tended to use monolingual or bilingual dictionaries when writing essays, letters, reports, etc. The findings reveal that most students used both monolingual and bilingual dictionaries (58%) for writing, 26% of the students used monolingual dictionaries, while only 14% of the students used bilingual dictionaries. 1% of the students said that they did not use dictionaries when producing written texts, whereas 1% of the respondents opted for the *other* choice. To sum up, deciding to use both monolingual and bilingual dictionaries, or at least having both dictionary types at one's disposal, seems to be the right choice. Nevertheless, it seems worrying that a small proportion of the students either did not resort to using a dictionary, or solely relied on the use of a bilingual dictionary, ignoring the additional benefits that monolingual dictionaries could bring to the process of writing.

Table 6: Responses to Question 6 (What dictionary type do you usually use for writing in English, for example, an essay, a letter, a report, or some other written form?)

Answer option	No. of responses	Percentage
Monolingual	44	26%
Bilingual	23	14%
Monolingual and bilingual	97	58%
I do not use a dictionary	2	1%
Other	2	1%

Given *Question 7*, the students were supposed to provide answers regarding the dictionary type they used when searching for the meaning of a newly encountered word. 25% of the students said that they used a monolingual dictionary, 23% admitted to using a bilingual dictionary, whereas 52% declared that they used both a monolingual and bilingual dictionary.

Table 7: Responses to Question 7 (What dictionary type do you use more often when looking for the meaning of an unknown word in English?)

Answer option	No. of responses	Percentage
I use a monolingual dictionary	42	25%
I use a bilingual dictionary	38	23%
I use both a monolingual and bilingual dictionary	88	53%

One question that has become relevant recently is whether dictionary users prefer to engage in dictionary consultation by accessing their desktop/laptop computers, smartphones or tablets, and importantly what the reasons are behind these choices. Interestingly, this question has already been partially answered in a recent survey completed by roughly 10,000 dictionary users and non-users (Kosem et al. 2019). The survey revealed that the participants clearly favored computer-based dictionaries over either smartphone-based dictionaries or dictionaries accessed from a tablet. The authors attributed this to the smaller display size of smartphones and the fact that it might be more convenient for dictionary users to access dictionaries from computers, given that most of their activities are computer-based and it seems reasonable to use the same device. It could be added that perhaps not all dictionary users own tablets and

hence the sporadic use of dictionaries accessed from tablets. Likewise, in the present survey, the vast majority of the students (74%) declared that they preferred to use computer-based dictionaries (either desktop or laptop computers), 26% preferred using dictionaries from their mobile phones, whereas 1% of the respondents said that they preferred to use dictionaries accessed from a tablet. This finding indicates that computer-based dictionaries are superior to both the smartphone and tablet format of dictionaries, and it seems that the reasons for this preference are once again not coincidental.

Table 8: Responses to Question 8 (What dictionaries do you prefer to use?)

Answer option	No. of responses	Percentage
Dictionaries accessed from a computer	124	74%
Dictionaries accessed from a smartphone	43	26%
Dictionaries accessed from a tablet	1	1%

To elicit more detailed responses on this particular topic, the participants were asked to elaborate on why they would rather use a dictionary that can be accessed from either a computer (desktop or laptop computer), smartphone or tablet (*Open-ended question: Explain why you prefer to use a dictionary which can be accessed either from a computer, smartphone, or tablet.*). Some of the typical responses are given below.

I prefer computer-based dictionaries because I spend most of my time in front of a computer.

It is much easier to search for relevant information in dictionaries via a computer.

More information can be displayed on a computer screen.

Text from a computer screen is more readable than text from a smaller mobile device.

I always carry a smartphone with me and can use it at any single moment, using smartphone-based dictionaries is simply more convenient.

I usually work using my laptop, it is more convenient for me to access dictionaries from a laptop.

I prefer computer-based dictionaries because I can quickly access information from a few dictionaries by opening multiple tabs in Chrome.

Using computers is simply more convenient.

A laptop allows me to use multiple dictionaries at the same time.

I usually use dictionaries for writing, and I write using Microsoft Word, so it is more convenient to use computer-based dictionaries.

Display size is important, a larger display size is better for my eyesight.

When I encounter an unknown word, I can use my mobile phone and quickly find the word's meaning.

I prefer smartphones to computers, so I use smartphone-based dictionaries. Smartphones are my favorite device.

Whenever classes are taught online (due to COVID-19) I always use my computer, so I prefer to access dictionaries from a computer.

I can type faster using a computer keyboard.

A smartphone allows me to check a word's meaning regardless of where I am.

I prefer smartphone-based dictionaries because I am not always at home and my laptop is turned off.

I don't use my smartphone too often, so I prefer to access dictionaries from a computer.

The advertisements in smartphones are annoying, they take up too much space on my smartphone screen, I prefer computer-based dictionaries.

I always use my laptop, there is no point in using two different devices at the same time.

I do not have a tablet.

Having two different devices (laptop and smartphone) at your disposal is more convenient because you can be involved in two different activities at the same time.

I am more accustomed to using a traditional computer keyboard rather than a touch screen.

The text is larger on a computer screen.

Dictionary entries displayed on computer screens present more information.

It depends on the situation. I prefer to use smartphone-based dictionaries during my classes. I use computer-based dictionaries at home.

In another question of the survey, the respondents were asked how they normally learned the pronunciation of a word. The participants had the opportunity to select either one or two answers depending on their own personal strategies for pronunciation search. As many as 96% of the students opted for online dictionaries and audio recordings of British and American English native speakers. It must be admitted that the fact that students discover the pronunciation of selected vocabulary items in this manner means that students are well aware of where reliable information about the pronunciation of words can be found, they are skilled at solving problems, and are resourceful enough to find the information they need independently of anyone's help. 27% of the students reported looking for the pronunciation of a word on a random Internet website, 11% would ask a friend for help, 10% would ask their teachers, while only 7% of the students said that they would use a paper dictionary for phonetic transcription. 11% of the survey participants admitted to employing other strategies when searching for a word's pronunciation.

Table 9: Responses to Question 9 (How do you usually check the pronunciation of a word in English?)

Answer option	No. of responses	Percentage
Online dictionary	162	96%
Internet website	45	27%
I ask a friend	18	11%
I ask a teacher	17	10%
Paper dictionary	12	7%
Other	19	11%

To analyze the students' decision-making when encountering an unknown word, the respondents were asked what it is that they specifically do when faced with such a situation. The following responses were elicited: 59% of the students claimed that they used a dictionary, 20% said that they browsed the Internet for information, 19% said that they tried to guess the word's meaning from context, 1% said that they asked someone what the word means, 1% said that they simply ignored the word, and 1% of the students chose the *none of the above* option. The students were asked to select only one answer, which means that it is thought-provoking why as many as 41% of the students replied that they would not engage in dictionary consultation. On the one hand, it seems obvious that some students may already have some general idea about what the word means and so they resort to the Internet, or they simply decide to guess a word's meaning from context in order to receive confirmation of the meaning of the item in question. Nowadays browsing the Internet for information is nothing out of the ordinary, which partially explains why 20% of the students chose this option. On the other hand, students of English are expected to find reliable information about a word so that they not only know what it means, but that they also manage to learn more about this word's grammatical or collocational environment. This is where dictionaries should come into play, given that the main function of dictionaries is explaining word meanings. In other words, more proficient students should endeavor to discover how to use a word in the language. Hence, it seems reasonable to assume that in such situations students should not only rely on intuition and common sense, but they ought to strictly adhere to common practices which in this specific scenario clearly involve the need for dictionary consultation. Even if such a student intends to only obtain information about what the word means, regardless of acquiring information about how this word is used in the language, dictionaries still seem to be the most logical solution in the context of a student of English.

Table 10: Responses to Question 10 (What do you usually do when you do not know what a word means in English?)

Answer option	No. of responses	Percentage
I use a dictionary	99	59%
I browse the Internet for information	33	20%
I guess the meaning from context	32	19%
I ask someone what it means	1	1%
I ignore the word	1	1%
None of the above	2	1%

To see how important illustrations are to dictionary users, the respondents in the current survey were asked whether they thought that the presence of illustrations in dictionaries was beneficial to the dictionary user. Significantly, Lew et al. (2018) focused in their eye-tracking study on pictorial illustrations in English monolingual learners' dictionaries and their effect on dictionary users' attention, as well as extraction and retention of meaning. In other words, it could be said that the researchers sought to assess the usefulness of the presence of illustrations in monolingual dictionary entries for learners of English. In the present survey, the respondents were simply asked what their opinions were on the inclusion of illustrations in dictionary entries. 33% of the respondents said that illustrations were probably useful, 16% said that they were useful, 19% said that they were probably not useful, 8% were of the opinion that they were not needed in dictionaries. 24% of the students said that it was difficult to say. Taking these percentages into account, it appears that more than half of the students (51%) thought that illustrations were either not needed in dictionary entries or the students were simply undecided. Put another way, these data illustrate that the students could be divided into two separate groups which represent opposite ends of the spectrum. Roughly 50% of the students rather perceive illustrations as beneficial in dictionary use, whereas the other half are rather doubtful or unconvinced that they are needed.

Table 11: Responses to Question 11 (Do you think that illustrations in dictionaries are useful?)

Answer option	No. of responses	Percentage
Yes	27	16%
Probably	55	33%
Probably not	31	19%
No	14	8%
Difficult to say	41	24%

Example sentences provide dictionary users with additional information about words. For instance, one can derive collocational or grammatical information from examples and learn how a particular word should be used in the language in speech or writing. The author of the present survey attempted to see whether examples² are beneficial to dictionary users. To be more precise, the participants were asked in the survey to say whether they read example sentences in English monolingual learners' dictionaries. 74% of the students admitted to the fact that they usually consulted examples in dictionaries, 21% said that they sometimes read them, 5% said that they usually did not even browse through them, whereas 1% of the students said that it was difficult to say whether they read examples in dictionaries or not. Assuming that the students are reporting here what they are actually doing during dictionary consultation, it must be contended that the data are promising, and that the students recognize the importance of reading example sentences in monolingual dictionaries for learners of English.

Table 12: Responses to Question 12 (Do you read the example sentences in English monolingual learners' dictionaries once you have familiarized yourself with the definition of a word?)

Answer option	No. of responses	Percentage
I usually do	124	74%
I sometimes do	35	21%
I usually do not	8	5%
Difficult to say	1	1%

In the following question the respondents had to say whether they thought that dictionary users could find frequency information about words in an English monolingual learner's dictionary. 117 of the respondents (70%) answered *yes*, while 51 of the respondents (30%) were of the opinion that it was not possible to find such information in a dictionary. This finding suggests that there are still learners of English (51 students, 30% of the survey participants) who need to be made more aware of what type of information English monolingual learners' dictionaries have to offer to the average dictionary user.

Table 13: Responses to Question 13 (Do you think that it is possible to find out from an English monolingual learner's dictionary whether a particular word is used frequently or infrequently in the English language?)

Answer option	No. of responses	Percentage
Yes	117	70%
No	51	30%

Given that monolingual dictionaries for learners of English are a source of grammatical information for dictionary users, the author of the survey attempted to analyze the frequency of dictionary use for getting information about grammar from dictionaries. 25% of the students claimed that they *often* tried to obtain grammatical information from English monolingual learners' dictionaries, while 7% declared that they tried to acquire this information *very often*. This finding suggests that 32% of the students were rather keen to learn more about grammar from dictionaries and endeavored to get such information regularly. On the contrary to this finding, there were students who either *rarely* (16%) or *very rarely* (10%) tried to acquire grammatical information from dictionaries, or there were students who even did not search for this type of information (*I do not look for grammatical information in dictionaries* — 5%). These data indicate that 30% of the students infrequently engage in dictionary use for the purpose of extending their grammar knowledge. One reason behind this statistic could be that dictionary users are not interested in obtaining this type of information from dictionaries. Another possibility is that they prefer to use grammar books and do not perceive dictionaries as a potential source of grammatical information. 38% of the students admitted to the fact that they *sometimes* decided to try to learn grammar from a monolingual dictionary for learners of English. To sum up, more or less a third of the students use dictionaries regularly with the aim of also obtaining grammatical information, roughly a third of the students occasionally use dictionaries for the very same purpose, whereas approximately a third of the students seldom use dictionaries to get information about grammar, or they do not use dictionaries for this purpose.

Table 14: Responses to Question 14 (How often do you try to obtain grammatical information (part of speech, countable and uncountable nouns, transitive and intransitive verbs, irregular verbs, etc.) from English monolingual learners' dictionaries?)

Answer option	No. of responses	Percentage
Very often	11	7%
Often	42	25%
Sometimes	64	38%
Rarely	27	16%
Very rarely	16	10%
I do not look for grammatical information in dictionaries	8	5%

An additional question attempted to check how frequently dictionary users consulted either the front or back matter of their print dictionaries. Not a single respondent admitted to using a dictionary for the purpose of obtaining information from the front or back matter frequently (answer option — *often*). 6% of the respondents admitted to using dictionaries in order to get information from either the front or back matter of their dictionaries *sometimes*, 24% declared that they engaged in such dictionary consultation *rarely*, whereas 70% of the students did not read or even browse through their dictionaries for such information.

Table 15: Responses to Question 15 (How often do you read the front matter or back matter of a paper dictionary?)

Answer option	No. of responses	Percentage
Often	0	0%
Sometimes	10	6%
Rarely	41	24%
I do not read the front or back matter of a paper dictionary	117	70%

As for the type of information that students usually search for in a monolingual dictionary for learners of English, the main finding of the current survey is that students primarily consult dictionaries for *meaning*. In the present survey, 95% of the students selected this answer (the students were allowed to select more than one answer for this question). Furthermore, it appears that the second reason for which dictionary users reach for a monolingual dictionary is to acquire *pronunciation* information (67%), and what is interesting is that this specific type of information in dictionaries plays a more important role for dictionary users than the presence of *examples* (49%), information about *spelling* (46%), *collocations* (39%), or *idioms* (23%) in a dictionary. Pessimistically, acquiring *grammatical information* from dictionaries is even less important for the students (17%), not to mention the information that can be obtained from either the *front* or *back matter* of a print dictionary.

Table 16: Responses to Question 16 (What type of information do you usually look for in an English monolingual learner's dictionary?)

Answer option	No. of responses	Percentage
Meaning	160	95%
Pronunciation	112	67%
Examples	82	49%
Spelling	77	46%
Collocations	65	39%
Idioms	39	23%
Grammar	28	17%
Front/back matter	1	1%
Other	1	1%

A recent empirical study by Dziemianko (2020) explored the relation between online advertisements and language reception, production and learning. Interestingly, another aim was to analyze what dictionary users had thought about being exposed to online dictionary entries appearing with advertisements. In the current survey, the respondents were asked to give their opinion on the presence of advertisements in online dictionary entries. 45% of the respondents were of the opinion that such advertisements in online dictionary entries were disturbing, 39% reported that they were not annoyed by the presence of advertisements, whereas 16% of the students chose the *difficult to say* answer. In general, the conclusion that can be drawn from the survey is that the opinions of the students vary and that one possible factor which comes into play in the present context is one of individual preferences of the students.

Table 17: Responses to Question 17 (Do you find advertisements in online dictionaries disturbing?)

Answer option	No. of responses	Percentage
Yes	76	45%
No	65	39%
Difficult to say	27	16%

In the third open-ended question (*Is there anything you do not like in dictionaries? Is anything missing from dictionaries?*), the respondents had the opportunity to

express their opinions regarding the lexicographic content of dictionaries. Most responses were positive, as the majority of students expressed their general satisfaction with dictionaries. However, roughly 10% of the students complained that they did not always manage to find the information they were looking for in an English monolingual learner's dictionary, and attributed this to an insufficient amount of information in dictionaries about how a particular word can be used in different contexts. This problem in their view could be resolved by making more example sentences available to users in online dictionaries. Additionally, a handful of respondents accentuated the need for the incorporation of more example sentences in monolingual learners' dictionaries for collocations.

6. Discussion and conclusions

The results indicate that, when it comes to bilingual dictionaries, English majors favor *Diki.pl*, *Bab.la* and *Google Translate* over the remaining online dictionaries, when dealing with either Polish–English or English–Polish translation tasks. The 'other dictionary' option (beyond one of: *Diki.pl*, *Bab.la*, *Google Translate*, *PONS*, *Ling.pl*, *Edict.pl*) was only supplied by 7% of the participants. The most frequent additions here were: *Reverso Context*, *Linguee.pl*, and *Cambridge Dictionary*. *Cambridge Dictionary* was also the students' dictionary of choice as a monolingual learner's dictionary (93%). It is, however, possible that the students chose this dictionary as their favorite monolingual dictionary not only because the dictionary gives the students access to the *Cambridge Advanced Learner's Dictionary*, but because the portal also includes bilingual dictionaries (such as the *Cambridge English–Polish Dictionary* and *Cambridge Polish–English Dictionary*) selectable from a drop-down list. In other words, this online resource guarantees students exposure to the target language alone and at the same time creates the opportunity for students to engage in either Polish–English or English–Polish translation, which appears to be an added advantage to the monolingual learner's dictionary. The *Macmillan Dictionary* was rated by the students as their second favorite monolingual dictionary (59%), whereas the remaining three monolingual learners' dictionaries (*Oxford Advanced Learner's Dictionary*, *Collins Online Dictionary*, *Longman Dictionary of Contemporary English*) were all selected by less than 50% of the students.

The data obtained from the current survey confirm earlier findings from Lew and Szarowska's study (2017), in which a systematic evaluation framework for online bilingual dictionaries was proposed and applied to the English–Polish language pair. That study found that *Diki.pl* was the most frequently used and highest-ranked extant dictionary. In the present survey, *Diki.pl* also turned out to be the most-frequently consulted dictionary, and it is reassuring that this high popularity corresponds with the outcome of Lew and Szarowska's expert quality evaluation: it appears that the users know what they are doing when they elect to use a specific dictionary and are able to

appreciate the quality.

It is by no means surprising that the overwhelming majority of the students resort to the use of online dictionaries, given the benefits of this specific dictionary format. Also, this finding is in line with the study of Kosem et al. (2019: 109) in which it was found that the younger participants clearly favored the digital format of dictionaries over the paper format. It suggests that the *digital revolution* (hailed by Lew and De Schryver 2014) has all but completed its course. However, the collected evidence points to the fact that print dictionaries cannot yet be brushed aside by lexicographers, as there are students who use the book-form dictionary, for example, for their writing assignments (essays, BA and MA theses) or when the students focus on paper-based tasks. Another reason for the use of a paper dictionary is to simply enhance the learning process, as some students reported that they believed the use of a paper dictionary to be an effective memorization technique. Whether this is true or not would need to be empirically tested; however, studies (Bishop 2000; East 2008) have demonstrated that the availability of dictionaries in the process of writing during examinations and tests brings students psychological reassurance and alleviates stress. By the same token, it is possible that given individual preferences of students and their habits of using print dictionaries rather than online dictionaries for specific tasks, the mere use of a paper-form dictionary could result in some sort of psychological support to such dictionary users, which in turn could perhaps lead to better memorization of words. In general, the use of a paper dictionary requires more time and effort, and so perhaps the fact that one devotes more attention to dictionary consultation via the paper medium can contribute to enhanced memorization of vocabulary.

The findings suggest that most English majors use dictionaries on a regular basis: 63% of the students reported that they consulted dictionaries (*nearly every day*), while 31% of the students declared that they engaged in dictionary use *a few times a week*. What needs to be stressed in the present context is that it seems reasonable to assume that the frequency of dictionary consultation must be related to students' individual preferences and hence it would be unwise to say that learners of English should engage in dictionary consultation on a daily basis. Some students simply have to use dictionaries more often than other students given their heavy workload. As long as students manage to bring back the right meaning when deciding to use a dictionary, and provided that students use a dictionary when the task at hand requires them to use one, it seems that a less sporadic use of dictionaries should not be frowned upon. When it comes to looking up more sophisticated words in the dictionary, the majority of the survey participants tend to use at least two dictionaries for consultation, which seems to be a positive trend. Nevertheless, using one dictionary can also be perceived as an optimal strategy on the condition that dictionary users find what they are looking for in a single source.

The findings show that the average dictionary user prefers to access dictionaries from a desktop or laptop computer as opposed to either a smartphone

device or tablet. This finding appears to be transferable to the larger context of European dictionary users as these observations are in agreement with those of Kosem et al. (2019), who found that their participants (roughly 10,000 respondents from nearly thirty countries) clearly favored computer-based dictionaries over smartphone-based dictionaries.

In addition, upper-intermediate and advanced-level dictionary users have a tendency to search for pronunciation information in online dictionaries, which seems to be the most reasonable choice. This should not come as a surprise as audio recordings of native speakers in digital dictionaries are easily accessible online and they provide users of the language with quick and easy answers as to how a word should be properly pronounced in the target language.

More than two fifths of the participants in the survey admitted to either browsing the Internet for information or guessing a word's meaning from context when encountering an unknown word in English, rather than using a dictionary. On the one hand, this finding is rather surprising given that English majors participated in the survey. However, taking into account the fact that the sample included 19–25-year-olds, who can safely be called digital natives growing up in a society of digital technologies, it becomes more clear why dictionary users tend to resort to simpler strategies when encountering unknown words.

The majority of the participants seemed to recognize the importance of example sentences in dictionaries. In their open-ended feedback, the students strongly emphasized their need for even more examples in dictionaries, notwithstanding the numerous examples that are already available, for instance, in the current English monolingual learners' dictionaries. Given that storage space in the context of online dictionaries is no longer a problem for the lexicographer, perhaps it would be possible to at least rethink the approach of dictionary-making in the context of online dictionaries, regarding the inclusion of the most optimal number of example sentences in dictionary entries. Studies conducted by Frankenberg-Garcia (2012, 2014, 2015) are a promising start.

Finally, I would like to make a comment with regards to the third research question posed in the survey (*Research question 3*: Do the findings of the current survey carry any pedagogical implications for dictionary use by English majors?). 30% of the participants reported that they thought that English monolingual learners' dictionaries did not provide dictionary users with frequency information, which in turn means that 70% of the participants were aware that frequency information was available. Still, a third of the participants were apparently misinformed or simply had never been taught by their teachers that this type of information could be obtained from a monolingual pedagogical dictionary for learners of English. The aim of including frequency information in such a dictionary is, for example, to teach learners which words are frequently used in the language and, consequently, which specific words students should learn to be able to communicate in the target language fluently. This

salient feature of monolingual learners' dictionaries also allows one to distinguish between words and expressions which are frequently used in either spoken or written English. To sum up, in the author's view, teaching English majors that frequency information is available in monolingual learners' dictionaries should be made a priority in schools and dictionary classes held at universities. Likewise, I would like to make a further comment in relation to grammatical information which is incorporated into dictionary entries. Roughly two thirds of the survey participants admitted to consulting dictionaries for grammar rather occasionally, rarely, or very rarely. It is true that, in general, dictionaries are not perceived as the primary source of grammatical information for students. However, it is the author's belief that the importance of the mere possibility of acquiring syntactic information from monolingual learners' dictionaries should at least be stressed by English teachers and lecturers of English departments at universities, as modern dictionaries are much more than just books or digital products consisting of word lists and their meanings. In the author's view, dictionaries constitute repositories of knowledge, or are simply reference sources packed with information about word meanings, grammar, collocations, and other types of information.

Acknowledgements

I thank Prof. Robert Lew for valuable suggestions, constructive comments and discussions.

Endnotes

1. Open-ended question responses generally require more time and hence respondents can also be reluctant to elaborate on their answers, let alone provide responses for such question items. For more information on open-ended questions in survey-based research see Muijs (2004).
2. To learn more about empirical research dealing with the usefulness of examples in dictionary use, see Frankenberg-Garcia's studies (2012, 2014, 2015) in which an elaborate research design was adopted.

References

- Bishop, G.** 2000. Dictionaries, Examinations and Stress. *Language Learning Journal* 21: 57-65.
- Blair, J., R.F. Czaja and E.A. Blair.** 2014. *Designing Surveys: A Guide to Decisions and Procedures*. Third edition. Thousand Oaks, California: SAGE Publications.
- Blaxter, L., C. Hughes and M. Tight.** 2006. *How to Research*. Third edition. Maidenhead: Open University Press.
- Brace, I.** 2018. *Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research*. Fourth edition. London: Kogan Page.
- Brown, J.D.** 2001. *Using Surveys in Language Programs*. Cambridge, UK: Cambridge University Press.

- Chon, Y.V.** 2009. The Electronic Dictionary for Writing: A Solution or a Problem? *International Journal of Lexicography* 22(1): 23-54.
- Cohen, L., L. Manion and K. Morrison.** 2007. *Research Methods in Education*. Sixth edition. London/ New York: Routledge.
- Debois, S.** 2019. *10 Advantages and Disadvantages of Questionnaires*. <https://surveyanyplace.com/blog/questionnaire-pros-and-cons/>. Accessed 22 July, 2021.
- De Schryver, G.-M., S. Wolfer and R. Lew.** 2019. The Relationship between Dictionary Look-up Frequency and Corpus Frequency Revisited: A Log-file Analysis of a Decade of User Interaction with a Swahili-English Dictionary. *GEMA Online Journal of Language Studies*. 19(4): 1-27.
- Dillman, D.A., J.D. Smyth and L.M. Christian.** 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, New Jersey: Wiley.
- Dziemianko, A.** 2020. Smart Advertising and Online Dictionary Usefulness. *International Journal of Lexicography* 33(4): 377-403.
- East, M.** 2008. *Dictionary Use in Foreign Language Writing Exams: Impact and Implications*. Amsterdam/Philadelphia: John Benjamins.
- Frankenberg-Garcia, A.** 2012. Learners' Use of Corpus Examples. *International Journal of Lexicography* 25(3): 273-296.
- Frankenberg-Garcia, A.** 2014. The Use of Corpus Examples for Language Comprehension and Production. *ReCALL* 26(2): 128-146.
- Frankenberg-Garcia, A.** 2015. Dictionaries and Encoding Examples to Support Language Production. *International Journal of Lexicography* 28(4): 490-512.
- Harris, D.F.** 2014. *The Complete Guide to Writing Questionnaires: How to Get Better Information for Better Decisions*. Durham, North Carolina: I&M Press.
- Hatherall, G.** 1984. Studying Dictionary Use: Some Findings and Proposals. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 183-189. Tübingen: Max Niemeyer.
- Humblé, P.** 2001. *Dictionaries and Language Learners*. Frankfurt am Main: Haag und Herchen.
- Jackson, C.J. and A. Furnham.** 2000. *Designing and Analysing Questionnaires and Surveys. A Manual for Health Professionals and Administrators*. London: Whurr Publishers.
- Kosem, I., R. Lew, C. Müller-Spitzer, M.R. Silveira, S. Wolfer, et al.** 2019. The Image of the Monolingual Dictionary across Europe. Results of the European Survey of Dictionary Use and Culture. *International Journal of Lexicography* 32(1): 92-114.
- Lew, R.** 2002. Questionnaires in Dictionary Use Research: A Reexamination. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002, Vol.1*: 267-271. Copenhagen: Center for Sprogteknologi, Copenhagen University.
- Lew, R.** 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.
- Lew, R.** 2015. Opportunities and Limitations of User Studies. *OPAL — Online publizierte Arbeiten zur Linguistik* 2(2015): 6-16. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/3772/file/Research_into_dictionary_use-2015.pdf. Accessed 10 September 2022.
- Lew, R.** 2016. Can a Dictionary Help you Write Better? A User Study of an Active Bilingual Dictionary for Polish Learners of English. *International Journal of Lexicography* 29(3): 353-366.
- Lew, R. G.-M. and De Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.

- Lew, R., M. Grzelak and M. Leszkowicz.** 2013. How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-tracking Study. *Lexikos* 23: 228-254.
- Lew, R., R. Kaźmierczak, E. Tomczak and M. Leszkowicz.** 2018. Competition of Definition and Pictorial Illustration for Dictionary Users' Attention: An Eye-tracking Study. *International Journal of Lexicography* 31(1): 53-77.
- Lew, R. and A. Szarowska.** 2017. Evaluating Online Bilingual Dictionaries: The Case of Popular Free English–Polish Dictionaries. *ReCALL* 29(2): 138-159.
- Macer, T. and S. Wilson.** 2013. *A Report on the Confrimit Market Research Software Survey*. <https://www.quirks.com/articles/a-report-on-the-confrimit-market-research-software-survey-1>. Accessed 23 July, 2021.
- Mackey, A. and S.M. Gass.** 2005. *Second Language Research: Methodology and Design*. Mahwah, New Jersey: Lawrence Erlbaum.
- Muijs, D.** 2004. *Doing Quantitative Research in Education with SPSS*. London: SAGE Publications.
- Nesi, H.** 2000. *The Use and Abuse of EFL Dictionaries. How Learners of English as a Foreign Language Read and Interpret Dictionary Entries*. Lexicographica Series Maior 98. Tübingen: Niemeyer.
- Schierholz, S.J.** 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25: 323-352.
- Simonsen, H.K.** 2009. Vertical or Horizontal? That is the Question: An Eye-track Study of Data Presentation in Internet Dictionaries. *Paper presented at the Eye-to-It Conference on Translation Processes, Sentence Processing and the Bilingual Mental Lexicon, Business School, Copenhagen, Denmark, April 25–29, 2009*.
- Simonsen, H.K.** 2011. User Consultation Behaviour in Internet Dictionaries: An Eye-tracking Study. *Hermes* 46: 75-101.
- Sudman, S. and N.M. Bradburn.** 1982. *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass Series in Social and Behavioral Sciences. San Francisco: Jossey-Bass.
- Sue, V.M. and L.A. Ritter.** 2012. *Conducting Online Surveys*. Thousand Oaks, California: SAGE Publications.
- Tall, G. and J. Hurman.** 2002. Using Dictionaries in Modern Language GCSE Examinations. *Educational Review* 54(3): 205-217.
- Tono, Y.** 2001. *Research on Dictionary Use in the Context of Foreign Language Learning. Focus on Reading Comprehension*. Lexicographica Series Maior 106. Tübingen: Max Niemeyer.
- Tono, Y.** 2011. Application of Eye-tracking in EFL Learners' Dictionary Look-up Process Research. *International Journal of Lexicography* 24(1): 124-153.
- Welker, H.A.** 2013. Methods in the Research of Dictionary Use. 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*: 540-547. Berlin/Boston: De Gruyter.

APPENDIX A: Original text of the questionnaire

ANKIETA (SŁOWNIKI)

Drogi Studente! Uprzejmie proszę wypełnij ankietę, która składa się łącznie z 20 pytań otwartych i zamkniętych. Czas jest nieograniczony, chociaż wypełnienie ankiety nie powinno ci zająć więcej niż 20 minut. W nawiasach podano informację ile odpowiedzi powinieneś zaznaczyć w wybranym pytaniu zamkniętym, natomiast w pytaniach otwartych nie ma żadnych limitów słów. Pomocne w ankiecie będą hiperłącza do wybranych słowników internetowych. Klikając na nie zostaniesz automatycznie przekierowany do strony internetowej danego słownika (proszę skorzystać z hiperłączy jeżeli np. chcesz sobie przypomnieć z którego słownika korzystasz). Dziękuję za uczestnictwo w ankiecie!

1. Z których słowników częściej korzystasz? (proszę zaznaczyć jedną odpowiedź)
 - słowniki papierowe
 - słowniki internetowe
2. Wyjaśnij proszę w jakiej sytuacji wolisz użyć słownik papierowy aniżeli słownik internetowy.
3. Z których słowników internetowych (angielsko-polskich, polsko-angielskich) korzystasz do tłumaczenia? (możesz zaznaczyć więcej niż jedną odpowiedź)
 - diki.pl: Słownik angielsko-polski/polsko-angielski online
 - bab.la: Słownik angielsko-polski/polsko-angielski online
 - edict.pl: Słownik angielsko-polski/polsko-angielski online
 - ling.pl: Słownik angielsko-polski/polsko-angielski online
 - pons.pl: Słownik angielsko-polski/polsko-angielski online
 - translate.google.com
 - Korzystam z innych słowników
4. Z których niżej wymienionych angielskich słowników internetowych (brytyjskich słowników jednojęzycznych) korzystasz? (możesz zaznaczyć więcej niż jedną odpowiedź)
 - ldoconline.com (*Longman Dictionary of Contemporary English*)
 - dictionary.cambridge.org (*Cambridge Dictionary*)
 - macmillandictionary.com (*Macmillan Dictionary*)
 - oxfordlearnersdictionaries.com (*Oxford Advanced Learner's Dictionary*)
 - collinsdictionary.com (*Collins COBUILD Dictionary*)
5. Jak często korzystasz ze słownika? (proszę zaznaczyć jedną odpowiedź)
 - raz w tygodniu
 - dwa razy w tygodniu
 - kilka razy w tygodniu
 - prawie codziennie
 - codziennie
6. Z ilu słowników najczęściej korzystasz, aby sprawdzić znaczenie trudnego wyrazu w języku angielskim? (proszę zaznaczyć jedną odpowiedź)
 - z jednego słownika
 - z dwóch słowników
 - więcej niż dwóch słowników
7. Jakiego słownika najczęściej używasz do pisania w języku angielskim np. eseju, rozprawki, listu, sprawozdania bądź innej formy pisemnej? (proszę zaznaczyć jedną odpowiedź)
 - jednojęzyczny
 - dwujęzyczny
 - jednojęzyczny i dwujęzyczny
 - nie używam słownika
8. Z których słowników częściej korzystasz, gdy szukasz znaczenia nieznanego ci wyrazu w języku angielskim? (proszę zaznaczyć jedną odpowiedź)
 - raczej korzystam ze słowników jednojęzycznych
 - raczej korzystam ze słowników dwujęzycznych
 - korzystam zarówno ze słowników jednojęzycznych i dwujęzycznych

9. W jaki sposób wolisz korzystać ze słownika internetowego? (proszę zaznaczyć jedną odpowiedź)
 - wolę używać słownik korzystając z komputera stacjonarnego albo laptopa
 - wolę używać słownik korzystając ze smartfona
 - wolę używać słownik korzystając z tabletu
10. Wyjaśnij proszę dlaczego wolisz używać słownik korzystając z komputera stacjonarnego (lub laptopa) albo ze smartfona albo z tabletu.
11. W jaki sposób najczęściej sprawdzasz wymowę wyrazu w języku angielskim? (proszę zaznaczyć dwie odpowiedzi)
 - sprawdzam wymowę wyrazu w słowniku internetowym (nagrania wymowy brytyjskich i amerykańskich lektorów))
 - sprawdzam wymowę wyrazu w słowniku papierowym
 - pytam się nauczyciela
 - pytam się kolegi bądź koleżanki
 - szukam wymowy wyrazu na dowolnej stronie internetowej
12. Co z reguły robisz, gdy nie wiesz co dany wyraz w języku angielskim oznacza? (proszę zaznaczyć jedną odpowiedź)
 - sprawdzam znaczenie wyrazu w słowniku
 - zgaduję, domyślam się z kontekstu
 - pytam się innej osoby
 - szukam informacji w internecie
 - nie z tym nie robię
13. Czy uważasz, że ilustracje w słownikach są przydatne? (proszę zaznaczyć jedną odpowiedź)
 - tak
 - raczej tak
 - raczej nie
 - nie
 - trudno powiedzieć
14. Czy czytasz zdania przykładowe w angielskich słownikach jednojęzycznych po zapoznaniu się z definicją danego wyrazu? (proszę zaznaczyć jedną odpowiedź)
 - raczej tak
 - czasami
 - raczej nie
 - trudno powiedzieć
15. Czy myślisz, że można dowiedzieć się z angielskiego słownika jednojęzycznego czy dany wyraz np. należy do wyrazów często lub rzadko stosowanych w języku angielskim? (proszę zaznaczyć jedną odpowiedź)
 - tak
 - nie
16. Jak często korzystasz z informacji gramatycznej (np. część mowy, rzeczowniki policzalne i niepoliczalne, czasowniki przechodnie, nieregularne formy czasowników, itp.), która jest dostępna w hasłach angielskich słowników jednojęzycznych? (proszę zaznaczyć jedną odpowiedź)
 - bardzo często
 - często
 - czasami
 - rzadko
 - bardzo rzadko
 - raczej nie korzystam
17. Jak często zaglądasz do wstępu (strony początkowe) bądź zakończenia (strony końcowe) słownika papierowego? (proszę zaznaczyć jedną odpowiedź)
 - często
 - czasami
 - rzadko
 - nie zaglądam do wstępu ani zakończenia słownika

18. Jakiej informacji najczęściej szukasz w angielskim słowniku jednojęzycznym? (proszę zaznaczyć trzy odpowiedzi)

- znaczenie wyrazu
- wymowa wyrazu
- pisownia wyrazu
- zdania przykładowe
- gramatyka
- kolokacje
- idiomy
- zaglądam do wstępu/zakończenia słownika

19. Czy coś ci się nie podoba w słownikach, czy czegoś brakuje w słownikach?

20. Czy korzystając ze słownika online przeszkadzają ci wyświetlające się reklamy internetowe?

- tak
- nie
- trudno powiedzieć

APPENDIX B: English translation of the questionnaire items, grouped into close-ended and open-ended

Closed-ended questions

Question 1: What dictionaries do you usually use? (paper dictionaries; online dictionaries)

Question 2: Which of the following online dictionaries do you use for Polish–English or English–Polish translation? (*Diki.pl*; *Bab.la*; *Edict.pl*; *Ling.pl*; *PONS*; *Google Translate*; I use other dictionaries)

Question 3: Which of the following British English monolingual learners' dictionaries do you use? (*Longman Dictionary of Contemporary English*; *Cambridge Dictionary*; *Macmillan Dictionary*; *Oxford Advanced Learner's Dictionary*; *Collins Online Dictionary*)

Question 4: How often do you use a dictionary? (once a week; twice a week; a few times a week; almost every day; every day)

Question 5: How many dictionaries do you usually use when you are looking for the meaning of a difficult word in English? (one dictionary; two dictionaries; more than two dictionaries)

Question 6: What dictionary type do you usually use for writing in English, for example, an essay, a letter, a report, or some other written form? (monolingual; bilingual; monolingual and bilingual; I do not use a dictionary, other)

Question 7: What dictionary type do you use more often when looking for the meaning of an unknown word in English? (I use a monolingual dictionary; I use a bilingual dictionary; I use both a monolingual and bilingual dictionary)

Question 8: What dictionaries do you prefer to use? (dictionaries accessed from a computer; dictionaries accessed from a smartphone; dictionaries accessed from a tablet)

Question 9: How do you usually check the pronunciation of a word in English? (I check the pronunciation of a word in an online dictionary — audio recordings of British and American English speakers; I check the pronunciation of a word in a paper dictionary; I ask my teacher; I ask a friend; I look for the pronunciation of a word on a random Internet website; other)

Question 10: What do you usually do when you do not know what a word means in English? (I use a dictionary; I guess the meaning from context; I ask someone what it means; I browse the Internet for information; I ignore the word; None of the above)

Question 11: Do you think that illustrations in dictionaries are useful? (yes; probably; probably not; no; difficult to say)

Question 12: Do you read the example sentences in English monolingual learners' dictionaries once you have familiarized yourself with the definition of a word? (I usually do; I sometimes do; I usually do not; difficult to say)

Question 13: Do you think that it is possible to find out from an English monolingual learner's dictionary whether a particular word is used frequently or infrequently in the English language? (yes; no)

Question 14: How often do you try to obtain grammatical information (part of speech, countable and uncountable nouns, transitive and intransitive verbs, irregular verbs, etc.) from English monolingual learners' dictionaries? (very often; often; sometimes; rarely; very rarely; I do not look for grammatical information in dictionaries)

Question 15: How often do you read the front matter or back matter of a paper dictionary? (often; sometimes; rarely; I do not read the front matter or back matter of a paper dictionary)

Question 16: What type of information do you usually look for in an English monolingual learner's dictionary? (meaning; pronunciation; spelling; example sentences; grammar; collocations; idioms; front matter or back matter; other)

Question 17: Do you find advertisements in online dictionaries disturbing? (yes; no; difficult to say)

Open-ended questions

Question 1: Explain in what situations you would prefer to use a paper dictionary rather than an online dictionary.

Question 2: Explain why you prefer to use a dictionary which can be accessed either from a computer, smartphone, or tablet.

Question 3: Is there anything you do not like in dictionaries? Is anything missing from dictionaries?

APPENDIX C: Online dictionaries

Bab.la. (<https://pl.bab.la/>).

Cambridge Dictionary Online. (<https://dictionary.cambridge.org/>).

Collins Online Dictionary. (<https://www.collinsdictionary.com/dictionary/english>).

DICT: English–Polish Dictionary. (<https://edict.pl/dict?lang=EN>).

Diki: Słownik angielskiego Online. (<https://www.diki.pl/>).

Google Translate. (<https://translate.google.pl/?hl=pl>).

Ling.pl. (<https://ling.pl/>).

Linguee.pl. (<https://www.linguee.pl/>).

Longman Dictionary of Contemporary English Online. (<http://www.ldoceonline.com/>).

Macmillan English Dictionary Online. (<http://www.macmillandictionary.com>).

Oxford Advanced Learner's Dictionary. (<https://www.oxfordlearnersdictionaries.com/>).

PONS. (<https://pons.pl/>).

Reverso Context. (<https://context.reverso.net/translation/>).

Stigmatised Dictionaries Housing a Stigmatised Variety of English: The Use of Korean English Online Dictionaries as a Teaching Tool within the EFL Classroom*

Alex Baratta, *Manchester Institute of Education, University of Manchester, Manchester, United Kingdom* (alex.baratta@manchester.ac.uk)

Abstract: This paper adopts a practice-based focus, explaining how online dictionaries can be used as pedagogic tools for teachers within the English as a Foreign Language (EFL) classroom. A discussion is provided of online dictionaries which are created by, broadly speaking, 'laypeople', and not professional lexicographers; and the vocabulary in question pertains to expanding circle Englishes, such as those tied to the Far East, with Korean English presented here as representative of this circle. However, online dictionaries, notably the variety described here, are often viewed negatively, based on judgements made regarding their accuracy, and lacking the prestige of being compiled by professional lexicographers. Further, varieties of English which reside outside the context of standard inner-circle varieties (e.g. American English), are often regarded as incorrect and/or inferior to the standard variety spoken by native speakers of English; indeed, the standard inner-circle variety is often regarded as *the* standard. However, this paper seeks to demonstrate how both online dictionaries and expanding circle Englishes have a role to play in the EFL classroom. In doing so, this can help to address broader issues tied to linguistic prejudice and hegemony, providing a more equal approach to language, to include its online compilation.

Keywords: ONLINE DICTIONARIES, STANDARD ENGLISH, INNER-CIRCLE ENGLISH, NON-INNER CIRCLE ENGLISHES (NICE), EXPANDING CIRCLE OF ENGLISH, KOREAN ENGLISH, INTERCULTURAL COMMUNICATION, TRANSLATION

Opsomming: Gestigmatiseerde woordeboeke bevat 'n gestigmatiseerde variasie van Engels: Die gebruik van aanlyn woordeboeke van Koreaanse Engels as onderrighulpmiddel in die EVT-klaskamer. In hierdie artikel word daar op die praktykgefokus: Daar word uiteengesit hoe aanlyn woordeboeke as pedagogiese hulpmiddels vir onderwysers in die Engels-as-Vreemde Taal- (EVT)-klaskamer gebruik kan word. Aanlyn woordeboeke wat oor die algemeen geskep word deur 'leke', en nie professionele leksikograwe nie,

* A version of this article was presented as a paper at the 26th Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by Stellenbosch University, Stellenbosch, South Africa, 27–29 July 2022.

word bespreek, en die tersaaklike woordeskat het betrekking op die uitbreidende kring van Engels, soos dié wat verbind word met die Verre Ooste, met Koreaanse Engels wat hier aangebied word as verteenwoordigend van hierdie kring. Aanlyn woordeboeke, veral die soorte wat hier beskryf word, word egter dikwels, gebaseer op oordele rakende hul akkuraatheid en weens die gebrek aan prestige aangesien hulle nie saamgestel word deur professionele leksikograwe nie, as negatief beskou. Boonop word die variëteite van Engels wat buite die konteks van die standaard binnekring variëteite (bv. Amerikaanse Engels) val, dikwels beskou as onjuis en/of minderwaardig aan die standaard variëteit wat deur moedertaalsprekers van Engels gebruik word. Die standaard binnekring-variëteit word inderdaad dikwels beskou as *die* standaard. In hierdie artikel word daar egter gepoog om te demonstreer hoe beide aanlyn woordeboeke en die uitbreidende kring van Engels in die EVT-klaskamer 'n rol kan speel. Sodoende kan wyer kwessies rakende linguïstiese vooroordeel en oorheersing aangespreek word om 'n meer gelyke benadering tot taal te verskaf wat die aanlyn samestelling daarvan kan insluit.

Slutelwoorde: AANLYN WOORDEBOEKE, STANDAARD ENGELS, BINNEKRING ENGELS, NIE-BINNEKRING ENGELS (NBKE), UITBREIDENDE KRING VAN ENGELS, KOREAANSE ENGELS, INTERKULTURELE KOMMUNIKASIE, VERTALING

Introduction

The central theme within this paper is that of inclusivity, applied to both dictionaries and the particular languages that they house. Specifically, this paper focuses on online dictionaries, defined broadly, and those which present the vocabulary of expanding circle Englishes (e.g. China, Brazil, Russia). This circle represents countries in which English is a foreign language, yet widely used, and by implication, not spoken by native speakers. The common link both dictionary and language have in this case is that they can be viewed negatively, seen as lacking authority. However, this paper discusses the ways in which online dictionaries which present the vocabulary of expanding circle Englishes have a central role to play in the EFL classroom, using Korean English as the focus here. In doing so, a more egalitarian approach can be taken to both language and the more modern means to present vocabulary, demonstrating how both can be a useful tool to teach an international language.

A typology of online dictionaries

Dictionaries are no longer tied solely to print-bound creations, given the proliferation of online dictionaries (Tarp 2019) which represent an innovation as part of the suggested Fourth Industrial Revolution (Schwab 2015). This refers to the ways in which technology has spread its influence within society, involving, as Tarp (2019: 226) puts it, a "disruptive explosion of technological innovations". Dictionaries are a part of this, seen with web-based dictionaries of course, but also including DVDs and CD-ROMs. While Tarp exhorts us to address the ways in which dictionaries can move forward within this technological context, he points out particular issues with online dictionaries. These issues will be

addressed, after first considering the variety within the otherwise singular category of 'online dictionary'.

Indeed, it is important to point out that online dictionaries can take several forms. In the first instance, they must, by definition, include a list of vocabulary from a particular language or language variety, and as an online source, they must be available via the Internet through a web browser. This, I would suggest, is an accurate definition of both 'online' and 'dictionary'. While I further suggest that this definition is hardly controversial, its broadness allows us to understand online dictionaries from a more inclusive perspective. On the one hand, an online dictionary could indeed be a copy of a print-based variety, providing the same content as its print-based component (such as the OED). However, online dictionaries could equally apply to word lists or blogs, which nonetheless serve the same purpose: to provide a list of lexis. Thus, I use the term online dictionary broadly for this paper, but not inaccurately. This is because the term would, by definition, include online lists of lexis in many forms and, given the spirit of inclusivity that this paper champions, we should not rule out any form of online source whose purpose is indeed to present the lexis of a language variety.

This points toward a typology of online dictionaries, from those which represent online versions of prestigious dictionary varieties, to those compiled by individual lexicographers as part of, say, a passion project, or indeed a blog which presents lexis of a language variety which is designed by a language enthusiast in their free time. Nesi (2000: 842) proposes a typology which involves the Internet dictionary, glossaries of courseware, dictionaries for CD learners, and pocket electronic dictionaries. The Internet dictionary itself is hardly monolithic, and can be realised as different varieties, a point already made. Campoy Cubillo (2004), for example, points to varied aspects of online dictionaries, such as those which include links to the cultural aspects of a given language, interactive features such as mechanisms for online feedback, and online sources in which a dictionary is but one of the site's resources (such as a blog). Given the variety inherent within online dictionaries, it can be a category that is indeed "difficult to classify" (Campoy Cubillo 2004: 49), with Campoy Cubillo further discussing typology in terms of users, needs and skills.

Even further, we can consider *genres* of online dictionaries (Luzón Marco 2004), thus allowing for subcategories to be presented of the type 'online dictionary'. Luzón Marco (2004) explains how genres are dynamic and subject to change and thus, while there are common features which indeed identify a genre as just that, there are differences and variation to be expected, based on the purposes of the individuals and communities who create online dictionaries. Thus, we can consider online dictionaries as "genres in evolution" (Luzón Marco 2004: 75). Berkenkotter and Huckin (1995: 17) elaborate on this, explaining that there is a need "to take a more articulated approach (to genre) in which individual texts are seen to contain heterogeneous mixtures of elements, some of which are recognizably more generic than others". Given the inherent variety within online dictionaries, and considering the focus on inclu-

sivity, it makes sense to include examples of dictionaries which can indeed range from the OED to word lists compiled by language enthusiasts. Both examples, and those in between, nonetheless share the same function of presenting the lexis of a given language variety.

It is beyond the scope of this paper to suggest revisions, refinements and/or updates to the important category of online dictionary typology and/or genre. However, it is hoped that the inclusive approach that is suggested within the literature presented is sufficient for us to regard a broad category beyond more rigid definitions and understanding.

The legitimacy of online dictionaries

Web-based dictionaries are argued to be of dubious quality, given "inadequate and even incorrect information" (Tarp 2019: 227), with Green (2012) further suggesting that this may be due to the fact that they are not necessarily compiled by professional lexicographers. Essentially, anyone with the requisite technological knowledge can create their own online dictionary, to include a personal blog which incorporates such. Implicit in this is the fact that many individuals who rely on technology and seek a quick route to information retrieval may have to consider the quality, or lack thereof, regarding the online dictionaries they access. Nonetheless, it is precisely this category of online dictionaries — those created by laypeople, as it were — that comprises the focus within this paper. However, this broad term needs clarification.

First, 'laypeople' can be used as a default word to simply refer to those who are not professional lexicographers by trade. However, this would not exclude professionals, such as university lecturers and linguists, for example, those who we would associate with some degree of relevant knowledge. The issue, I would suggest, is that Dr Smith from X university might not be regarded as highly as a team of professional lexicographers who compile online versions of the more authoritative dictionaries such as the OED and Merriam Webster. Thus, online dictionaries per se are not the issue, at least not in terms of authority, respect and trust, if they are produced by such organisations. Nonetheless, there is evidence that the 'big names' in language education, such as the British Council, do not always propagate accurate information in terms of EFL learning on their website, including vocabulary usage (Baratta 2021). However, the British Council has the pedigree that a lone individual might not, even if he/she is a researcher in a relevant field such as Education.

Moreover, the term 'laypeople' arguably extends more to individuals who are not associated with academia per se, but perhaps have a personal interest in creating dictionaries. This is especially true when such creation involves the vocabulary of a language, or language variety, which these individuals have a passion for. This could include compiling the vocabulary of the artificial languages of Klingon or Elvish, used in the worlds of *Star Trek* and *Lord of the Rings* respectively. Or more recently we might consider the online dictionaries

dedicated to the language of *Harry Potter*, including words such as *muggle* (an individual without magical ability).

Granted, we could dismiss such online dictionaries as merely representing fan sites, and thus not 'academic' enough (especially given that they represent fictional worlds and/or artificial languages). However, such abrupt dismissal would be a case of not seeing the forest for the lexicographical trees, and I assert this for several reasons. First, though a narrow example, prestigious publishing houses are getting in on the act regarding fan fiction; Bloomsbury, for example, has published an online glossary for the vocabulary used in the *Harry Potter* series. This can help to not merely accommodate the fan base for this series, but it also adds a degree of linguistic respect perhaps, in that the online glossary is published by a major company such as Bloomsbury, who of course published the *Harry Potter* books. Such passion for creating dictionaries need not be tied solely to fan fiction; indeed, as I will demonstrate, there is ample evidence for online dictionaries devoted to a real language — that referred to as Konglish, or Korean English as I use the term (Baratta 2019; Baratta 2021). This is a variety of the English language as used in Korea and by Koreans, consisting of its own specific lexis. This is a second consideration regarding the importance of online dictionaries.

Third, individuals who compile such language, from students, to chefs, to world travellers, given their vested interest in the creation of online dictionaries, indeed have skin in the game, regardless of their 'day job'. This is based on personal experience with the language per se, whether as an everyday speaker, EFL teacher residing in the relevant country (e.g. South Korea) or simply as a language enthusiast (Baratta 2021). We should not discount the importance of such contributions from individuals who wish to share lexical information and with this, very often share cultural insights also regarding the country whose lexis is represented online. In the case of South Korea, given the current *hallyu*, or cultural wave regarding this country (Potayroi 2014; Touhami and Al-Haq 2017), there is ample evidence of a dedicated fan base, whose knowledge of Korean culture — based on cuisine, K-POP and/or K-drama — is accurate; this would include their knowledge of Korean English also, as such individuals are *users* of this variety of English. Touhami and Al-Haq (2017) in fact cite the use of Korean English — and Korean — amongst Algerian K-POP fans, to include its use as part of online forums, but also everyday speech (an example of the latter category includes non-Koreans adopting Korean pronunciation for certain words, so that *coffee* is realised more as *koppée*). Such use can include other varieties of English of course, such as Englishes tied to India, Ghana or Singapore.

Fourth, a linguistic dilemma is that to access the lexis of such a variety of English, it is online dictionaries which are arguably the only source in the first instance. Thus, while there is evidence that a trusted source such as the OED (whether print or online) incorporates borrowed words from Korean (e.g. *mukbang*), this is not Korean English of course. There is a notable excep-

tion, to be fair: the OED includes the Korean English words of *skinship* and the interjection *fighting* (though often pronounced as *hwaiting* or *paiting* by Koreans). *Skinship* is a blended term which derives entirely from English. However, for those who seek a more extensive list of Korean English vocabulary, they will need to look beyond the OED and seek out online dictionaries. The dilemma I refer to is not that online dictionaries (as described in this paper) are somehow second-best; rather, the issue is that we don't have wider representation in the first place, with the exception of a major dictionary including just two words of a given variety of English. Thus, online dictionaries are often the dominant source of information for those wishing to access, and learn, the vocabulary of Englishes beyond the context of native speaker varieties.

To illustrate the word *skinship* in the OED, its entry is provided below (found at www.oed.com/view/Entry/92474715):

skinship, *n.*

Pronunciation:

Brit. Hear pronunciation/'skɪnʃɪp/

U.S. Hear pronunciation/'skɪn_ʃɪp/

Origin: Formed within English, by compounding. **Etymons:** SKIN *n.*, KINSHIP *n.*

Etymology: Blend of SKIN *n.* and KINSHIP *n.*

Esp. in Japanese and Korean contexts: touching or close physical contact between parent and child or (esp. in later use) between lovers or friends, used to express affection or strengthen an emotional bond.

This is clearly an example of how Koreans have made English their own, and its inclusion in the OED arguably helps to cement a more 'official' status. However, I have argued extensively against the notion of more traditional means of linguistic codification being viewed as the *sole* means of codification (Baratta 2021). In other words, to regard the lexis of a particular language variety as only having linguistically arrived once it is compiled in dictionaries published by authoritative sources would be to ignore the societal reality of language. This societal codification is initiated with the speakers themselves on the ground level. After all, without a language in existence in the first instance, there would be nothing to compile in dictionaries regarding its lexis. From this starting point of the speakers themselves, the language can be further codified by public signage — consider the term *Grand Open* used in Korea, seen throughout the country on flyers for new cafes, tapestries draped over new department stores and printed on menus for new restaurants. Are these not also 'authoritative' sources helping to spread a language's vocabulary, given that they reflect the language being used in society? Going further, there is evidence of such vocabulary used within web-based media, such as the term *officetel* used on Korean real estate websites, and referring to a building which functions as both one's residence and place of business (a blend of 'office' and 'hotel').

Thus, speakers of a language initiate, spread and establish their language variety and in doing so, they help to codify it in the process. Kruger and Van Rooy (2017) use the term conventionalisation to refer to what I am calling (societal) codification. Kruger and Van Rooy, however, do not consider codification to have taken place until the relevant feature, here vocabulary, has been accepted for publication in, say, a national newspaper and thus, gained editorial approval. Such approval again is implied to represent some kind of linguistic officialdom, but the language in question is already in official use by its speakers. As Mufwene (2001: 106) explains, "it is those who speak a language on a regular basis — and in a manner they consider normal to themselves — who develop the norms for their communities".

Linking to the modern theory of lexicography, one postulate is the need for dictionaries to reflect the needs of a certain group of people in specific social situations (Bergenholtz and Tarp 2003). One particular group comprises individuals with a need to access the vocabulary of expanding circle Englishes within the extra-lexicographical situation of Internet usage. This may well include EFL teachers who seek to use such vocabulary for their own personal knowledge if residing in a certain country, or as a teaching and learning tool within their classroom — this latter point will be discussed in this paper. We should also consider that online dictionaries are a modern approach to lexicography, emphasising "democracy and equal access to meaning-making rights" (Damaso 2005: 4). This latter point also suggests a more egalitarian approach to language, not just in terms of the variety of language per se but also in terms of the means of its compilation. In addition, the societal codification which I have discussed indeed includes online dictionaries (Cotter and Damaso 2007), with such dictionaries a more modern approach to the traditional print-based dictionary.

Dolezal (2006: 395) acknowledges the power that dictionaries have to codify language, explaining that dictionaries "confer legitimacy on a language". This is mentioned in specific regard to the variety of Englishes that exist in the world, with this aforementioned legitimacy based on "identifying and establishing the varieties of Englishes that are used in various locations around the world". The authority of the dictionary that Dolezal also references should not be centred, however, on a small number of publishers which have been regarded as trusted and authoritative sources for over a hundred years, such as the OED; this authority should also apply to online dictionaries. And from here, the ways in which such dictionaries can be used within the teaching of English as a Foreign Language (or any foreign language, by extension), can be harnessed as an effective teaching tool. Given that Englishes used by native speakers are the focus within EFL pedagogy, varieties which fall outside this category are sometimes stigmatised; a point I will come back to.

Furthermore, there are additional points to consider regarding this particular variety of online dictionaries. Given that the Internet as a whole allows for a public forum, whether YouTube, websites or blogs, this can create a pop-

ulist dictionary (Damaso 2005). This can be seen in cases where websites allow for the viewing public to post comments, and this in turn can lead to a public discussion and debate. Cotter and Damaso (2007: 1) refer to this as an "emergent dictionary genre", and this involves another layer — public input — to consider as part of the intersection between dictionary and technology. Thus, creators and users can, in part, collaborate and cooperate in terms of meaning-making, with an authoritative editor replaced with public usage. In addition, whatever lack of authority some may attribute to laypeople in their compilation of online dictionaries, Green (1996: 11) reminds us that traditional dictionaries are also "the product of human beings [who] bring their prejudices and biases into the dictionaries they make". Whatever these biases might be, some online dictionaries allow for readers to post their comments and views nonetheless, to include addressing points which they indeed might take issue with. In this way, there is the potential for such online dictionaries to perhaps be considered 'the people's dictionary'.

Herein lies the focus of this paper: to discuss the ways in which a stigmatised variety of both dictionary and English can nonetheless join forces within the EFL classroom, as an effective means to help students acquire the native speaker variety of this international language.

The judgements made against non-inner circle Englishes

Beyond the need to recognise the importance of online dictionaries, and doing so from a broad perspective, we also need to consider the general societal push for equality and recognition of diversity as seen from a linguistic standpoint. Widdowson (1994: 385) sums this up well, notably in terms of the negative attitudes that suggest linguistic *inequality* regarding certain varieties of English:

How English develops in the world is no business whatsoever of native speakers in England, the United States, or anywhere else. They have no say in the matter, no right to intervene or pass judgement. They are irrelevant ... it (English) is not a possession which they lease out to others, while retaining the freehold. Other people actually own it.

This captures the essence of language development, in as much as all languages are of course dynamic, not static. A given language spoken in one specific area will change with each generation, so it is hardly surprising that a language will change when it is exported throughout the world. Further, Kermas (2012: 75) argues that "lexicographers need to address the culture-specific dimension of knowledge sharing in today's global village and broaden their cultural viewpoint". This point is made in direct relation to the relative lack of vocabulary within dictionaries (such as the OED) that reflects English varieties which exist outside the so-called inner circle (Kachru 1991), the circle devoted to native speaker varieties in countries such as the USA, UK, Canada, Australia, New Zealand and Ireland. While this quotation from Kermas points to a lack of cer-

tain English vocabulary within dictionaries, online dictionaries already fill this gap, and often include entries which discuss the culture of the country whose vocabulary is being presented. The importance, then, of being more inclusive of Englishes around the world and their vocabulary is being noted, and inclusivity needs to extend to language, given that language in turn acts as a proxy for culture. Beyond the inner circle of English, there is also the outer circle which consists of countries which have a historical presence of English, often due to colonialism, such as Indian English. The expanding circle incorporates countries in which English has a dominant presence, but lacks the historical hold seen with the outer circle; Korea is an example of the expanding circle. And all of these circles use English for both intra- and international purposes, with a lexis (and grammar) which serves their speakers well.

However, non-inner circle Englishes are often marginalised, based on lexis which diverges from prestigious English varieties such as the inner-circle standard (Galloway and Rose 2015; Meriläinen 2017; Matsuda 2020). A good example is Singapore's Speak Good English campaign, in which the government had tried to discourage the people's use of Singlish. While inner-circle standard is the variety which has the most prestige, given its usage in contexts such as education, government and business, the linguistic reality is that all languages and varieties within are equally legitimate. This legitimacy is based on the fact that the language in question has a systematic and predictable use of lexis (and grammar), and for its speakers gets the job done in terms of effecting communicative success. The societal reality, however, is often based on prejudiced notions of absolute correctness, encouraged perhaps by the proliferation of a one size fits all standard of English in the classroom, and indeed represented within prestigious dictionaries and school-based textbooks. However, there is simply no linguistic basis to consider a language variety as primitive at worst, or a failed attempt to learn correctly at best.

From this discussion, it is argued that all language varieties are legitimate, regardless of societal sanctions placed on them, and there are individuals who have a desire to know more about the many varieties of English, with Korean English one example. Online dictionaries are also a legitimate means to spread and reinforce the lexis of Korean English, created by individuals who are users, or certainly enthusiasts, of the language variety under discussion on their websites and blogs. The discussion now continues and explains the ways in which online dictionaries can be used within the context of EFL teaching and learning.

The implications for EFL pedagogy

With regard to EFL pedagogy, both online dictionaries and non-inner circle Englishes (NICE) have a part to play. My discussion in this section focuses on Korean English, but not to the exclusion of other NICES; indeed, the overall purpose is to approach language from an inclusive perspective. Nonetheless, the focus on Korean English essentially functions to place the spotlight on a

particular variety of English, while also acknowledging its function as a stand-in for NICE in general and their role within the EFL classroom.

First, we need to consider that to approach English from a monolithic perspective within teaching is not reflective of the real linguistic world. Current EFL pedagogy (Matsuda 2002) in fact emphasises the need to consider English from an international perspective, a means to consider the "socio-cultural functions of global English" (Lovtsevich and Sokolov 2020: 703). Rose (2017: 173) elaborates, arguing that "by not exposing learners to the diversity of English, teachers are doing their learners a disservice by ill-equipping them to use English in the future with a wide variety of speakers who will not conform to the unrepresentative standards promoted in traditional English as a foreign language (EFL) classrooms". The implication here is that even if a specific variety of English is prescribed for a given EFL classroom, such as standard inner-circle English, it is necessary to expose students to other varieties, more so if they intend to live overseas in which they will be exposed to speakers of multiple Englishes, some of whom are not native speakers. This could also include of course using a NICE as a pedagogical tool to help teach inner circle English. Moreover, my suggestion is more specifically focused on using the local variety of English that the students are already familiar with, as a means to teach the inner circle standard variety. Thus, Korean EFL students, already familiar with Korean English, can have their variety of English used to help teach the inner circle variety. Moreover, if we consider inner-circle dialects of English, then clearly there is much linguistic variety within just a single country, let alone globally.

However, the purpose is not merely to expose students to multiple Englishes in order for them to learn such varieties and go beyond a singular classroom focus on English, important though this is. Another important reason to bring NICE into the EFL classroom is to use them as a means to teach the prescribed variety of English. Thus, Korean English can be used to teach students inner-circle standard English, assuming that is the dominant classroom focus. In both cases, other varieties of English, which might otherwise be disregarded — even by the students who themselves speak such varieties — can be taught in their own right *and* used as a teaching tool as well. The former function is actually important for the teachers perhaps more so than the students, a means for them to avoid an otherwise singular pedagogic focus on one variety of English, and the culture(s) which is part of this. In fact, there is evidence that inner-circle English speakers, as part of their teaching preparation, are being exposed to NICE, as a means to develop respect for such varieties and the cultures within, and recognise their validity as varieties of English (Brown 1993; Baratta 2019). This pedagogy is used in universities in countries such as the UK, USA, Sweden, Germany, South Africa and Japan, and reflects the need for future EFL teachers to promote cultural and linguistic diversity in their future teaching, by first being exposed to such in their current learning (Friedrich 2002; McKay 2002; Matsuda 2002; Galloway 2017).

This is not to suggest that inner-circle EFL teachers would deliberately

champion *their* English over their students'; but the focus on just one variety of English can reinforce ideas that NICE are, by definition, incorrect, and something to be avoided. Thus, a classroom of linguistic and cultural respect can be forged by bringing NICE into the EFL classroom, to (a) help both teachers and students learn from each other; and (b) demonstrate the systematic nature — and communicative ability — of multiple varieties of English. Bamgbose (1998: 3) further asks a telling question, "why should a native-variety-based standard continue to license the norms of non-native Englishes?" This indicates a key issue regarding some people's attitudes toward NICE, and inner-circle dialects for that matter, in that they are sometimes viewed as failed attempts to master the standard. Again, I do not suggest that linguistic prejudice is an unfortunate reality that can be ignored, nor can the prestige afforded to some varieties of English over others be disregarded. But once again, the additional reality is that there are more non-native speakers of English than native speakers, and thus more NICE-speakers. While standard inner-circle English is indeed the variety required for writing academic essays and passing exams such as IELTS, for informal communication amongst friends, there is clear evidence of NICE being deployed, in Korea (Rüdiger 2019), Singapore (Leimgruber, Siemund and Terrassa 2018) and India (Trudgill and Hannah 1994).

This background which has discussed matters from a more linguistic perspective has been necessary in order to better situate the focus that now follows on online dictionaries, and how they can be used in the EFL classroom. This derives from my book publication in 2021, *The Societal Codification of Korean English*, in which online dictionaries were a prominent aspect of the aforementioned societal codification. Specifically, I had obtained a list of online dictionaries dedicated to Korean English, reflecting the specific type of web-based dictionary under discussion here. The search terms used were *Korean English*; *Konglish*; *Korean English dictionary*; and *Konglish dictionary*; the term *Konglish* was deemed necessary, given its common societal use over Korean English. The entries deriving from the first page of each search result were selected, and this resulted in a total of nineteen websites. I present three examples of the dictionaries which respectively derive from an EFL teacher in Korea; a Korean individual; and a dictionary compiled by an Australian individual with a keen interest in Korean culture, to include its use of English. Again, while these individuals are not professional lexicographers per se, they have accurate knowledge of the vocabulary of Korean English. And while such online dictionaries may lack the prestige afforded to certain print-based dictionaries (and their online versions), we cannot wait for the vocabulary of NICE to be compiled in a more 'traditional' manner given this will take more time, and we already have the vocabulary in abundance within online dictionaries. Further, Pikilnyak et al. (2020) advocate the usage of online dictionaries in pedagogical contexts, arguing that the benefit of dictionaries in learning a language is increased through the online medium. Specifically, they state that "the previously existing monopoly on knowledge has disappeared. Now everyone can use the knowledge and create it. This provides the main trends in education —

free access to education and using open resources" (page 71).

Samples of online dictionaries of Korean English

The online dictionary below is designed by Leon Priz, who spent ten years in Korea and who was an educator for twenty five years. His website offers a great deal of information on Korean English vocabulary, and more besides (to be described).

Accessory	jewelry	extra, secondary parts for anything, including jewelry, but not commonly used to mean jewelry
Angle	bracket (used in construction)	The space between any two lines that intersect or come together end-to-end, relative to the 360° (degrees) of a circle. i.e., an angle can be between 0° and 360°
Bond	super glue	any binding substance or an emotional attachment
Booking	introduction request at a night club	making a reservation
Burberry	overcoat or trench coat	Trademark of light, long, waterproof coat
can 맥주	canned beer (or can of beer)	can (n.) = metallic cylinder (not adjective)
Cider	sweetened soda water (like the brand: 7up)	[US] cider = apple juice, usually homemade [UK] cider=alcoholic apple juice [CAN] hard cider = alcoholic apple juice soft cider = non-alcoholic apple juice

Figure 1: Leon's Planet (About leonsplanet.com)

The next online dictionary is designed by Miri Choi, who is a Korean teacher.

Konglish	The Inferred Meaning by Britain students	The Real Meaning in Korea
스킨 (skin)	natural covering of your body	toner
선크림 (suncream)	?	sunblock
세트메뉴 (set menu)	?	combo
원룸 (one-room)	a room	studio apartment
화이트 (white)	color	whiteout or correction fluid
세일 (sale)	selling	discount
비주얼 (visual)	relating to sight, or to things that you can see	appearance
사인 (sign)	a mark or shape that always has a particular meaning	signature
바버리코트 (Burberry coat)	?	trench coat

Figure 2: English words you only hear in Korea (English Words You Only Hear in Korea | by Miri Choi | Story of Eggbun Education | Medium)

A final example of an online dictionary is designed by an Australian university student, who has a keen interest in many aspects of Korea and its culture, to include of course Korean English.

	Korean term	Meaning	
<i>Direct loan words</i>	팀	Team	Team
	버스	Bus	Bus
<i>Clipping</i>	셀프	Self	Self-service
	아파트	Apart	Apartment
<i>Semantic shifts</i>	미팅	Meeting	College student's blind date
	안전벨트	Ahn jeon belt	Seatbelt(Ahn jeon: Korean for safety)
<i>Mixed-code combinations</i>	감자 칩	Gam ja chip	Potato chip(Gam ja: Korean for potato)
	백미러	Back mirror	Rear view mirror
<i>Fabrications</i>	베이글 녀	Bagel nyeo	Used to describe a woman with a b aby face and g lamorous body(Nyeo: Korean for girl)

Figure 3: Sydney to Seoul dictionary (sydneytoseoul.wordpress.com)

From these three examples of online dictionaries, we can now consider the next step: how can they be used within the context of EFL teaching? One possibility is to provide students with online links to a variety of dictionaries, from which they must then learn the vocabulary. Following this, a classroom test might be

administered some time later. One useful example of testing involves having students *translate* words from one English into another English variety. This is one key way that students can be taught two different Englishes, but both can be used to understand the other in the process. Moreover, by focusing on translation, it avoids the notion that NICE vocabulary is incorrect or, at the very least, that inner-circle standard is the only variety of English worth pursuing. Therefore, rather than ask students to 'correct' the Korean English term *burberry*, they instead would be asked to translate this word into inner-circle English. Of course, even inner circle is far too broad at times, given that *burberry* could be translated into *trench coat* in the US, but many British individuals might well use the word *rain coat* as well.

For example:

Korean English	American/inner-circle English
<hr/> <i>Back mirror</i>	<i>Trench coat</i>
<i>Gagman</i>	<hr/> <i>Comedian</i>
<i>Grand open</i>	<i>Grand opening</i>

Therefore, students can be provided with a word in Korean English and asked to provide the meaning in American English, for example, or vice versa. In addition, by providing students with some otherwise unfamiliar NICE vocabulary, this can further help students apply their critical thinking skills, attempting to understand the logic inherent in one variety of English as much as it is equally 'logical' in another. This can be seen clearly in examples such as *cell phone* (American English), *mobile phone* (British English) and *hand phone* (Korean English). Moreover, using the online dictionaries as both a lexical, and visual, teaching aid, students can be taught the nature of language from a more purely-linguistic perspective. For example, if the Korean English word *burberry*, referring to a trench coat, is 'wrong', then so are the words *hoover* (British English) and *kleenex* (American English). None of these examples are in any way wrong of course, in large part because they have so much use within society and society decides for itself how it will communicate. But these particular examples, in which a brand name is attributed more broadly to an object, is nothing new in linguistics — this is simply an example of overextension. Such a focus can play a large part in TESOL programmes, helping future EFL teachers, regardless of which English circle they derive from, to understand how language works and develops; this can assist in EFL teaching, notably in terms of grammar classes; and ultimately, it can help students and teachers alike approach language in the manner it should be approached — with linguistic objectivity and not based on knee-jerk reactions revolving around notions of '(in)correct'.

I recognise, as some readers may have done already, that there is what we might call an error from an absolute sense in Choi's dictionary — the expres-

sion *Britain students*. Within the inner-circle of English, and perhaps beyond, we would expect an adjective to be used when describing one's nationality: *British students*. If we can determine, however, that *Britain students* would be the term used in Korean English (and all its possible extensions, such as *Britain teachers*, *America doctors*, *France pilots*, and so on), then this is not an error. I cannot claim to know whether this is a feature of Korean English or not, though if it is an error, then I would suggest it is based on negative transfer from Korean to English. This is because in Korean, the country name followed by the word for people (사람, *saram*) is how nationalities are expressed. Thus, in literal Korean, the correct expression would indeed be 'he is a Britain person'. Given that Choi also uses 'British students', however, then this suggests that her use of *Britain students* is indeed an error. It could be regarded as one of the weaknesses of online dictionaries compiled by individuals who, while having knowledge of the language in question, are not necessarily professionals and further, there is not the same kind of editorial scrutiny that would be involved with, say, the OED. This is entirely a legitimate claim of course, though I approach this from a different perspective, one that is entirely relevant for a classroom context.

Specifically, in the context of a language-learning class, such as EFL, a discussion of errors, negative transfer, cultural influence on language and indeed, determining an error from an innovation (Jenkins 2009; Gut 2011), is a useful approach. First, by having such discussions, it can help students to think critically, and ponder relevant topics for a language class, such as intercultural communication, translation and how one language can be influenced by another. Second, this can help to remove negative judgements overall, whether pertaining to language proficiency and/or online dictionaries. This is because the use of such dictionaries in the classroom fully pertains to language learning in the first instance, not language judgement or questioning someone's linguistic ability (and here, the linguistic ability pertains to the writer's inner-circle English, and not the absolute focus of *Korean English*).

The aforementioned interactive aspects of online dictionaries are also reflected with the three online dictionaries. Priz, for example, provides an email address for people to contact him, and the Sydney to Seoul website also allows for comments to be posted, as part of a 'leave a reply' option. Choi also invites people to follow her. This can generate discussion, questions, debates and potentially forge friendships, all based on a shared interest. Beyond this, we should also consider that the level of interest in Korea often goes far beyond Koreans' use of English on many online dictionaries. Indeed, such online dictionaries often comprise more than just a list of vocabulary. Leon's Planet, for example, offers a range of topics which people can access with the click of their mouse, such as the Korean language, Korean food, Korean history and the origins of Korean itself. As an educator, Priz also makes clear on his site that a central purpose of his website is indeed to educate people, given his background as a teacher, and Choi is also a teacher. While this is only two individuals,

it might suggest that some teachers are indeed taking lexicographical matters into their own hands, by creating their own dictionaries for the purposes of sharing information and ultimately based on wishing to inform and educate. This alone is another important reason why we need to consider such dictionaries as indeed having a degree of authority, precisely because they are sometimes created and compiled by educators, who would have an incentive to ensure that their dictionary's content is accurate. Nonetheless, on his website, Priz acknowledges that despite his interest in various languages and the content he has created for such (beyond Korean and Korean English), mistakes might be found. He offers two specific 'apologies', which arguably help to promote a sense of professionalism, and not incompetence:

Apology 1: Since I am not a native Korean, Chinese, Mongolian, or Spanish speaker, there may be linguistic errors. There might even be some English typographical errors. I apologize for those.

Apology 2: I have a plethora of links on my collective pages. Sometimes links get "broken", because other websites delete their pages. I would appreciate it if you would report broken links to me: contact me.

The Sydney to Seoul website also covers topics such as Korean culture and food, and also provides an extensive resources list pertaining largely to Korean (and even Polish!) language learning. A brief look at the comments posted on the resources section of this website demonstrates the positive features that can derive from online dictionaries in this regard. For example, there are queries from individuals looking for links to past TOPIK exam papers (Test of Proficiency in Korean), with the queries addressed. Also, an individual shares his own online dictionary, and others even post comments which ask for links to be updated, as they can no longer be accessed. But overall, there is a real sense of appreciation for Korea, to include its use of English. While I discuss the contents of only three online dictionaries out of a total of nineteen whose content I analysed, there are literally thousands more to consider. In fact, at the time of conducting my online search for online dictionaries devoted to Korean English, a search term for 'Korean English dictionary' resulted in 'about 396,000,000 results'. I do not have the time to analyse each and every online Korean English dictionary, and I can't ignore the possibility that quality and content will differ from one dictionary to the next. Indeed, some of the sites that came up in my search provided very little in terms of content. Using the additional search term of 'Konglish dictionary', one site, www.yourdictionary.com/Konglish, provided just a definition: 'A disparaging term for various varieties of Korean English having distinctive lexis, syntax and phonology'. In this instance, no examples of vocabulary were actually provided. Nonetheless, when considering the more specific examples demonstrated here regarding online dictionaries per se, it is clear that there is often more on offer beyond a dictionary: links to additional online resources on varied, but related, subjects;

a chance to engage with the public through their online posts; and many additional topics on offer which link with Korea. If we consider just one example of such an online Korean English dictionary, there is a wealth of information which could be used for an entire class. Students are offered a chance to learn about Korea from multiple perspectives, helping to reflect a focus on cultural knowledge, and they can learn this within what might be a more immediately accessible context — the Internet.

Moreover, prospective teachers can be exposed to further linguistic categories regarding lexis, which apply to all languages of course; this can further be deployed for their future (or current) students. That some of the online dictionaries provide such information is an added benefit, referring to categories such as clipping and semantic shift. An interesting feature of Choi's dictionary is the inclusion of the word 'real' in the context of 'the real meaning in Korea'. This might not appear to be of particular importance, but it arguably suggests that in this context, word meanings in Korea take precedence — as they should — regardless of the origins of a given word. Hence, the word *white* in Korean English is as meaningful as are its counterparts in countries such as the USA and UK, which would respectively use words such as *whiteout* and *Tipp-ex*. Here, then, are examples of three Englishes for the price of one, and initiated with the inclusion of an online dictionary in the classroom. Likewise, the word *saida* (cider) refers to a drink in Korea similar to 7-UP, but in the UK cider is an alcoholic beverage made from apples, whereas in the US it is non-alcoholic.

Going further, the inclusion of online dictionaries within the classroom can be the dominant focus per se, and not a supplement. Students can be asked to create their own online dictionaries, reflecting the ways in which technology is being used in the EFL classroom by the students and not just the teacher. Lee (2019), for example, discusses the use of IDLE (Informal Digital Learning of English), in which students can be asked to use their skills with social media and the virtual community, creating websites or blogs for assessment. This serves a twofold purpose: first, to tap into students' technological abilities and allow them to use such abilities and skills for a practical purpose, which can help facilitate learning; and second, to give students an opportunity to apply their own knowledge by creating their own online dictionary, thus being required to consider linguistic difference as just that, difference (and not deficit). Lee goes on to explain that "if Korean students practice IDLE activities to better understand diverse cultures, they are likely to improve their perceived ability to employ cross-cultural communication strategies" (page 294). This points to a suggested additional benefit, which derives from respecting the speakers within a culture who have made English their own, by first being exposed to the ways in which they have adapted English for national, and international, usage. Moreover, for EFL teachers who are planning to teach overseas, then learning that country's variety of English would certainly make life easier, facilitate communication and probably be easier to learn than a new language per se. For example, it is easier to

learn Korean English as an American than to learn Korean, though learning Korean if residing in Korea is of course entirely worthwhile.

From this brief sample of potential classroom ideas, it is hoped that the practical inclusion of online dictionaries, as well as NICE housed within such dictionaries, serves an entirely useful application for teaching and learning. But the larger framework for the use of such is tied to a need to go beyond rather singular notions of 'correct' English and 'authoritative' dictionaries. Instead, the inclusion of NICE helps to reflect the linguistic reality outside the classroom, foster empathy within future EFL teachers regarding their students' English varieties (Matsuda 2002, 2020; Renandya 2012; Dinh 2017; Joo, Chik and Djonov 2020), and likewise help EFL students accept such linguistic diversity in the process. Furthermore, the use of online dictionaries as part of pedagogy helps students to regard the online dictionary as wholly legitimate regarding its usage in the classroom, a current, and perhaps future, means to capture the vocabulary of NICE, expanding circle Englishes in particular. I leave it to the readers of this article to consider additional ways in which online dictionaries can be used as part of pedagogy, though such usage need not be tied to EFL of course, and can involve a multitude of languages, topics and disciplines. For example, online dictionaries can also have a part to play if discussing fandom vocabulary, as was discussed, as part of a media course. Online dictionaries can also be used in the classroom to teach diplomats a given language variety before moving overseas, such as British diplomats who are learning English vocabulary as used in India (Wrenn 2012). And online dictionaries can be used as the main focal point per se in classes focused on the history of lexicography, in which case their inclusion would allow once again for a more concrete and non-judgemental approach, which is entirely necessary in the context of higher education.

In closing, I again invite readers to consider their own ideas and innovations regarding the use of online dictionaries in the classroom, regardless of the language used (be it Afrikaans or Burushaski), as well as the classroom focus for which the dictionary has a natural and meaningful inclusion. But certainly within the current societal use of technology, online dictionaries are here to stay, and their use within classrooms reflecting all manner of subjects should be considered, a means to determine how we can "inject new blood into the lexicographical veins" (Tarp 2019: 226). This need not relate solely to the ways in which we approach the development of lexicography from a more modern technological perspective, but also how we can use online dictionaries in their current form within our various pedagogies. But when online dictionaries are deployed within language classes, here EFL, there is the added benefit of helping students recognise the sheer variety of the English language, and its validity beyond a singular pedagogic model; and when presented within the context of online dictionaries, then a modern approach to lexicography can also be harnessed as a powerful teaching and learning tool.

References

- Bamgbose, A.** 1998. Torn between the Norms: Innovations in World Englishes. *World Englishes* 17(1): 1-14.
- Baratta, A.** 2019. *World Englishes in English Language Teaching*. London: Palgrave Macmillan.
- Baratta, A.** 2021. *The Societal Codification of Korean English*. London: Bloomsbury.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Berkenkotter, C. and T.N. Huckin.** 1995. *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. Hillsdale, NJ: Lawrence Erlbaum.
- Brown, K.** 1993. World Englishes in TESOL Programs: An Infusion Model of Curricular Innovation. *World Englishes* 12(1): 59-73.
- Campoy Cubillo, M.C.** 2004. Computer-mediated Lexicography: An Insight into Online Dictionaries. Campoy Cubillo, M.C. and P.S. Jordà (Eds.). 2004. *Computer-Mediated Lexicography in the Foreign Language Learning Context*: 47-72. Castelló de la Plana: Universitat Jaume.
- Choi, M.** 2017. *English Words You Only Hear in Korea*. Available at: <https://medium.com/story-of-egg-bun-education/english-words-you-only-hear-in-korea-5bd3eed8f9c3>.
- Cotter, C. and J. Damaso.** 2007. Online Dictionaries as Emergent Archives of Contemporary Usage and Collaborative Codification. *Queen Mary's OPAL #9 Occasional Papers Advancing Linguistics*: 1-10.
- Damaso, J.** 2005. *The New Populist Dictionary: A Computer-Mediated, Ethnographic Case Study of an Online, Collaboratively Authored English Slang Dictionary*. M.A. dissertation. London: Queen Mary University of London.
- Dinh, T.N.** 2017. Preparing Preservice Teachers with EIL/WE-oriented Materials Development. Matsuda, A. (Ed.). 2017. *Preparing Teachers to Teach English as an International Language*: 131-146. Bristol: Multilingual Matters.
- Dolezal, F.** 2006. World Englishes and Lexicography. Kachru, B.B., Y. Kachru and C.L. Nelson (Eds.). 2006. *The Handbook of World Englishes*: 694-708. Chichester: Blackwell.
- Friedrich, P.** 2002. Teaching World Englishes in Two South American Countries. *World Englishes* 21(3): 441-444.
- Galloway, N.** 2017. *Global Englishes and Change in English Language Teaching*. London: Routledge.
- Galloway, N. and H. Rose.** 2015. *Introducing Global Englishes*. London: Routledge.
- Green, J.** 1996. *Chasing the Sun: Dictionary-Makers and the Dictionaries they Made*. London: Jonathan Cape.
- Green, J.** 2012. Dictionaries are not Democratic. *The Guardian*, 13 September 2012.
- Gut, U.** 2011. Studying Structural Innovation in New English Varieties. Mukherjee, J. and M. Hundt. 2011. *Exploring Second-Language Varieties of English and Learner Englishes*: 101-124. Amsterdam: John Benjamins.
- Jenkins, J.** 2009. *World Englishes: A Resource Book for Students*. London: Routledge.
- Joo, S.J., A. Chik and E. Djonov.** 2020. The Construal of English as a Global Language in Korean EFL Textbooks for Primary School Children. *Asian Englishes* 22(1): 68-84.
- Kachru, B.B.** 1991. Liberation Linguistics and the Quirk Concern. *English Today* 7(1): 3-13.
- Kermas, S.** 2012. Culture-specific Lexis and Knowledge Sharing in the Global Village. Facchinetti, R. (Ed.). 2012. *English Dictionaries as Cultural Mines*: 73-94. Newcastle-upon-Tyne: Cambridge Scholars.

- Kruger, H. and B. van Rooy.** 2017. Editorial Practice of the Progressive in Black South African English. *World Englishes* 36(1): 20-41.
- Lee, J.S.** 2019. An Emerging Path to English in Korea: Informal Digital Learning of English. Dressman, M. and R.W. Sadler (Eds.). 2019. *The Handbook of Informal Language Learning*: 289-302. Chichester: Wiley-Blackwell.
- Leimgruber, J.R.E., P. Siemund and L. Terassa.** 2018. Singaporean Students' Language Repertoires and Attitudes Revisited. *World Englishes* 37(2): 282-306.
- Lovtsevich, G.N. and A.A. Sokolov.** 2020. World Englishes and Learner Lexicography: View from the Expanding Circle. *Russian Journal of Linguistics* 24(3): 703-721.
- Luzón Marco, M.J.** 2004. Dictionary Websites: Transforming and Adding Value to the Dictionary Genre. Campoy Cubillo, M.C. and P.S. Jordà (Eds.). 2004. *Computer-Mediated Lexicography in the Foreign Language Learning Context*: 73-94. Castelló de la Plana: Universitat Jaume.
- Matsuda, A.** 2002. "International Understanding" through Teaching World Englishes. *World Englishes* 21(3): 436-440.
- Matsuda, A.** 2020. World Englishes and Pedagogy. Nelson, C.L., Z.G. Proshina and D.R. Davis (Eds.). 2020. *The Handbook of World Englishes*: 686-702. Second edition. Chichester: John Wiley.
- McKay, S.** 2002. *Teaching English as an International Language*. Oxford: Oxford University Press.
- Meriläinen, L.** 2017. The Progressive Form in Learner Englishes: Examining Variation across Corpora. *World Englishes* 36(4): 760-783.
- Mufwene, S.S.** 2001. *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Nesi, H.** 2000. Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: The State of the Art. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2002. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000. Vol. 2*: 839-847. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Pikilnyak, A.V., N.M. Stetsenko, V.P. Stetsenko, T.V. Bondarenko and H.V. Tkachuk.** 2020. Comparative Analysis of Online Dictionaries in the Context of the Digital Transformation of Education. Semerikov, S.O. and M.P. Shyshkina (Eds.). 2020. *Proceedings of the 8th Workshop on Cloud Technologies in Education (CTE 2020), 18 December 2020, Kryvyi Rih, Ukraine*: 70-81. Academy of Cognitive and Natural Sciences (ACNS).
- Potayroi.** 2014. *Fandom as a Discourse Community*. Available at: <https://thansong123.wordpress.com/2014/04/24/fandom-as-a-discourse-community/>.
- Priz, L.** 2011. *Konglish*. Available at: <http://leonsplanet.com/konglish.htm>.
- Renandya, W.A.** 2012. Teacher Roles in EIL. *The European Journal of Applied Linguistics and TEFL* 1(2): 65-80.
- Rose, H.** 2017. A Global Approach to English Language Teaching: Integrating an International Perspective into a Teaching Methods Course. Matsuda, A. (Ed.). 2017. *Preparing Teachers to Teach English as an International Language*: 169-180. Bristol: Multilingual Matters.
- Rüdiger, S.** 2019. *Morpho-Syntactic Patterns in Spoken Korean English*. Amsterdam/Philadelphia: John Benjamins.
- Schwab, K.** 2015. The Fourth Industrial Revolution: What It Means and How to Respond. *Foreign Affairs*, 12th December 2015. Available at: www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution.
- Tarp, S.** 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249.

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/32-1-1755> (Volle uitgawe / Full volume)
<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/32-1-1726> (Article)

- Touhami, B. and F. Al-Abed Al-Haq.** 2017. The Influence of the Korean Wave on the Language of International Fans: Case Study of Algerian Fans. *Sino-US English Teaching* 14(10): 598-626.
- Trudgill, P. and J. Hannah.** 1994. *International English: A Guide to Varieties of Standard English*. London: Edward Arnold.
- Widdowson, H.G.** 1994. The Ownership of English. *TESOL Quarterly* 28(2): 377-389.
- Wrenn, E.** 2012. British Diplomats to India Must Learn 'Hinglish' as Mix of Hindi and English Becomes the Most Used Language. *Daily Mail*, 11 October 2012.
- Z.** 2020. *SydneytoSeoul: A Self-Study Korean Language Blog*. Available at: <https://sydneytoseoul.wordpress.com/2013/08/02/lexical-borrowings-from-english-in-konglish/>.

The Influence of Grammatical Gender on the Sequence of Near-synonyms in Serbian Dictionaries in Contrast to English Thesauri

Dragana Čarapić, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro (draganac@ucg.ac.me)*

Abstract: It has been noticed that Serbian dictionaries of synonyms do not provide a clear insight into the lexical relations among the words bearing similar meaning (Prčić 2010). Therefore, we have devised the multi-faceted approach to the research which incorporates the collocational, componential, and contrastive analysis of the descriptive adjectives *hrabar*, *-a*, *-o*, in Serbian, and its English counterpart *brave*. Specific emphasis has been given to the presence of grammatical gender (masculine, feminine, neuter) in Serbian in contrast to natural gender in English. Electronic databases and electronic corpora are used for the analysis of the most frequent collocational framework of the chosen samples. The results of the research employed illustrate how semantic (and grammatical) aspects of words are reflected onto and within their collocational range indicating idiomatic meaning of the adjectives analysed in both languages. Furthermore, the analysis applied has shown that the most frequent collocations of the adjective *hrabar*, *-a*, *-o*, have varied depending on the grammatical gender implied (masculine, feminine, neuter), as well as the sequence of its near-synonyms. The same changes have not occurred in English due to its lack of grammatical gender.

Keywords: NEAR-SYNONYMS, DESCRIPTIVE ADJECTIVES, COLLOCATIONAL ANALYSIS, COMPONENTIAL ANALYSIS, CONTRASTIVE ANALYSIS

Opsomming: Die invloed van grammatikale gender op die opeenvolging van ampersinonieme in Serwiese woordeboeke in teenstelling met Engelse tesourusse. Daar is opgemerk dat Serwiese sinoniemwoordeboeke nie 'n duidelike insig verskaf in die leksikale verwantskappe tussen die woorde wat soortgelyke betekenis dra nie (Prčić 2010). Daarom het ons die veelvlakkige benadering tot die navorsing gevolg wat die kollokasionele, komponensiele, en kontrastiewe analise van die deskriptiewe adjektiewe *hrabar*, *-a*, *-o*, in Serwies, en die Engelse ekwivalent daarvan, *brave*, inkorporeer. Spesifieke klem is geplaas op die aanwesigheid van grammatikale gender (manlik, vroulik, onsydig) in Serwies in teenstelling met natuurlike gender in Engels. Elektroniese databasisse en elektroniese korpora word vir die analise van die mees frekwente kollokasionele raamwerk van die gekose voorbeelde gebruik. Die resultate van die navorsing wat gedoen is, illustreer hoe semantiese (en grammatikale) aspekte van woorde op en binne hul kollokasionele reeks gereflekteer word wat dui op die idiomatiese betekenis van die adjektiewe wat in albei tale geanaliseer is. Verder het die analise wat uitgevoer is, getoon dat die

mees frekwente kollokasies van die adjektief *hrabar*, *-a*, *-o*, weens die geïmpliseerde grammatikale gender (manlik, vroulik, onsydig), asook die opeenvolging van die ampersinonieme, gevarieer het. As gevolg van die afwesigheid van grammatikale gender het soortgelyke veranderings nie in Engels voorgekom nie.

Sleutelwoorde: AMPERSINONIEME, DESKRIPTIEWE ADJEKTIEWE, KOLLOKASIONELE ANALISE, KOMPONENSIËLE ANALISE, KONTRASTIEWE ANALISE

1. Introduction

The aim of this paper is to provide a multi-faceted method to analysing the proximity of meaning among near-synonyms in Serbian and English dictionaries. Presently, in Serbian and English dictionaries synonyms are listed either randomly or alphabetically without pointing out proximity of meaning among near-synonyms. Moreover, only adjectives in masculine gender are stated in Serbian dictionaries of synonyms and there is no insight into the specific semantic features of adjectives in female and neuter gender.¹ Therefore, we have applied a specific approach to deal with the lack of nuances in meaning among synonyms in two Serbian dictionaries, namely *Rečnik sinonima i tezaurus srpskog jezika* compiled by Pavle Ćosić and *Sinonimi i srodne reči srpskohrvatskoga jezika* compiled by Miodrag Lalević. The semantic analysis employed in this paper comprises the most frequent collocational range for the selected descriptive adjectives *hrabar*, *-a*, *o* in Serbian and *brave* in English and samples of their near-synonyms were proposed. The results of the research confirm previous findings that the semantic (and grammatical) aspects of a word are reflected in their most frequent collocational range (Harris (1954), Stubbs (2001), Sinclair (2004), Gries and Stefanowitsch (2004), Hoey (2005), Xiao and McEnery (2006), Turney and Pantel (2010), Liu (2020)).

The first dictionaries of synonyms emerged at the outset of the 17th century (Petrović 2005: 65). Those were the dictionaries of Latin synonyms. In the 18th century dictionaries of French, German, English and Spanish synonyms were published, and in the 19th century Nikola Tomasseo compiled the first Italian dictionary of synonyms which can still be found in bookshops today. In this century the first Danish, Dutch and Swedish dictionaries of synonyms appeared. Petrović (2005: 68) suggests that there have been no crucial advances in the area of the lexicography on synonymy. The first dictionary of synonyms of the Serbian language *Sinonimi i srodne reči srpskohrvatskog jezika* was compiled by Lalević in 1974. The photo type edition of this dictionary was published in 2004 (see Dragičević 2010). Petrović (2005: 73-83) acknowledges a difference between differential and cumulative dictionaries. Differential dictionaries of synonyms are characterized by stating the nuances in meaning among the synonyms. In this type of dictionary, an orderly cluster of synonyms is placed next to the node word which in certain cases may be set out in alphabetical order and in some cases is set out based on their proximity of meaning. On the other hand, a

cumulative dictionary contains its node word, the orderly cluster of synonyms that it belongs to as well as examples of its use. The differences in meaning among synonyms are not explained (Dragičević 2010: 261).

The lack of clear insight into the nuances in meaning among the near-synonyms renders impositions in other linguistic areas such as translation, especially when we need to decide whether a certain lexeme can be translated by a greater number of its synonymous equivalents. Bearing in mind all these issues, we can realize why certain European languages, such as Croatian, saw its first dictionary of synonyms published as late as 2008. A similar situation occurred in the Serbian language, with the latest dictionary of synonyms published in the same year, 2008, as a new supplemented edition when compared to the one compiled by M. Lalević in 1974. Further to the aforementioned data, the following quotes account for the delay in the compilation of the dictionary of Croatian synonyms:

One of the possible reasons of the delayed publishing of the Thesaurus could stem out of lack of agreement among the Croatian linguists what could be considered under the very notion of synonymy. There are a few theoretical definitions of synonyms in research papers of the linguists, what is the least expected occurrence due to the complexity of this issue (Opašić 2010: 165).

The lexicographic relevance of our research study lies in proposing a specific method of presenting near-synonymous samples of descriptive adjectives in the future Serbian and English dictionaries of synonyms. The method was devised in an attempt to delineate fine boundaries in meaning among near-synonyms of the node word. A detailed account of the method applied and its results will be given in the subsequent sections of the paper.

2. Theoretical background

Lexicographical compilations of synonyms in the 18th century contained explanations related to the meaning of words and their nuances in meaning. Nevertheless, Miller and Charles (1991) noticed that the definitions of those words were so close in meaning that the words were frequently mixed and interchanged, e.g., *austerity*, *severity* and *rigour*. However, in these examples the existence of any absolute synonyms cannot be noticed, but relations among them would be rather defined as: "similarity of meaning", "semantic differential" and "semantic similarity". Introducing such expressions enhanced synonymy analysis as a constant variable, within which identity of meaning is defined through grading of meaning (Miller and Charles 1991).

Following the same approach to lexical semantics' analysis, Edmonds (Edmonds and Hirst 2002) concludes that synonymy relies on grading as the main issue underlining definitions of words and their slight differences in meaning, while investigating machine translation and having dealt with the near-synonymy analysis based on computer and corpus linguistics, e.g., if *fir*

and *pine* can be defined as words denoting *coniferous trees*, in such case they can be considered absolute synonyms. Constant recognition of near-synonyms depends on the principle of meaning level, above which all words listed, are considered not to be synonyms, and on the opposite side of the same level near-synonyms can be found. Edmonds (Edmonds and Hirst 2002) suggests a model of lexical meaning, which consists of three levels:

- the ultimate level which includes a specific concept, converging with clusters of near-synonyms which, at the same time, belong to a different level. The concept contains essential content which is shared among all near-synonyms;
- at the second level near-synonyms differ in their semantic, stylistic and expressive features;
- whereas at the third level syntactic characteristics can be recognized.

Edmonds (Edmonds and Hirst 2002) shares the same opinion as Hirst (Edmonds and Hirst 2002) who states that the ultimate level can operationally be identified as the level highly dependent on a language: a concept which is lexicalized within different languages must be an ultimate concept. However, near synonymy is an increasingly changeable category in all languages. For example, an ultimate concept, known as a *generic mistake* is connected with an inconsistent cluster of near-synonyms in different languages: English *error, blunder, slip, mistake, lapse, howler* and so on; French *faute, erreur, faux, pas, bévue, bêtise, bavure, imapir*, and so on.; and German *Irrtum, Fehler, Mißgriff, Versehen, Schützer*, and so on.

Therefore, synonymy can be construed as the highest level of mutual substitution, without any crucial changes in the meaning of a certain statement within a context (Miller and Charles 1991; Miller 1999), by which the phenomenon of synonymy can be characterised as near-synonymy (or plesynonyms), which can be described as "contextually defined synonyms" (Miller 1999: 24).

By the application of lexical choices among semantically similar words, especially near-synonyms, in a theoretical approach which is based on computer linguistics, Edmonds and Hirst (2002) suggested at least three levels of meaning complexity, as follows:

- (1) The Conceptual-semantic level;
- (2) The Subconceptual/stylistic-semantic level;
- (3) The Syntactic-semantic level.

All these levels are in accordance with far more detailed accounts of meaning, known as the granularity of word meaning (Edmonds and Hirst 2002: 117-124).

Divjak and Gries (2006) show how the internal structure of a group of near-synonyms can be revealed. Second, they deal with the problem of distinguishing the sub-clusters and the words in those sub-clusters from each other. Finally, they illustrate how these results identify the semantic properties that should be mentioned in lexicographic entries. They illustrate their methodol-

ogy with a case study on nine near-synonymous Russian verbs that, in combination with an infinitive, express *try*. Their approach is corpus-linguistic and quantitative. Assuming a strong correlation between semantic and distributional properties, Divjak and Gries (2006) analyse 1,585 occurrences of these verbs taken from the Amsterdam Corpus and the Russian National Corpus, supplemented where necessary with data from the Web. They code each particular instance in terms of 87 variables (ID tags), i.e., morphosyntactic, syntactic and semantic characteristics that form a verb's behavioural profile (BP). Liu (2010, 2013) and Liu and Espino (2012) have conducted BP (behavioural profile) studies on English synonymous adjectives, nouns, and adverbs by focusing on key co-occurrences instead of all the contextual distribution patterns (Liu 2010: 80).

Inkpen and Hirst (2006) and Phoocharoensil (2020) support the idea that collocational patterns within structural variation can be an indicator for distinguishing near-synonyms. A collocation is defined as a co-occurrence of certain words (O'Dell and McCarthy 2008; Lewis 2000). In terms of collocational preference, near synonymous words may co-occur with different collocations from each other according to collocational restrictions (Palmer 1981). Edmonds and Hirst (2002) provide examples of the differences in collocational patterns of the synonym *task* and *job* in which one can *face a daunting task* but not *a daunting job*. This concept of collocational patterns is related to other criterion used to differentiate near-synonyms, namely semantic preference. While collocational patterns concern lexical restriction, the word's occurrence is defined within the semantic environment (Flowerdew 2012). For example, Edmonds and Hirst (2002) mention that the synonymous verbs *die* and *pass away* have a different semantic preference in which *pass away* is only used with people not with animals or plants.

There are a number of studies on synonym differentiation investigating their semantic preference based on data drawn from language corpora, e.g., the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA). In these studies near-synonyms have been differentiated by examining their distribution across genre, collocations, and semantic preference within different parts of speech:

- nouns (Jirananthiporn (2018), Phoocharoensil (2020), Jarunwaraphan and Mallikamas (2020), Sumonsriworakun (2022));
- verbs (Gu (2017), Song (2021), Phoocharoensil (2021), Kruawong and Phoocharoensil (2022))
- adverbs (Desagulier (2014) Yang (2020), Stoyanova-Georgieva (2020) and Tao (2021));
- adjectives (Cai (2012), Crawford and Csomay (2016), Petcharat and Phoocharoensil (2017), Kamiński (2017), Sayyed and Al-Khanji (2019), Selmistraitis (2020) and Thongpan (2022).

In the collocational analysis of selected adjectives Cai (2012) examined the synonyms *awesome*, *excellent*, *fabulous*, *fantastic*, *great*, *terrific*, and *wonderful*

based on the data drawn from COCA. The results of the study revealed that the most frequently occurring word is *great* which was also reported to be more general. Moreover, regarding the frequency distribution across genres, it was found that *fabulous*, *fantastic*, *great*, *terrific*, and *wonderful* occur more in the spoken genre, while *awesome* and *excellent* are prevailing in magazines. Collocation was also the major criterion used in distinguishing two adjective synonyms *equal* and *identical* in a study by Crawford and Csomay (2016). Generally, it was shown that both synonyms are interchangeable in certain contexts; however, in an analysis of collocations derived from COCA, different collocational behaviour was revealed in that there is more likelihood of abstract concepts such as *opportunities*, *rights*, and *protection* accompanying *equal*, and concrete nouns such as *twins*, *copies*, and *houses* co-occurring with *identical*. Such an observation is only possible by referencing large amounts of texts through corpus data, as it allows linguists and language researchers to access language use in authentic contexts.

Petcharat and Phoocharoensil (2017) examined three synonymous adjectives *appropriate*, *proper* and *suitable* based on their meaning, collocations, formality and grammatical patterns. The data were collected from three dictionaries, namely the *Longman Dictionary of Contemporary English* 6th edition (2014), *Longman Advanced American Dictionary* 3rd edition (2013), and *Macmillan Collocations Dictionary* (2010) and COCA. It was found that even though *appropriate*, *proper* and *suitable* shared the same core meaning, they differed in detailed meanings in which *suitable* tends to be used more with person. In addition, *appropriate* occurred more in formal context than *suitable* and *proper*. Moreover, by examining collocations and grammatical patterns, it was found that they shared only a small number of collocations and grammatical patterns in which they occur. This indicated that the three synonymous adjectives are not absolute synonyms as they cannot be used interchangeably in the same context. Kamiński (2017) explores the potential usefulness of two techniques that visualise collocational preference for the purpose of synonym discrimination. Given the fact that collocation is one of the most important markers of meaning difference, it is used as the criterion for distinguishing between near-synonyms. Collocational preferences for a set of near-synonyms (*artificial*, *fake*, *false*, and *synthetic*) were visualised using two techniques: correspondence analysis plot and collocational network. The collocations were retrieved from BNC corpus by using a distributional method. An advantage of the graphs is that they allow lexicographers to spot similarities and differences in collocational preference of several words in a single diagram. Sayyed and Al-Khanji (2019) investigated the similarities and differences that exist between *afraid*, *scared*, *frightened*, *terrified*, *startled*, *fearful*, *horrified* and *petrified*. Specifically, they compared and contrasted the words in terms of dialectal differences, frequency of occurrence, distribution in different genres and core meanings. The findings have shown that nearly all adjectives appear to be mostly used in fiction and spoken genres. Furthermore, the results also unveiled that both the Americans and the British tend to avoid

using such adjectives of fear in academic contexts. In a more recent study, Selmistraitis (2020) examined semantic preference among three pairs of synonymous adjectives, namely *succinct* and *concise*, *coherent* and *cohesive*, and *precise* and *accurate* in the academic texts of COCA and found that *succinct* and *concise* are more similar in semantic preference than other two pairs of synonyms.

Thongpan (2022) analyses the similarities and differences between the three synonymous adjectives *far*, *distant*, and *remote*, concentrating on the degree of formality in their distribution across eight genres, as well as their collocations with the semantic preference combination. The data for this study was drawn from COCA. The results revealed that the word *remote* has the most formal degree, followed by *distant* and *far*, respectively, since it is most commonly used in academic texts. In terms of collocations and semantic preference, the data pointed out that the three synonyms share only a few collocates and, as a result, they vary in semantic preference.

Investigating possibilities of determining proximity in meaning among near-synonyms, various scholars have proposed specific methods of quantifying near-synonymy relations among selected words. Cappelli (2011) presented a cluster analysis approach to quantify near-synonymy relations and compare non-parametric and parametric methods. The first approach is model free since it does not assume an underlying model of lexical knowledge but it uncovers the group structure in the set of near synonyms of a target word by comparing the list of synonyms of the given entry with those of its near synonyms as contained in available thesauri. Then, in order to validate the results provided by the cluster analysis, a statistical model is introduced for analyzing human judgments of perceived degree of synonymy, also by a relationship with the subjects' characteristics. Piits (2013) carried out a comparative study on words for *human beings* and their Estonian collocates. The study was inspired by the distributional hypothesis by Zellig Harris (1957), which states that words occurring in similar contexts tend to have a similar meaning. The word *collocate* is used in a neo-Firthian sense, covering all the words that co-occur with the node word the most often. The comparison involved the 30 most frequent collocates for each node word. Assuming that a bigger number of shared collocates means a greater semantic closeness, intersections of collocates of the Estonian words for 'human being' were computed. It turned out that antonymous words had the highest number of collocates in common, which indicates that syntagmatic relations of words may also reflect some of their paradigmatic relations. As antonymy is interpreted as a similarity relation (which is quite natural considering that the opposite value applies to just one semantic component out of several), Piits (2013: 157) agreed with Harris (1957) that similarity between words is manifested in their contextual coincidence. Fachrutdinov, Khisamova and Khisamatullina (2017) conducted a comparative analysis of semantic distinctions between synonymous adjectives in Tatar and English. Herda (2020) contrasted the near-synonymous Polish classifiers *kupa* 'heap', *sterta* 'pile', and *stos*

'stack', all of which encode upward-oriented arrangements of objects or substances and thus prototypically combine with concrete inanimate nouns, by means of a collocational analysis conducted on naturally-occurring data derived from the National Corpus of Polish. As part of a larger research project on expert assessment of synonymic rows in RuWordNet thesaurus, Gimaletdinova, Khalitova, Solovyev and Bochkarev (2021) pointed out that the problem of determining semantic similarity between words affects the understanding of synonymy and creates obstacles to the work of lexicographers.

In the present study we have analysed descriptive adjectives *hrabar*, *-a*, *-o* in Serbian and *brave* in English within their most frequent collocational framework using different corpora: Corpus of the Contemporary Serbian language, Faculty of Mathematics and Natural Sciences, University of Belgrade, SrWac (Serbian Web), the British National Corpus (BNC) and ukWac (British Web). The results of the collocational analysis of these adjectives were the basis for proposing their near-synonymy samples while investigating their proximity in meaning. We have noticed that grammatical gender influences the sequence of near-synonymy samples in Serbian, but not in English. The approach to the lexical-semantic analysis employed in this paper relies on the idea of the granularity of word meaning applying corpus analysis. This idea is relevant to the present study because it explains the nature of the complexity of near-synonymy as a lexical relation. It points to the way in which near-synonymy affects the structure of the lexicon and lexicographic representation as both fine-grained aspects of near-synonymous adjectives and their collocational patterns are significant for discerning their subtle nuances in meaning.

3. Methodology of the research

The starting hypothesis of this paper rests on the agreement among the majority of semanticists that absolute synonymy is rare, while in the opinion of certain linguists it is in fact non-existent (Quine (1951); Palmer (1981: 59)). In the opinion of philosophers such as Quine (1951) and Goodman (1952) absolute synonymy is impossible, though it may be limited to mostly technical terms (distichous, two-ranked; groundhog, woodchuck) (Hirst 1995: 51). Cruse (1986: 270) claims that "natural languages abhor absolute synonyms just as nature abhors vacuum," as the meanings of words are constantly changing.

The multi-faceted approach applied in this paper features the use of the collocational approach to lexical analysis and componential analysis (Nida (1975); Bendix (1996); Zhang (2002)) of the collocates of the extracted descriptive adjectives, and as the final step of the research, applying contrastive analysis.

The results of the analysis should prove another hypothesis, namely that the collocational framework is a better generator of the precise meaning of a word than can be inferred through its denotation. These results will indicate findings that the semantic (and grammatical) aspects of words are reflected

onto and within their own collocational framework. It is expected that the collocational framework (the most frequent collocates) of the adjective *hrabar*, *-a*, *-o* in Serbian will change depending on the grammatical gender implied (masculine, feminine, neuter), as well as the sequence of its near-synonyms. It is claimed that there are inflectional selectional differences among synonyms in a morphologically rich language such as Serbian. The same changes are not expected in the English language due to its lack of grammatical gender.

Componential analysis is carried out to show that the denotational meaning of the lexemes is not enough to illustrate the comprehensive scale of the meaning of lexemes analysed within context. In dictionaries of synonyms mentioned earlier in this paper, mainly the denotational meaning of the lexemes is presented. It is essential to point out how much denotational meaning is insufficient and how much other aspects of the meaning of the lexeme are dependent upon the collocational range.

Consequently, componential analysis has been applied to look for the semantic features of the descriptive adjectives analysed within their most frequent collocational framework. This approach appears very similar to the profile-based usage-feature analysis (Geeraerts et al. (1994), Gries (2006), Arppe (2008), Divjak (2010), Glynn (2014)).

Certain authors, like Taylor (1991), have a very negative attitude towards componential analysis, which they consider an outdated and obsolete method. In contrast, Geeraerts (2010: 4) claims that: "there can be no semantic description without some sort of decompositional analysis".

The process of contrasting (or analysing) presupposes the comparison of nominal collocates of descriptive adjectives. For the analysis of synonyms within the most frequent collocational framework, electronic dictionaries and electronic databases, such as the corpus of the Contemporary Serbian language, Faculty of Mathematics and Natural Sciences, University of Belgrade, SrWac (Serbian Web), the British National Corpus (BNC), and ukWac (British Web), are used.

4. Analysis

In the analysis undertaken we have applied a collocational method aiming at determining the proximity of meaning among the samples of near synonyms. First, we extracted the most frequent collocations of the node word *hrabar*, *-a*, *-o*. We then have selected the first 10/15 synonyms of the adjective *hrabar*, *-a*, *-o* from the *Dictionary of synonyms* by Pavle Ćosić and his assistants (2008) (see Addendum 1). In the next phase of analysis we analysed co-occurrences of the extracted samples of synonyms within the most frequent collocational framework of the adjective *hrabar* while using the SrWac Corpus (554.627.647 tokens) (see Addendum 2). The co-occurrences and concordance of the near synonymy samples within the most frequent collocational framework of the node word implied the proximity of meaning of near-synonyms to the node word.

4.1 The analysis of the descriptive adjectives *hrabar*, *-a*, *-o* and their near-synonyms in Serbian

After analyzing the first ten synonyms² (extracted from the dictionary of synonyms: *Rečnik sinonima*, Pavle Ćosić et al. (2008: 293)) through their co-occurrences within the most frequent nominal collocations of the node word *hrabar* *~čovek* (eng. *man*), *~potez* (eng. *move*), *~korak* (eng. *step*), *~čin* (eng. *act*)³ the following four synonyms can be considered to be near-synonyms of the descriptive adjective *hrabar*, *-a*, *-o* (eng. *brave*): *herojski* (eng. *heroic*), *odvažan* (eng. *bold*), *junački* (eng. *stout*), *smeo* (eng. *courageous*), (*Rečnik sinonima*, Pavle Ćosić i saradnici 2008: 659).

Table 1: Near-synonymy samples of the adjective *hrabar*

The most frequent four collocates of the descriptive adjective <i>hrabar</i> analysed in corpus data of the Contemporary Serbian language of the Faculty of Mathematics in Belgrade					
	<i>Čovek</i>	<i>Potez</i>	<i>Korak</i>	<i>Čin</i>	
Descriptive adjective	Concordance number of the collocates analysed on the srWaC corpus				Total number of concordances for the given examples
<i>Hrabar</i>	567	224	88	43	922
<i>(tokens per million words)</i>	<i>(1.02)</i>	<i>(0.40)</i>	<i>(0.16)</i>	<i>(0.08)</i>	<i>(1.66)</i>
Near-synonymy samples of the adjective <i>hrabar</i>					
<i>Herojski</i>	6	/	23	108	137
<i>Odvažan</i>	6	43	11	2	62
<i>Smeo</i>	15	38	32	3	78
<i>Junački</i>	2	7	26	9	44

In the second stage of the research, frequency analysis of the four most recurrent collocates of the adjective *hrabar* (*~potez* (eng. *move*), *~čovek* (eng. *man*), *~korak* (eng. *step*), *~čin* (eng. *act*)) was carried out within the collocational framework of the suggested near synonyms (*herojski*, *odvažan*, *smeo*, *junački*) (see Table 1).

Having examined the results, we can conclude that the near synonyms of the adjective *hrabar* are: *herojski* (137) *smeo* (78), *odvažan* (62), *junački* (44) (srWaC corpus, taken on July 28, 2022).

Table 2: Near-synonymy samples of the adjective *hrabra*

The most frequent four collocates of the descriptive adjective <i>hrabra</i> analysed on corpus data of the Contemporary Serbian language of the Faculty of Mathematics in Belgrade					
	<i>Žena</i>	<i>Odluka</i>	<i>Devojka</i>	<i>Igra</i>	
Descriptive adjective	Concordance number of the collocates analysed on the srWaC corpus				Total number of concordances for the given examples
<i>Hrabra</i> (tokens per million words)	107 (0.19)	39 (0.07)	22 (0.04)	18 (0.03)	186 (0.33)
Near-synonymy samples of the adjective <i>hrabra</i>					
<i>Herojska</i>	/	13	/	3	17
<i>Odvažna</i>	10	1	1	2	14
<i>Smela</i>	12	4	2	1	20
<i>Junačka</i>	1	/	2	1	4

The third stage of the research includes frequency analysis of the four most recurrent collocates of the adjective *hrabra* (~*žena* (eng. *woman*), ~*odluka* (eng. *decision*), ~*devojka* (eng. *girl*), ~*igra* (eng. *play*)) within the collocational framework of the suggested near synonyms (*herojska*, *odvažna*, *smela*, *junačka*) (see Table 2). The results of the analysis suggest that the near synonyms of the adjective *hrabra* are: *smela* (20), *herojska* (17), *odvažna* (14), *junačka* (4) (srWaC corpus, taken on July 28, 2022) (see Table 2).

The final stage of the research involves frequency analysis of the four most recurrent collocates of the adjective *hrabro* (~*srce* (eng. *heart*), ~*suočavanje* (eng. *confrontation*), ~*novinarstvo* (eng. *journalism*), *svedočanstvo* (eng. *testimony*)) within the collocational framework of the suggested near-synonyms (*herojsko*, *odvažno*, *smelo*, *junačko*) (see Table 3). The results of the analysis suggest that the near synonyms of the adjective *hrabro* are: *herojsko* (20), *junačko* (18), *odvažno* (7), *smelo* (5) (srWaC corpus taken on July 28, 2022).

Table 3: Near-synonymy samples of the adjective *hrabro*

The most frequent four collocates of the descriptive adjective <i>hrabar</i> analysed in corpus data of the Contemporary Serbian language of the Faculty of Mathematics in Belgrade					
	<i>Srce</i>	<i>Suočavanje</i>	<i>Novinarstvo</i>	<i>Svedočanstvo</i>	
Descriptive adjective	Concordance number of the collocates analysed on the srWaC corpus				Total number of concordances for the given examples
<i>Hrabro</i> (tokens per million words)	104 (0.19)	26 (0.04)	13 (0.03)	7 (0.01)	150 (0.27)
Near-synonymy samples of the adjective <i>hrabro</i>					
<i>Herojsko</i>	16	/	1	3	20
<i>Odvažno</i>	1	2	4	/	7
<i>Smelo</i>	2	1	1	1	5
<i>Junačko</i>	15	3	/	/	18

4.2 The analysis of adjective *brave* and its near synonyms in English

The most frequent collocations of the node word *brave* were extracted from the corpus data of the website www.just-the-word.com. In the next phase of analysis we analysed co-occurrences of the extracted samples of synonyms within the most frequent collocational framework of the adjective *brave* while using the ukWac Corpus (2.135.658.231 tokens).

A frequency analysis of the four most recurrent collocates of the adjective *brave* (*man, face, attempt, fight*) was carried out within the collocational framework of the suggested near synonyms (*courageous, bold, fearless, dauntless*) (see Table 4). The results of the analysis suggest that the near synonyms of the adjective *brave* are: *bold* (326), *courageous* (176), *fearless* (26), *dauntless* (4) (ukWac corpus, taken on July 28, 2022).

Table 4: Near-synonymy samples of the adjective *brave*

The most frequent four collocates of the descriptive adjective <i>brave</i> analysed on the corpus data of the following website: www.just-the-word.com					
	<i>Man</i>	<i>Face</i>	<i>Attempt</i>	<i>Fight</i>	
Descriptive adjective	Concordance number of the collocates analysed on the srWaC corpus				Total number of concordances for the given examples
<i>Brave</i> (tokens per million words)	1.222 (0.57)	379 (0.18)	278 (0.13)	146 (0.07)	2.025 (0.95)
Near-synonymy samples of the adjective <i>brave</i>					
<i>Courageous</i>	115	1	27	33	176
<i>Bold</i>	56	129	141	/	326
<i>Fearless</i>	19	4	1	2	26
<i>Dauntless</i>	2	/	1	1	4

5. Contrastive analysis⁴

In the process of contrastive analysis application the semantic features of the adjective *hrabar*, *-a*, *-o* and its near-synonyms *herojski* (eng. *heroic*), *odvažan* (eng. *audacious*), *smeo* (eng. *bold*), *junački* (eng. *stout*) have been compared according to the frequency of their appearance with their most recurrent collocates ((~*potez*, ~*čovek*, ~*korak*, ~*čin*, ~*žena* (eng. *woman*), ~*odluka* (eng. *decision*), ~*devojka* (eng. *girl*), ~*igra* (eng. *play*), (~*srce* (eng. *heart*), ~*suočavanje* (eng. *confrontation*), ~*novinarstvo* (eng. *journalism*), *svedočanstvo* (eng. *testimony*)).

The most frequent semantic features of the adjective *hrabar*, *-a*, *-o* in Serbian and the adjective *brave* in English, as well as their near synonyms analysed in the range of their most frequent collocates are as follows:

- (a) the most frequent semantic components of the adjective *hrabar* and its near synonyms (*herojski* (eng. *heroic*), *odvažan* (eng. *audacious*), *smeo* (eng. *bold*),

junački (eng. *stout*) analysed in the following collocational framework (~*potez*, ~*čovjek*, ~*korak*, ~*čin*):

[+MALE±ANIMATE±ADULT]

[+KURAŽAN+NEPOKOLEBLJIV+VITEŠKI] 4: Eng. [+COURAGEOUS+UNDAUNTED+CHIVALROUS]

- (b) the most frequent semantic components of the adjective *hrabra* and its near synonyms (*herojska*, *odvažna*, *smela*, *junačka*) analysed within the given collocational framework (~*žena* (eng. *woman*), ~*odluka* (eng. *decision*), ~*devojka* (eng. *girl*), ~*igra* (eng. *play*)):

[-MALE±ANIMATE±ADULT]

[+SRČANA+KURAŽNA+NEUSTRAŠIVA] 5: Eng. [+BOLD+COURAGEOUS+DAUNTLESS]

- (c) the most frequent semantic components of the adjective *hrabro* and their near synonyms (*herojsko*, *odvažno*, *smelo*, *junačko*) analysed in the range of the following collocates (~*srce* (eng. *heart*), ~*suočavanje* (eng. *confrontation*), ~*novinarstvo* (eng. *journalism*), ~*svedočanstvo* (eng. *testimony*)):

[±MALE±ANIMATE±ADULT]

[+NEUSTRAŠIVO+NEPOKOLEBLJIVO+VITEŠKO] 10: Eng. [+DAUNTLESS+UNDAUNTED+CHIVALROUS]

- (d) the most frequent semantic components of the adjective *brave* and their near synonyms (*courageous*, *bold*, *fearless*, *dauntless*) analysed in the following collocational framework: (*man*, *face*, *attempt*, *fight*):

[-MALE±ANIMATE±ADULT]

[+NEUSTRAŠIV+SRČAN+NEPOKOLEBLJIV] 8: Eng. [+DAUNTLESS+UNDAUNTED+CHIVALROUS]

The most frequent semantic components of the adjective *hrabar*, *-a*, *-o* in Serbian and the adjective *brave* in English are as follows:

[-MALE±ANIMATE±ADULT]

[+NEUSTRAŠIV+SRČAN+NEPOKOLEBLJIV]: Eng. [+DAUNTLESS+BOLD+UNDAUNTED]

In the next stage of our analysis the frequency of use of the descriptive adjective *hrabar*, *-a*, *-o* was analysed within the collocational framework of the descriptive adjective *brave* (its English counterparts) (see Table 5).

Table 5: The frequency of use of the descriptive adjective *hrabar*, *-a*, *-o* within the collocational framework of the descriptive adjective *brave* (its English counterparts)

COLLOCATES OF THE DESCRIPTIVE ADJECTIVE <i>HRABAR</i> , <i>-A</i> , <i>-O</i> extracted from the corpus of contemporary Serbian language, Faculty of Mathematics and Natural Sciences, University of Belgrade (113,000,000 words)		Number of concordances of the descriptive adjective <i>brave</i> analysed within the given collocates of its counterpart <i>hrabar</i> , <i>-a</i> , <i>-o</i> in Serbian, extracted from the database of the British National corpus (112,181,015 words)	Number of concordances of the adjective <i>brave</i> analysed within the given collocates of its counterpart in Serbian <i>hrabar</i> , <i>-a</i> , <i>-o</i> , extracted from the database of ukWac corpus (British Web) (2.135.658.231 tokens)
1.	Potez 38	+Move 5	+186
2.	Čovek 39	+Man 81	+1.222
3.	Korak 13	+Step 7	+148
4.	Čin 13	+Act 1	+37
5.	Žena 21	+Woman 10	+120
6.	Odluka 12	+Decision 15	+248
7.	Devojka 11	+Girl 20	+69
8.	Igra 5	-Game 0	/
9.	Srce 30	+Heart 3	+106
10.	Suočavanje 2	-Coping 0	1
11.	Novinarstvo 1	-Journalism 0	/
12.	Svedočanstvo 1	-Testimony 0	1
TOTAL			
186		142	2.138

The frequency of the use of the descriptive adjective *brave* in the most frequent collocational framework of the adjective *hrabar*, *-a*, *-o* (*move*, *man*, *step*, *act*, *woman*, *decision*, *girl*, *game*, *heart*, *coping*, *journalism*, *testimony*) was carried out (Table 5).

After the analysis, it is determined that there is similarity in the collocation framework of the most common collocations of the adjectives *hrabar*, *-a*, *-o* (186), and its translation equivalents, analyzed in the collocation framework of the

adjective *brave* (142). After carrying out the same analysis based on the larger corpus ukWac (2.135.658.231 tokens), the number of concordances increased to 2.138 (Table 5).

In the continuation of the analysis, the frequency of using the adjectives *hrabar*, *-a*, *-o* in the collocation framework of the descriptive adjective *brave* (its translation equivalents in the Serbian language) was compared (Table 6).

Table 6: The frequency of use of the descriptive adjective *hrabar*, *-a*, *-o* within the collocational framework of the descriptive adjective *brave*

COLLOCATES OF THE DESCRIPTIVE ADJECTIVE <i>brave</i> analysed within the given collocates of its counterpart <i>hrabar</i> , <i>-a</i> , <i>-o</i> in Serbian, extracted from the database of the British National corpus (112,181,015 words)	Number of concordances of the descriptive adjective <i>HRABAR</i> , <i>-A</i> , <i>-O</i> extracted from the corpus of contemporary Serbian language, Faculty of Mathematics and Natural Sciences, University of Belgrade (113,000,000 words)	Number of concordances of the adjective <i>brave hrabar</i> , <i>-a</i> , <i>-o</i> analysed within the given collocates of its counterpart in English <i>brave</i> , extracted from SrWac corpus (554.627.647 tokens)
1. Man 94	-Čovek 0	+567
2. Face 75	+Lice 2	+3
3. World 63	-Svet 0	+4
4. Attempt 29	+Pokušaj 4	+20
5. Girl 22	+Devojka 11	+22
6. Boy 16	-Dečak 0	+31
7. Smile 13	+Osmeh 1	+3
8. Soul 12	+Duša 1	+4
9. Effort 12	-Zalaganje 0	-
10. Fight 10	-Borba 0	+4
11. Soldier 8	+Vojnik 8	+118
12. Warrior 7	+Ratnik 3	+93
TOTAL		
361	30	869

The results of the analysis indicate that the collocational framework of the adjective *brave* (its most frequent collocates) is more specific and in a very small number (30) comparable to the collocational framework of the adjective *hrabar, -a, -o* (361). However, the results are different when the analysis is performed on a larger corpus, such as the SrWac (554.627.647 tokens) (869 concordances) (Table 6).

In the next stage of the analysis a specific collocational framework of both the adjectives *brave* and *hrabar, -a, -o* was contrasted in order to reveal whether there are the same collocates among them, i.e., their counterparts (see Table 7). Therefore, we have applied contrastive analysis in both directions (Serbian to English and English to Serbian).

The results of the analysis imply the existence of two mutual most frequent collocates of the adjective *hrabar, -a, -o* and the most frequent collocates of the adjective *brave* (see Table 7).

Table 7: Comparison of specific collocational framework of both the adjectives *brave* and *hrabar, -a, -o*

COLLOCATES OF THE DESCRIPTIVE ADJECTIVE <i>HRABAR, -A, -O</i> extracted from the corpus of contemporary Serbian language, Faculty of Mathematics and Natural Sciences, University of Belgrade (113,000,000 words)	COLLOCATES OF THE ADJECTIVE <i>BRAVE</i> extracted from the British National corpus (112,181,015 words)
Potez 38	Man 94
Čovek 39	Face 75
Korak 13	World 63
Čin 13	Attempt 29
Žena 21	Girl 22
Odluka 12	Boy 16
Devojka 11	Smile 13
Igra 5	Soul 12
Srce 30	Effort 12
Suočavanje 2	Fight 10
Novinarstvo 1	Soldier 8
Svedočanstvo 1	Warrior 7
186	361

Throughout this analysis it has been noticed that there are the following mutual collocates: *čovjek/man* and *devojka/girl*, which are the translation equivalents of the descriptive adjectives *brave* and *hrabar, -a, -o*.

The most frequent collocates of the near-synonyms of the adjective *brave* were also analyzed on the basis of data taken from the British National Corpus (112,181,015 words) (Table 8).

COURAGEOUS	MAN (11)	DECISION (9)	PEOPLE (6)	EFFORT (5)
BOLD	MOVE (21)	STEP (15)	STATEMENT (16)	ATTEMPT (12)
FEARLESS	PATRIOT (2)	INVENTORY (1)	CREATIVITY (1)	HUNTER (1)
DAUNTLESS	JAVELOT (4)	CAVALRYMAN (1)	CONVERSATION (1)	BOY (1)
BRAVE	MAN (94)	FACE (75)	WORLD (63)	ATTEMPT (29)

Table 8

The analysis has shown that the mutual collocates of the adjective *brave* and its near-synonyms are as follows:

courageous/brave man
bold/brave attempt

Among the most frequent collocates of the adjective *brave* and its near-synonyms the following ones have a transferred meaning:

courageous decision/effort
bold move/step/statement/attempt
fearless inventory/creativity
dauntless conversation
brave face/world/attempt

We have performed an analysis of the most frequent collocates of near-synonyms of the adjective *hrabar, -a, -o* (*herojski, -a, -o, junački, -a, -o, odvažan, -a, -o, smeo, -la, -lo*) endorsing the corpus data of the Contemporary Serbian language, Faculty of Mathematics, University of Belgrade (113,000,000) (see Table 9).

HEROJSKI	OTPOR (24)	NAROD (11)	ČIN (9)	PODUHVAT (5)
JUNAČKI	PODVIG (3)	OTPOR (2)	ČIN (2)	POKLIČ (2)
ODVAŽAN	ČOVEK (3)	PODUHVAT (3)	MOMAK (2)	VOJSKOVOĐA (2)
SMEO	/	/	/	/
HRABAR	ČOVEK (39)	POTEZ (38)	KORAK (13)	ČIN (13)

¹⁰ HEROJSKA	DELA (9)	BORBA (9)	ODBRANA (8)	POBEDA (4)
¹¹ JUNAČKA	DELA (10)	KRV (3)	SMRT (3)	PESMA (2)
¹² ODVAŽNA	DAMA (2)	DEVOJKA (1)	TINEJDŽERKA (1)	STARICA (1)
¹³ SMELA	/	/	/	/
¹⁴ HRABRA	ŽENA (21)	ODLUKA (12)	DEVOJKA (11)	IGRA (5)
¹⁵ HEROJSKO	DELO (12)	DOBA (10)	VREME (2)	SAMOŽRTVOVANJE(5)
¹⁶ JUNAČKO	DELO (12)	DRŽANJE (10)	ZDRAVLJE (7)	SRCE (4)
¹⁷ ODVAŽNO	DELO (2)	ZAUZIMANJE (1)	SUČELJAVANJE (1)	SRCE (1)
¹⁸ SMELO	/	/	/	/
¹⁹ HRABRO	SRCE (30)	SUOČAVANJE (2)	NOVINARSTVO (1)	SVEDOČANSVO (1)

Table 9

The mutual collocates of the adjective *hrabar*, *-a*, *-o* and its near-synonyms are as follows:

herojski/junački otpor (Eng. *heroic/stout resistance*);
herojski/junački čin (Eng. *heroic/stout act*);
odvažan/hrabar čovek (Eng. *audacious/brave man*);
herojska/junačka dela (Eng. *heroic/stout deeds*);
odvažna/hrabra devojka (Eng. *audacious/brave girl*);
herojsko/junačko/odvažno delo (Eng. *heroic/stout/audacious deed*);
odvažno/hrabro srce (Eng. *audacious/brave heart*);

While analysing the most frequent collocates of the adjective *hrabar*, *-a*, *-o* and its near-synonyms it has been noticed that the following ones have the transferred meaning:

herojski otpor/čin/poduhvat (Eng. *heroic/resistance/act/achievement*);
junački otpor/čin/poklič (Eng. *heroic resistance/act/cry*);
odvažan poduhvat (Eng. *audacious achievement*);
hrabar potez/korak/čin (Eng. *brave move/step/act*);
herojska dela/borba/odbrana/pobeda (Eng. *heroic deeds/battle/defence/victory*);
hrabra odluka/igra (Eng. *brave decision/game*);
junačka krv/smrt (Eng. *stout blood/death*);
herojsko delo/doba/vreme/samožrtvovanje (Eng. *heroic deed/age/time/self-sacrifice*);
junačko delo/držanje/zdravlje/srce (Eng. *stout deed/holding/health/heart*);
odvažno delo/zauzimanje/sučeljavanje/srce (Eng. *audacious deed/advocating/confrontation/heart*);
hrabro srce/suočavanje/novinarstvo/svedočanstvo (Eng. *brave heart/confrontation/journalism/testimony*)

The analysis employed has shown that in addition to the denotative meaning of steadfastness, heartiness, fearlessness, as human characteristics, this adjective, as well as its near synonyms, indicate the heartiness, fearlessness and steadfastness of certain inanimate concepts (*resistance, act, endeavor, shout, step, move, deeds, fight, defense, victory, decision, game, blood, death, time, heart, journalism, confrontation ...*). The analysis of the adjective *brave* and its near-synonyms in the stated collocational framework: *courageous decision/effort; bold move/step/statement/attempt; fearless inventory/creativity; dauntless conversation; brave face/world/attempt* also shows that some traits characteristic of human beings have been attributed to certain objects and phenomena in English. The conducted research indicates prevailing idiomatic meaning of the adjectives analysed in both languages.

6. Conclusion

The multi-faceted approach to the research employed in this paper incorporates the collocational, componential, and contrastive analysis of the adjectives' extracted collocates and the semantic content of their near synonyms in English and Serbian.

The first significant result of the approach applied in this paper is related to the specific method of selecting near synonyms that consists of four stages of the analysis:

- (1) After analyzing the first ten synonyms (extracted from the dictionary of synonyms; see Addendum 2) through their co-occurrences within the most frequent nominal collocations of the node word (corpus-based analysis; see Addendum 1) four synonyms were considered to be near synonyms of the selected descriptive adjective.
- (2) In the second stage of the research, frequency analysis of the four most recurrent collocates of the adjective *hrabar* (masculine gender) was carried out within the collocational framework of the suggested near synonyms.
- (3) The third stage of the research includes frequency analysis of the four most recurrent collocates of the adjective *hrabra* (female gender) within the collocational framework of the suggested near synonyms.
- (4) The final stage of the research involves frequency analysis of the four most recurrent collocates of the adjective *hrabro* (neuter gender) within the collocational framework of the suggested near synonyms.

Within this frequency-based approach to entrenchment within the usage-based understanding of language (Glynn 2014: 15) the following valid result of the analysis highlights the influence of grammatical gender (masculine, female, neuter) on the various most frequent collocates of the analysed descriptive adjective when seen from the perspective of each gender:

For example:

hrabar (masculine gender): *potez* (567); *čovjek* (224); *korak* (88); *čin* (43) (see Table 1)

hrabra (female gender): *žena* (107); *odluka* (39); *devojka* (22); *igra* (18) (see Table 2)

hrabro (neuter gender): *srce* (104); *suočavanje* (26); *novinarstvo* (13); *sučeljavanje* (7) (see Table 3)

In the research undertaken it has been found that besides the influence of the most frequent collocational framework of the adjective on the choice of its near synonyms, this interrelation depends on the adjective's grammatical gender and also the near synonymy choice, especially regarding the near-synonyms' proximity-of-meaning order.

Namely, it has been noticed that this proximity-of-meaning order of near synonyms varies with the gender implied (masculine, feminine, neuter):

For example:

Near-synonyms of the adjective *hrabar* (masculine gender): *herojski* (137) *smeo* (78), *odvažan* (62), *junački* (44) (see Table 1)

Near-synonyms of the adjective *hrabra* (female gender): *smela* (20), *herojska* (17), *odvažna* (14), *junačka* (4) (see Table 2)

Near-synonyms of the adjective *hrabro* (neuter gender): *herojsko* (20), *junačko* (18), *odvažno* (7), *smelo* (5) (see Table 3)

The same interrelation between the word order of near-synonyms' proximity-of-meaning and grammatical gender was not noticed in the English language featuring natural language.

In this respect, the importance of the research lies in emphasizing the disparity between the English and Serbian language in the issues of grammatical gender.

An analysis of the descriptive adjective and its near-synonyms carried out within the most frequent collocational framework reveals a great number of samples implying idiomatic meanings in English and Serbian. The stated result is in line with the conclusion drawn by Firth (1957) claiming that: "you shall know the word by the company it keeps" and Dragičević (2010: 156) saying that: "the meaning of the lexemes is determined and shaped by the context, thus each new context features new semantic components of the lexeme, while the other components are shadowed." Therefore, this result proves that the collocational framework of the lexeme determines the meaning of the lexeme with more precision than its denotational meaning does.

The samples of collocates indicate a prevailing idiomatic meaning of the adjective *hrabar*, *-a*, *-o* and its near-synonyms while being analysed in the most frequent collocational framework. It can be concluded that besides a denotational meaning, this adjective and its near-synonyms imply other nuances of transferred meaning indicating the heartiness, fearlessness and steadfastness of

certain inanimate concepts (*resistance, act, endeavor, shout, step, move, deeds, fight, defense, victory, decision, game, blood, death, time, heart, journalism, confrontation ...*).

As a result of the contrastive analysis applied it has been noticed that the collocational framework of the descriptive adjectives in Serbian is more comparable to the collocational framework of the analysed adjectives in English (see Table 5), whereas the most frequent collocational framework of the descriptive adjective in English is rather specific and a lot less comparable to the collocational framework (of its Serbian counterparts) in Serbian (see Table 6):

<i>hrabar, -a, -o</i>	<i>brave</i>
Serbian (186)/ English (142)	English (361) /Serbian (30)

This result is in accordance with the further analysis which showed a small number of mutual collocates (their counterparts) in both English and Serbian. These results confirm the starting hypothesis of this paper that the collocational frameworks of words are highly dependent on the semantic-syntactic specificity of the language.

However, contrastive analysis of the most frequent semantic features of descriptive adjectives and their near-synonyms in Serbian and English showed that there are a great number of their most frequent semantic features (see Table 7). This finding poses an interesting question: "How can it be possible that in the analysis of the most frequent collocational frameworks of the adjectives analysed in Serbian and its counterparts in English, a small degree of mutual comparability has been noticed, and on the other hand, a separate analysis showed a great number of their most frequent semantic features? The answer to this question is related to the specific way of classifying semantic features into primary semes, and distinctive and peripheral features, as presented by Dragićević (2010).

Applying this method in future thesauri of English and Serbian dictionaries of synonyms would undoubtedly improve their quality by making differences in the meaning of near-synonyms more transparent, taking into account the obvious lack of presentation of the nuances in meaning among the given synonyms of the selected node word. At present, synonyms in the English and Serbian dictionaries are listed either in alphabetical order or randomly without paying attention to their proximity in meaning. Therefore, it has been proposed that list of synonyms be included next to the node word following the principle of proximity of meaning, not randomly or alphabetically. More precisely, the most frequent collocational framework of the node word should be extracted, within which the near-synonymy samples should be analysed, thus measuring their proximity of meaning to the node word. Moreover, alongside the suggested method, the use of synonyms within example sentences should be added, showing the nuances of their meaning within the context. In addition, as only adjectives in masculine gender are stated in Serbian dictionaries of synonyms and there is no insight into the specific semantic features of adjectives

tives in female and neuter gender, these samples of the adjectives analysed should be included.

Endnotes

1. Fuertes-Olivera and Tarp (2022), Ledinek and Michelizza (2021), Norri (2019), Mathiasen (2017), Westveer, Sleeman and Aboh (2018) investigated issues of gender-biased dictionaries.
2. Method of choosing the first four near-synonymy samples has been presented in Addendum 2.
3. Method of choosing the most frequent collocational framework of the adjective analysed has been presented in Addendum 1.
4. See Fisiak (1971); Filipović (1984); Mair and Markus (1992); Hawkins (1986); House (1996); Johansson (2007); Xiao and McEnery (2006).
5. Eng. HEROIC (male) RESISTANCE (24); PEOPLE (11); ACT (9); ENDEAVOUR (5)
6. Eng. STOUT (male) ACHIEVEMENT (3); RESISTANCE (2); ACT (12) CRY (2)
7. Eng. AUDACIOUS (male) MAN (3); ENDEAVOUR (3); BOY (2); MILITARY LEADER (2)
8. Eng. BOLD (male)
9. Eng. BRAVE (male) MAN (39); MOVE (38); STEP (13); ACT (5)
10. Eng. HEROIC (female) DEEDS (9); BATTLE (9); DEFENSE (8); VICTORY (4)
11. Eng. STOUT (female) DEEDS (10); BLOOD (3); DEATH (3); POEM (2)
12. Eng. AUDACIOUS (female) LADY (2); GIRL (1); TEENAGER (1); OLD WOMAN (1)
13. Eng. BOLD (female)
14. Eng. BRAVE (female) WOMAN (21); DECISION (12); GIRL (11); GAME (5)
15. Eng. HEROIC (neuter) DEED (12); AGE (10); TIME (2); SELF-SACRIFICE (5)
16. Eng. STOUT (neuter) DEED (12); HOLDING (10); HEALTH (7); HEART (4)
17. Eng. AUDACIOUS (neuter) DEED (2); ADVOCATING (1); CONFRONTATION (1); HEART (1)
18. Eng. BOLD (neuter)
19. Eng. BRAVE (neuter) HEART (30); CONFRONTATION (2); JOURNALISM (1); TESTIMONY (1)

References

Dictionaries

- Ćosić, P. 2008. *Rečnik sinonima i tezaurs srpskog jezika*. Beograd: Kornet.
- Lalević, M.S. 1974. *Sinonimi i srodne reči srpskohrvatskoga jezika*. Beograd: Leksikografski zavod Sveznanje.

Electronic corpora

- British National Corpus (BNC). www.natcorp.ox.ac.uk [Accessed on October 2, 2021].
- Corpus of Contemporary Serbian language, Faculty of Mathematics and Natural Sciences, University of Belgrade. korpus.matf.bg.ac.rs [Accessed on July 28, 2022].
- SrWac (Serbian Web). https://www.clarin.si/noske/run.cgi/corp_info?corpname=srwac&struct_attr_stats=1 [Accessed on July 28, 2022].

ukWac (British Web). https://www.clarin.si/noske/run.cgi/corp_info?corpname=ukwac&struct_attr_stats=1 [Accessed on July 28, 2022].

Other literature

- Arppe, A.** 2008. *Univariate, Bivariate, and Multivariate Methods in Corpus-based Lexicography — A Study of Synonymy*. Doctoral Thesis. Helsinki: University of Helsinki.
- Bendix, E.H.** 1996. *Componential Analysis of General Vocabulary. The Semantic Structure of a Set of Verbs in English, Hindi and Japanese*. Bloomington: Indiana University Press.
- Cai, J.** 2012. *Is it "great" Enough? A Corpus-based Study of "great" and its Near Synonyms*. Master Thesis. Muncie, Indiana: Ball State University.
- Cappelli, C.** 2011. Grouping Near-synonyms of a Dictionary Entry: Thesauri and Perceptions. *Quaderni di Statistica* 13: 53-67.
- Crawford, W.J. and E. Csomay.** 2016. *Doing Corpus Linguistics*. London, UK: Routledge.
- Cruse, A.** 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Desagulier, G.** 2014. Visualizing Distances in a Set of Near-synonyms. *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy* 43: 145-178.
- Divjak, D.** 2010. *Structuring the Lexicon. A Clustered Model for Near-synonymy*. Cognitive Linguistics Research 43. Berlin/New York: Mouton de Gruyter.
- Divjak, D. and Stefan Th. Gries.** 2006. Ways of Trying in Russian: Clustering Behavioural Profiles. *Corpus Linguistics and Linguistic Theory* 2(1): 23-60.
- Dragičević, R.** 2010. *Leksikologija srpskog jezika*. Beograd: Zavod za udžbenike.
- Edmonds, P. and G. Hirst.** 2002. Near-synonymy and Lexical Choice. *Journal of Computational Linguistics* 28(2): 105-144.
- Fachrutdinov, R.R., V.N. Khisamova and L.G. Khisimatullina.** 2017. Comparative Analysis of Semantic Distinctions between Synonymous Adjectives in Tatar and English *European Research Studies Journal* XX, Special Issue: 447-457.
- Filipović, R.** 1984. What are the Primary Data for Contrastive Analysis. Fisiak, J. (Ed.). 1984. *Contrastive Linguistics*: 107-108. Berlin: Mouton de Gruyter.
- Firth, J.R.** 1957. A Synopsis of Linguistic Theory 1930–1955. *Studies in Linguistic Analysis. Special Volume of the Philological Society*: 1-32. Oxford: Blackwell.
- Fisiak, J.** 1971. The Poznań Polish–English Contrastive Project. Filipović, R. (Ed.). 1971. *Zagreb Conference on English Contrastive Projects*: 87-96. Zagreb: Institute of Linguistics, University of Zagreb.
- Fuertes-Olivera, P.A. and S. Tarp.** 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132.
- Geeraerts, D.** 2010. *Theories of Lexical Semantics*. New York/Oxford: Oxford University Press.
- Geeraerts, D., S. Grondelaers and P. Bakema.** 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin: Mouton de Gruyter.
- Gimaltdinova, G., L. Khalitova, V. Solovyev and V. Bochkarova.** 2021. Lexicographic Study of Synonymy: Clarifying Semantic Similarity Between Words. *Computacion y Sistemas* 25(3): 667-675.
- Glynn, D.** 2014. Corpus-driven Lexical Semantic Evidence for Conceptual Structure of ANGER. Novakova, I., P. Blumenthal and D. Siepmann (Eds.). 2014. *New Directions in Lexical Semantics and Discourse Organization*: 69-83. Frankfurt: Peter Lang.

- Gries, Stefan Th.** 2006. Corpus-based Methods and Cognitive Semantics: The Many Senses of *to run*. Gries, Stefan Th. and A. Stefanowitsch (Eds.). 2006. *Corpora in Cognitive Linguistics. Corpus-based Approaches to Syntax and Lexis*: 57-99. Berlin/New York: Mouton de Gruyter.
- Gries, Stefan Th. and A. Stefanowitsch.** 2004. Extending Collocational Analysis: A Corpus-based Perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1): 97-129.
- Gu, B.J.** 2017. Corpus-Based Study of Two Synonyms Obtain and Gain. *Sino-US English Teaching* 14(8): 511-522.
- Harris, Z.S.** 1954. Distributional Structure. *Word* 10(2-3): 146-162.
- Harris, Z.S.** 1957. Co-occurrence and Transformation in Linguistic Structure. *Language* 33(3): 283-340.
- Hawkins, J.A.** 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. London: Croom Helm.
- Herda, D.** 2020. Arrangement Classifiers, Collocations, and Near-synonymy: A Corpus-based Study with Reference to Polish. *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 137(4): 245-258.
- Hirst, G.** 1995. Near-synonymy and the Structure of Lexical Knowledge. Papers from the 1995 AAAI Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity, 27-29 March 1995, Stanford University: 51-56. Stanford, CA: Stanford University.
- Hoey, M.** 2005. *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge.
- House, J.** 1996. Contrastive Discourse Analysis and Misunderstanding: The Case of German and English. Hellinger, M. and U. Amon. 1996. *Contrastive Sociolinguistics*: 345-361. Berlin/New York: Mouton de Gruyter.
- Inkpen, D. and G. Hirst.** 2006. Building and Using a Lexical Knowledge Base of Near-synonym Differences. *Computational Linguistics* 32(2): 223-262.
- Jarunwaraphan, B. and P. Mallikamas.** 2020. A Corpus-based Study of English Synonyms: 'chance' and 'opportunity'. *REFlections* 27(2): 218-245.
- Jiranthiporn, S.** 2018. Is this Problem Giving you Trouble? A Corpus-based Examination of the Differences between the Nouns 'problem' and 'trouble'. *Thoughts* 2018(2): 1-25.
- Johansson, S.** 2007. Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies. Facchinetti, R. (Ed.). 2007. *Corpus Linguistics 25 Years On*: 51-71. Amsterdam/New York: Rodopi.
- Kamiński, M.P.** 2017. Visualisation of Collocational Preferences for Near-Synonym Discrimination. *Lexikos* 27: 237-251.
- Kruawong, T. and S. Phoocharoensil.** 2022. A Genre and Collocational Analysis of the Near-Synonyms *teach, educate* and *instruct*: A Corpus-Based Approach. *TEFLIN Journal* 33(1): 75-97.
- Ledinek, N. and M. Michelizza.** 2021. Gender in Slovenian Monolingual General Explanatory Dictionaries. *Collegium antropologicum* 45(4): 329-339.
- Lewis, M.** 2000. There is Nothing as Practical as a Good Theory. Lewis, M. (Ed.). 2000. *Teaching Collocation: Further Developments in the Lexical Approach*: 10-27. Hove, UK: Language Teaching Publications.
- Liu, D.** 2010. Is it a Chief, Main, Major, Primary, or Principal Concern: A Corpus-based Behavioral Profile Study of the Near-synonyms. *International Journal of Corpus Linguistics* 15(1): 56-87.
- Liu, D.** 2013. Using Corpora to Help Teach Difficult-to-distinguish English Words. *English Teaching* 68(3): 27-50.
- Liu, D. and M. Espino.** 2012. Actually, Genuinely, Really, and Truly: A Corpus-Based Behavioral Profile Study of Near-synonymous Adverbs. *International Journal of Corpus Linguistics* 17(2): 198-228.

- Liu, W. 2020. A Corpus-Based Study of Semantic Prosody and Semantic Preference of "seem". *Open Journal of Social Sciences* 8: 282-291.
- Mair, C. and M. Markus (Eds.). 1992. *New Departures in Contrastive Linguistics. Proceedings of the Conference Held at the Leopold-Franzens-University of Innsbruck, Austria, 10-12 May 1991*. 2 Vols. Innsbruck: University of Innsbruck Press.
- Mathiasen, E.J. 2017. A Study of Gender in a Bilingual Law Dictionary (English/Spanish). *Revista Española de Lingüística Aplicada* 30(1): 370-394.
- Miller, G.A. 1999. On Knowing a Word. *Annual Review of Psychology* 50(1): 1-19.
- Miller, G.A. and W.G. Charles. 1991. Contextual Correlates of Semantic Similarity. *Language and Cognitive Processes* 6(1): 1-28.
- Nida, E.A. 1975. *Componential Analysis of Meaning*. The Hague/Paris: Mouton.
- Norri, J. 2019. Gender in Dictionary Definitions: A Comparison of Five Learner's Dictionaries and Their Different Editions. *English Studies* 100(7): 866-890.
- O'Dell, F. and M. McCarthy. 2008. *English Collocations in Use: Advanced*. Cambridge: Cambridge University Press.
- Opašić, M. 2010. Hrabar, odvažan, izazovan ... pothvat Ljiljane Šarić i Wiebke Wittschen. *Fluminensia: časopis za filološka istraživanja* 22(2): 175-180 (review).
- Palmer, F.R. 1981. *Semantics*. Cambridge: Cambridge University Press.
- Petcharat, N. and S. Phoocharoensil. 2017. A Corpus Based Study of English Synonyms: appropriate, proper and suitable. *LEARN Journal: Language Education and Acquisition Research Network* 10(2): 10-24.
- Petrović, B. 2005. *Sinonimija i sinonimičnost u hrvatskome jeziku*. Zagreb: Hrvatska sveučilična naklada.
- Phoocharoensil, S. 2020. A Genre and Collocational Analysis of Consequence, Result, and Outcome. *3L: Southeast Asian Journal of English Language Studies* 26(3): 1-16.
- Phoocharoensil, S. 2021. Semantic Prosody and Collocation: A Corpus Study of the Near-synonyms *persist* and *persevere*. *Eurasian Journal of Applied Linguistics* 7(1): 240-258.
- Piits, L. 2013. Distributional Hypothesis: Words for 'human being' and their Estonian Collocates. *Trames* 17(2): 141-158.
- Prčić, T. 2010. Sinonimija u teoriji i praksi: isto ali ipak različito. *Jezik danas*. Available at: www-gewi.uni-graz.at/gralis/Liguistikarium/Lexikologie/BKS/Lexikologie_BKS/Prcic_Sinonimija.htm [Last accessed 10 March 2021].
- Quine, W.v.O. 1951. Two Dogmas of Empiricism. *The Philosophical Review* 60: 20-43. (Reprinted in Quine, W.v.O. 1953. *From a Logical Point of View*. Cambridge: Harvard University Press.)
- Sayyed, S.W. and R.R. Al-Khanji. 2019. A Corpus Based Analysis of Eight English Synonymous Adjectives of Fear. *International Journal of Linguistics* 11(1): 111-138.
- Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. (Edited with Ronald Carter.) London/ New York: Routledge.
- Song, Q. 2021. Effectiveness of Corpus in Distinguishing Two Near-Synonymous Verbs: 'Damage' and 'Destroy'. *English Language Teaching* 14(7): 8-20.
- Stoyanova-Georgieva, I. 2020. Expressing Attitude and Evaluation through Intensifiers Examined through the Prism of Semantic Prosody. *Knowledge — International Journal* 38(6): 1403-1408.
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Sumonsriworakun, P. 2022. A Corpus-based Investigation of English Synonyms: Disadvantage, Downside and Drawback. *LEARN Journal: Language Education and Acquisition Research Network* 15(2): 649-678.

- Thongpan, N.** 2022. A Corpus-Based Study of English Synonyms of the Adjectives 'Far', 'Distant' and 'Remote'. *Journal of Positive School Psychology* 6(6): 3986-4001.
- Turney, P.D. and P. Pantel.** 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research* 37: 141-188.
- Westveer, T., P. Sleeman and E.O. Aboh.** 2018. Discriminating Dictionaries? Feminine Forms of Profession Nouns in Dictionaries of French and German. *International Journal of Lexicography* 31(4): 371-393.
- Xiao, R. and T. McEnery.** 2006. Collocation, Semantic Prosody and Near Synonymy: A Cross-linguistic Perspective". *Applied Linguistics* 27(1): 103-129.
- Yang, B.** 2020. A Corpus-based Study of Synonymous Epistemic Adverbs Perhaps, Probably, Maybe and Possibly. *Research Journal of Education* 6(8): 158-168.
- Tao, Y.** 2021. Study of Near Synonymous Mental-State Verbs: A MARVS Perspective. *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*: 432-441. Shanghai, China: Association for Computational Linguistics.
- Zhang, F.** 2002. *Componential Analysis Method*. Munich: GRIN Publishing GmbH.

Addendum 1: A method of determining the most frequent collocational framework on the example of the descriptive adjective *hrabar*

The first 250/526 extracted collocates of the adjective *hrabar* in the Corpus of the Contemporary Serbian language of the Faculty of Mathematics and Natural Sciences in Belgrade (113,000,000 words). The concordance list of extracted co-occurrences for the adjective *hrabar* is as follows:

Oficir (eng. *officer*) 2
Potez (eng. *move*) 24
Borac (eng. *fighter*) 10
Lik (eng. *face*) 1
Komandant (eng. *commander*) 1
Čovek (eng. *man*) 38
Vojnik (eng. *soldier*) 4
Političar (eng. *politician*) 2
Narod (eng. *nation*) 5
Korak (eng. *step*) 13
Državnik (eng. *stateman*) 1
Pristup (eng. *approach*) 3
Čin (eng. *act*) 12
Borac (eng. *warrior*) 2
Kritičar (eng. *critic*) 1
Sudija (eng. *judge*) 1
Pokušaj (eng. *attempt*) 4
Odgovor (eng. *reply*) 1
Novinar (eng. *journalist*) 1
Život (eng. *life*) 4
Roman (eng. *novel*) 5
Duh (eng. *spirit*) 1
Iskorak (eng. *stride*) 2
Dječak (eng. *boy*) 2
Rez (eng. *cut*) 1
Skok (eng. *jump*) 2
Kurs (eng. *course*) 1
Tekst (eng. *text*) 1
Mačak (eng. *tomcat*) 1
Program (eng. *programme*) 1
Muzičar (eng. *musician*) 1

According to the analysis applied the most frequent collocates of the adjective *hrabar* are as follows: *čovek* (eng. *man*) (38), *potez* (eng. *move*) (24), *korak* (eng. *step*) (13) and *čin*, (eng. *act*) (12).

In the conducted analysis extracted tokens were not lemmatized and only exact words were considered.

The same method has been applied for the feminine gender (*hrabra*) and neuter gender (*hrabro*).

Addendum 2: A method of determining near-synonymy samples of the adjective on the example of the descriptive adjective *hrabar*

We have extracted the first 10/15 synonyms of the adjective *hrabar* from the *Dictionary of synonyms* by Pavle Ćosić and his assistants (2008). The co-occurrence of the extracted samples within the most frequent collocational framework of the adjective *hrabar* has been determined while using the SrWac Corpus (554.627.647 tokens):

	move	man	step	act	total
Junački (eng. <i>stout</i>)	2	7	26	9	44
Smeo (eng. <i>bold</i>)	15	38	32	3	78
Herojski (eng. <i>heroic</i>)	6	/	23	108	137
Junačan (eng. <i>manly</i>)	2	2	2	6	12
Viteški (eng. <i>chivalrous</i>)	2	1	/	9	11
Kuražan (eng. <i>courageous</i>)	2	5	2	1	10
Srčan (eng. <i>stout-hearted</i>)	3	4	/	3	10
Odvažan (eng. <i>audacious</i>)	6	43	11	2	62
Neustrašiv (eng. <i>dauntless</i>)	/	9	2	/	9
Nepokolebljiv (eng. <i>undaunted</i>)	/	2	1	2	5

According to the conducted analysis near-synonymy samples of the adjective *hrabar* are as follows: *herojski* (137), *smeo* (78), *odvažan* (62) and *junački* (44).

In the conducted analysis extracted tokens were not lemmatized and only exact words were considered.

The same method has been applied for the feminine gender (*hrabra*) and neuter gender (*hrabro*).

'n Prototipe-aanlyn Afrikaanse akademiese redigeerwoordeboek*

Maret Blom de Wet, *Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Suid-Afrika (maretblom@sun.ac.za)*

Opsomming: Die nagraadse studente aan hoërondewysinstellings in die Suid-Afrikaanse konteks het volgens Van Aswegen (2007: 1141) onvoldoende akademiese skryfvaardighede en 'n tekort aan navorsingsvaardighede; daarom groei die aanvraag na die diens van redigeerders wat spesifiek akademiese tekste redigeer. Die redigeerders van Afrikaanse akademiese tekste ondervind egter probleme met naslaanbronne (stylgidse of standaardiseringsbronne) wat hulle kan gebruik om konsekwentheid in die akademiese tekste te verseker. Gevolglik het hierdie redigeerders 'n behoefte aan 'n Afrikaanse woordeboek wat spesiaal gerig is op die behoeftes van akademiese redigeerders (Blom 2020: 18). 'n Model vir die ontwerp van 'n aanlyn Afrikaanse akademiese redigeerwoordeboek (Blom se 2018-woordeboekmodel) is opgestel as die eerste deel van 'n groter projek wat daarop gemik is om 'n volledige aanlyn Afrikaanse akademiese redigeerwoordeboek saam te stel. Ten einde Blom se 2018-woordeboekmodel te implementeer om 'n volledige woordeboek saam te stel, moet 'n prototipewoordeboek ontwikkel word en dan getoets word deur sy teikengebruikers (akademiese redigeerders).

In hierdie artikel is die beginsels van die funksieteorie (Fuentes-Olivera en Tarp 2014), Gouws (2014a; 2014b; 2018a; 2018b; 2018c; 2018d) se aanpassings van die algemene leksikografieteorie en bruikbaarheidsteorieë (ISO-standaard 9241 (1998a)(1998b)(2006)) gebruik om die genoemde prototipe-Afrikaanse akademiese redigeerwoordeboek saam te stel. Hierdie geïntegreerde teorie maak die samestelling van 'n gebruikersvriendelike prototipewoordeboek moontlik, wat in 'n opvolgstudie deur herhaalde bruikbaarheidstoetse met die eind- of teikengebruikers as toetsdeelnemers geëvalueer sal word.

Sleutelwoorde: WOORDEBOEKSAMESTELLING, AKADEMIESE REDIGEERWOORDEBOEK, PROTOTIPEWOORDEBOEK, ELEKTRONIESE LEKSIKOGRAFIE, FUNKSIETEORIE, ALGEMENE LEKSIKOGRAFIETEORIE, BRUIKBAARHEIDSTEORIEË, GEÏNTEGREERDE TEORIE

Abstract: A Prototype Afrikaans Online Dictionary for Academic Editing Purposes. The academic writing and research skills of postgraduate students at higher education institutions in the South African context are inadequate (Van Aswegen 2007: 1141), and therefore there is a growing demand for the services of editors of specifically academic texts. The editors of Afrikaans academic texts, however, experience problems in terms of reference sources (style guides

* 'n Weergawe van hierdie artikel is as referaat gelewer by die 26ste Jaarlikse Internasionale Konferensie van die African Association for Lexicography (AFRILEX), georganiseer deur Stellenbosch Universiteit, Stellenbosch, Suid-Afrika, 27–29 Julie 2022.

or standardisation sources) that they can use to ensure consistency in the academic texts. Consequently, these editors have a need for an Afrikaans dictionary that is specially aimed at the needs of academic editors (Blom 2020: 18). A model for the design of an Afrikaans online dictionary for academic editing purposes (Blom's 2018 dictionary model) was set up as the first part of a larger project that aims to compile a complete online Afrikaans academic editing dictionary. In order to implement Blom's 2018 dictionary model to compile a complete dictionary, a prototype dictionary must be developed and then tested by its target users (i.e. academic editors).

In this article the principles of the function theory (Fuentes-Olivera and Tarp 2014), Gouws' (2014a; 2014b; 2018a; 2018b; 2018c; 2018d) adaptations of the general lexicography theory and usability theories (ISO-standard 9241 (1998a)(1998b)(2006)) were used to compile the said prototype Afrikaans academic editing dictionary. This integrated theory enables the compilation of a user-friendly prototype dictionary that will be evaluated through repeated usability testing with the end or target users as test participants in a follow-up study.

Keywords: DICTIONARY COMPILATION, ACADEMIC EDITING DICTIONARY, PROTOTYPE DICTIONARY, ELECTRONIC LEXICOGRAPHY, FUNCTION THEORY, GENERAL LEXICOGRAPHY THEORY, USABILITY THEORIES, INTEGRATED THEORY

1. Inleiding

In die akademiese konteks lei studente se gebrekkige akademiese skryfvaardighede en 'n tekort aan navorsingsvaardighede (Van Aswegen 2007: 1141), asook hoëronderwysinstellings se verpligting dat studente se tesisse as deel van die afrondingsproses daarvan geredigeer moet word, tot 'n definitiewe aanvraag na die diens van redigeerders wat spesifiek akademiese tekste redigeer (Blom 2021: 1). Hierdie redigeerdiens sluit die redigering in van verskillende soorte tekste wat studente én dosente in die sfeer van akademiese instellings publiseer, insluitend akademiese of wetenskaplike artikels, referate en studies vir publikasiedoeleindes, konferensievoorleggings, proefskrifte en tesisse (Law 2011: 227). Volgens die omskrywing van die Universiteit Stellenbosch Taalsentrum (2019) se diens vir teksredigering en soos uitgelig in Lourens (2014: 267) moet akademiese redigeerders 'n veelheid van take kan verrig wat in die praktyk hoofsaaklik bestaan uit werk op die vlakke van kopieredigering en styl¹. Ten einde hierdie redigeertake te kan verrig, het akademiese redigeerders taalbronne (taalhulpmiddels) nodig om onder andere die taalgebruik na te gaan en te help om verskillende probleme op te los wat tydens die redigeerproses ontstaan (Carstens en Van de Poel 2012: 406). 'n Stylgids of normeringsbron moet ook gebruik word om kopieredigeringsaspekte soos spel-, tik en basiese grammatikale foute volgens die stylkonvensies uit te wys of te korrigeer (Blom 2021: 2). Aangesien teksredakteurs nie oor die volledige kennis van 'n taal beskik nie, moet redigeerders juis daarna streef om die beste ondersteuning (dus naslaan- en kontrolebronne) te kry om probleme in die redigeerproses te kan oplos (Carstens en Van de Poel 2012: 406-407). In die Afrikaanse akademiese redigeerpraktyk is dit egter 'n probleem dat van die naslaanbronne in Afrikaans verouderd is en van die vakwoordeboeke uit druk

en moeilik bekombaar is (Carstens en Van de Poel 2012: 408).

In Blom (2020) is daar ondersoek ingestel na die huidige Afrikaanse naslaanbronne wat vir akademiese redigeerders beskikbaar is, en 'n empiriese studie is gedoen om Suid-Afrikaanse vryskutredigeerders se behoeftes ten opsigte van Afrikaanse naslaanbronne te bepaal. Die webblaaie van die Suid-Afrikaanse Vertalersinstituut (SATI), die Professional Editors' Guild (PEG) en die Universiteit Stellenbosch Taalsentrum gee toegang tot heelwat beskikbare Afrikaanse stylgidse, terminologielyste, spesialiswoordeboeke vir spesifieke vakgebiede, asook gespesialiseerde woordeboeke wat handel oor die teoretiese aspekte en spesialisveld van teksredigering (Blom 2020: 7). Die probleem is dat hierdie bronne redelik wyd versprei is op die webblaaie van SAVI, PEG en die Universiteit Stellenbosch Taalsentrum en nie een van die bronne bevat al die aspekte wat 'n akademiese redigeerder benodig om 'n vinnige en akkurate redigeerdiens uit te voer nie (Blom 2020: 7). In Blom (2020) is daar tot die gevolgtrekking gekom dat daar 'n behoefte is vir 'n naslaanbron wat spesifiek gemik is op die redigeerders van Afrikaanse akademiese tekste. Die vryskutredigeerders het in die empiriese studie, as motivering vir 'n nuwe Afrikaanse naslaanbron, aangevoer dat indien al die inligting op een plek beskikbaar sal wees, dit 'n groter sin vir eenheid onder redigeerders sal skep en sal sorg vir opleiding van nuwelinge en konsekwentheid in akademiese tekste (Blom 2020: 13). Na afloop van hierdie ondersoek en terugvoer, is 'n projek begin om 'n omvattende aanlyn akademiese redigeerwoordeboek saam te stel, wat akademiese redigeerders se werk sal vergemaklik en konsekwentheid in die redigering van Afrikaanse akademiese tekste sal verseker (Blom 2020: 8). In die eerste gedeelte van die projek is 'n model vir 'n aanlyn Afrikaanse akademiese redigeerwoordeboek (Blom se 2018-woordeboekmodel²) opgestel sodat dit in verdere studies as raamwerk vir die beplanning en samestelling van die akademiese redigeerwoordeboek gebruik kan word (Blom 2020: 8).

Die kernidees in Fuertes-Olivera en Tarp (2014) se funksieteorie vir gespesialiseerde aanlyn woordeboeke is gebruik om die akademiese redigeerders se gebruikersbehoefte, asook die redigeerwoordeboek se funksies en datatipes te bepaal, en Blom se 2018-woordeboekmodel op grond hiervan saam te stel. Daar is vasgestel dat die teikengebruikers van die akademiese redigeerwoordeboek (voortaan ARW) gevorderde akademiese redigeerders is, sowel as studente wat nog opleiding ontvang om as redigeerders te praktiseer, en beide partye is semi-deskundiges of leke ten opsigte van verskillende vakgebiede (Blom 2020: 19). Verder het hierdie akademiese redigeerders hoofsaaklik 'n behoefte aan tegniese aspekte, taal-, spel- en interpunksiereëls, vakterme van verskillende vakgebiede, afkortings en akronieme, 'n beskrywing van die redigeerder se rol, eiename wat gereeld in die akademie gebruik word en uitvoerige voorbeeldmateriaal wat van toepassing is op akademiese Afrikaans (Blom 2020: 19). Die gebruikssituasies en ooreenstemmende funksies van die ARW is dus kommunikatiewe, kognitiewe en operatiewe situasies, waarin die woordeboek vir teksproduksie, teksresepse, teksregstelling, ekstra inligting oor spesifieke vakgebiede en riglyne oor etiese redigering geraadpleeg sal word (Blom 2020: 20).

Die datatipes wat moontlik in die leksikografiese databasis vir die ARW ingesluit kan word, sluit onder andere in: grammatikale data, kollokasies of voorbeeldsinne wat die redigeerder kan help om 'n lemma in 'n akademiese teks te gebruik of verstaan (Blom 2020: 21). Verder kan leksikografiese en proskriptiewe notas vir die redigeerder meer duidelikheid gee oor die korrekte gebruik van 'n lemma in byvoorbeeld 'n spesifieke vakrigting, en kruisverwysings kan die redigeerders na ander inskrywings in die ARW, of 'n eksterne bron soos 'n ander aanlyn woordeboek lei (Blom 2020: 21). (Sien Blom 2020 vir 'n gedetailleerde uiteensetting van die datatipes in die leksikografiese databasis vir die ARW.)

Die volgende stap in die samestellingsproses van die ARW, is om Blom se 2018-woordeboekmodel te verfyn, ekstra data by te voeg en aan teikengebruikers bloot te stel, sodat die bruikbaarheid daarvan getoets kan word (Blom 2020: 25). As deel van 'n MA-studie aan die Universiteit Stellenbosch is 'n prototipe-akademiese redigeerwoordeboek (voortaan prototipe-ARW) opgestel sodat redigeerders die konsepontwerp van die ARW aanlyn kon toets (Blom 2021). Hierdie studie het uit drie hooftake bestaan (Blom 2021): Eerstens is Blom se 2018-woordeboekmodel verder saamgestel in 'n prototipe-ARW, deur besluite te neem oor die moontlike woordeboekstrukture en tegnologiese kenmerke wat die woordeboek moet vertoon. Tydens die tweede taak is die effektiwiteit van die voorlopige woordeboek bepaal deurdat akademiese redigeerders die prototipe-ARW moes gebruik om redigeertake te voltooi. Die derde taak was om die akademiese redigeerders se probleme of tekortkominge ten opsigte van die datavoorlegging en databeskrywing in die prototipe van dié aanlyn woordeboek vas te stel. In hierdie artikel word daar verslag gedoen van die eerste taak, naamlik die opstel van 'n prototipe-ARW. Die funksieteorie (Fuentes-Olivera en Tarp 2014), Gouws (2014a; 2014b; 2018a; 2018b; 2018c; 2018d) se aanpassings van die algemene leksikografiëteorie en bruikbaarheidsteorieë (ISO-standaard 9241 (1998a)(1998b)(2006)) word geïntegreer om besluite oor die databeskrywing/-inhoud en datavoorlegging/-uitleg van die prototipe woordeboek te maak. Die bruikbaarheid van die prototipe-ARW se databeskrywing en datavoorlegging, en verbeteringsriglyne vir die verdere samestelling van die akademiese redigeerwoordeboek word in 'n opvolgartikel behandel.

2. Navorsingsdoelwit en rasionaal

Die navorsingsdoelwit van hierdie studie is om 'n prototipe-aanlyn Afrikaanse akademiese redigeerwoordeboek saam te stel, nadat Blom se 2018-woordeboekmodel aangepas is en die funksieteorie (Fuentes-Olivera en Tarp 2014), algemene leksikografiëteorie (Gouws 2014a; 2014b; 2018a; 2018b; 2018c; 2018d) en bruikbaarheidsteorieë (ISO-standaard 9241 (1998a)(1998b)(2006)) geïntegreer is om besluite oor die databeskrywing/-inhoud en data-aanbieding/-uitleg van die prototipe-ARW te neem.

'n Geïntegreerde teorie (waar die teoretiese beginsels van die ISO-standaard geïntegreer is met die leksikografiëteorieë), is gebruik aangesien die

funksie- en algemene leksikografieteorie, soos uitgelig in Du Plessis (2015: 2017), nie aandag gee aan die tegnologiese eienskappe en bruikbaarheid van elektroniese naslaanbronne nie, en die navorser die akademiese redigeerders se ervaring met die tegnologiese aspekte van die ARW wou verbeter. Die beginsels van 'n bruikbaarheidsbenadering is juis nuttig, aangesien hierdie benadering gebruikerservaring, produkdoeltreffendheid en produk-mens-interaksie gebruik om grootliks op die teikengebruiker (bv. akademiese redigeerder) te fokus, asook hoe effektief, doeltreffend en bevredigend hy/sy met die prototipe-ARW omgaan (Du Plessis 2017: 4).

3. Die prototipe-akademiese redigeerwoordeboek (ARW)

Die prototipe-ARW is op grond van leksikografie- en bruikbaarheidsteorieë ontwerp en saamgestel om die inligting vir die ARW se databasis, asook die aanbieding van die data in ooreenstemming met die akademiese redigeerders se leksikografiese en tegnologiese behoeftes te bepaal.

Eerstens is Fiertes-Olivera en Tarp (2014) se funksieteorie vir gespesialiseerde aanlyn woordeboeke gebruik om 'n verduideliking van die teikengebruiker, leksikografiese funksies en datatipes van die prototipe te verskaf. Die ARW moet ontwerp en saamgestel word om inligting te bied vir redigeerders in die praktyk asook studente-redigeerders, wat onder andere Afrikaanse akademiese tekste soos tesisse, proefskrifte en wetenskaplike artikels redigeer (Blom 2021: 82). Die ARW moet, soos vasgestel in Blom (2020: 25), die rol van 'n stylgids en normeringsbron aanneem en die volgende kwessies behandel om aan die akademiese redigeerders se reeds genoemde gebruikersbehoeftes in kommunikatiewe, kognitiewe en operatiewe situasies te voldoen: tegniese aspekte in akademiese tekste, taal-, spel- en interpunksiereëls soos die korrekte woordgebruik, skryf- of spelwyse wat die akademiese redigeerder nuttig sal vind, asook ander inligting oor taalgebruik in bepaalde vakgebiede. Verder moet die spesialisveld van akademiese redigering behandel word, 'n afdeling vir vakterminologie ingesluit word, asook skakels na ander reeds bestaande algemene en beperkte woordeboeke (insluitend vakwoordeboeke) verskaf word (Blom 2020: 25). Vervolgens is die inhoud van die prototipe-ARW in ooreenstemming met die teikengebruikers se behoeftes en die prototipe se funksies (kommunikatief, kognitief en operatief) gekies en in 16 verskillende afdelings verdeel.

3.1 Databeskrywing of inhoud in die prototipe-ARW

Die tegniese aspekte in akademiese tekste (insluitend bronverwysings) en die taal-, spel- en interpunksiereëls word onder andere in die afdelings "Afkortings en akronieme", "Akademiese verwysingstelsels", "Getalle, syfers en simbole", "Hoof- en kleinletters", "Interpunksie", "Kursivering en Romeinse syfers", "Spelling en skryfwyse — los en vas", "Tabelle en grafieke" en "Wiskundige notasie" behandel om die akademiese redigeerders met kopie-, stilistiese en strukturele

probleme in akademiese tekste te help. Verder bied die afdelings "Algemeenste foute in akademiese tekste" en "Spesialisveld van akademiese redigering" riglyne oor die hantering van verskillende redigeerkwessies, asook hulp met die gebruik van die "track changes"-funksie op die Microsoft Word-program. Die afdelings "Algemene woordeboeke", "Vakterminologie", "Vakwoordeboeke", "Wetenskap, wiskunde en rekenaars" en "Wette en verwysings na wette" is laastens nodig om redigeerprobleme in spesifieke vakgebiede te op te los, en in hierdie verband bied die afdelings oor "Algemene woordeboeke" en "Vakwoordeboeke" ook skakels na eksterne bronne wat die akademiese redigeerders kan raadpleeg om meer inligting oor die vakgebiede te verkry. Die inhoud van hierdie 16 verskillende afdelings word in tabel 1 weergegee:

Tabel 1: 'n Uiteensetting van data wat in elke afdeling van die prototipe-ARW voorkom (aangepas uit Blom, 2021: 97-99)

Afkortings en akronieme

Hierdie afdeling bevat 'n lys afkortings en akronieme wat tipies in akademiese tekste voorkom. Elke afkorting/akroniem word onder die "antwoord"-ikoon volledig uitgeskryf en daar word in sommige gevalle ook 'n omskrywing van die afkorting/akroniem gegee. Ekstra inligting oor byvoorbeeld die herkoms van die afkorting/akroniem word ook onder die "nota"-ikoon gegee. Die "voorbeeld"-ikoon gee voorbeeldsinne waarin die afkorting/akroniem in akademiese tekste gebruik word. Die redigeerder kan hier sien hoe om die afkorting/akroniem korrek in 'n akademiese konteks te gebruik.

Akademiese verwysingstelsels

Die basiese kenmerke van verskillende verwysingsisteme wat in akademia gebruik word (bv. die Harvard-stelsel) word in hierdie afdeling uiteengesit, insluitend die wyse waarop daar na verskillende soorte bronne, soos boeke, akademiese tydskrifte en internetbronne inteks en in die bronnelys verwys moet word.

Algemeenste foute in akademiese tekste

Hierdie afdeling bevat 'n lys foute wat akademiese redigeerders sal help om foute wat algemeen in akademiese tekste gemaak word, vinnig te kan opspoor. Hierdie afdeling sal ook nuwelingsredigeerders help om vertrouwd te raak met tipiese probleme wat in die praktyk voorkom.

Algemene woordeboeke

'n Lys skakels na algemene woordeboeke word in hierdie afdeling gegee. Dit sluit onder andere die *Elektroniese Aanlyn WAT, HAT, en Longman Aanlyn en Pharos Aanlyn in*. Daar word ook skakels na woordeboeke verskaf wat redigeerders aangedui het hulle in ander tale buiten Afrikaans benodig, byvoorbeeld *Deutsch-Englisch-Wörterbuch* (Duits/Engelse woordeboek) en *Van Dale* (Nederlandse woordeboek).

Getalle, syfers en simbole

In hierdie afdeling word daar onder andere gefokus op die korrekte formaat waarin die chemiese binding van 'n element aangedui moet word. Daar is ook voorbeeldsinne ter illustrasie daarvan. Daar word verduidelik hoe om super-/subskrif en die en-streep in 'n Word-dokument aan te bring. Verder word die korrekte manier gegee waarop grade Celsius (°C) en persentasies geskryf word.

Hoof- en kleinletters

Algemene beginsels oor die gebruik van hoof- en kleinletters word puntsgewys in hierdie afdeling uiteengesit. Daar word ook 'n skakel na hoofstuk 9 van die *AWS* gegee, waar daar meer uitgebreide verduidelikings van die gebruik van hoof- en kleinletters is.

Interpunksie

Hierdie afdeling bevat beginsels oor die gebruik van lees- en skryftekens, insluitend die aandagstreep, aanhalingstekens, komma en koppelteken. Daar word ook skakels na hoofstuk 12 en 13 van die *AWS* gegee, waar daar meer uitgebreide verduidelikings oor die gebruik van die koppelteken en leestekens is.

Kursivering en Romeinse syfers

Die algemene norme vir die gebruik van kursivering en Romeinse syfers word in hierdie afdeling uiteengesit. Die hoof funksie is onder andere beklemtoning, woorde of frases van vreemde herkoms, titels, asook ander gebruike.

Spelling en skryfwyse — los en vas

Die algemene norme oor los en vas skryf word in hierdie afdeling uiteengesit. Daar word ook 'n spellys gegee van woorde wat gereeld verkeerd gebruik word in akademiese tekste.

Spesialisveld van akademiese redigering

Die vier tipes redigering, naamlik kopie-, stilistiese, strukturele en inhoudelike redigering, soos verduidelik deur Mossop (2014), word in hierdie afdeling uiteengesit. Daar word ook verduidelik wat presies redigeerders van akademiese tekste tydens die redigeerproses mag doen. (In hierdie afdeling word 'n tabel gegee wat aandui watter aspekte die redigeerder in elke kategorie behoort te redigeer.)

'n Onderskeid tussen elektroniese redigering en redigering op gedrukte materiaal word getref en daar word ook 'n tabel gegee van proefleestekens, asook die stappe wat gevolg moet word om die "track changes"-funksie in die MS Word-program te gebruik.

Tabelle en grafieke

In hierdie afdeling word daar meer op die tegniese aspekte van 'n akademiese teks gefokus. Daar word met behulp van voorbeeldmateriaal gedemonstreer of tabelle en grafieke op- of onder-skrifte kry, en die formaat van elk word beskryf.

Vakterminologie

Hierdie afdeling bestaan uit 'n lys vakterme. Elke term word beskryf en dan word bykomende inligting en ander bronne (wat meestal 'n skakel na 'n video-demonstrasie van die term is) gegee.

Vakwoordeboeke

Hierdie afdeling bestaan uit aanlyn hulpbronne, soos die Professional Editors' Guild, Prolingua, die Suid-Afrikaanse Akademie vir Wetenskap en Kuns en VivA, wat skakels bied na vakwoordeboeke van verskillende vakrigtings.

Wetenskap, wiskunde en rekenaars

In hierdie afdeling word aspekte wat belangrik is vir akademiese skryfwerk in die wetenskap behandel. Daar word byvoorbeeld verduidelik hoe gene geskryf moet word, asook wat presies die periodieke tabel van elemente is. Daar is ook 'n skakel na 'n interaktiewe weergawe van die periodieke tabel in Afrikaans.

Wette en verwysings na wette

Hierdie afdeling verduidelik hoe kursivering en hoofletters in akademiese werk in die regsvelde gebruik moet word. Daar word ook 'n lys afkortings gegee wat algemeen in die regsvelde gebruik word, asook 'n lys met skakels na Suid-Afrikaanse wette. Verder word daar verduidelik hoe daar na sake en wetgewing in die hoofteks en in voetnote verwys moet word. Daar word as ekstra bronne 'n lys aanlyn hulpbronne gegee met met skakels na styl- en skryfgidse, asook verskeie webwerwe wat nuttig is vir die regsdisipline.

Wiskundige notasie

Hierdie afdeling bevat 'n lys van basiese wiskundesimbole wat in Wiskunde, die Fisiese Wetenskappe, Ingenieurswese en Ekonomie gebruik word.

Ná die beginsels van Fuentres-Olivera en Tarp (2014) se funksieteorie in Blom (2020; 2021) toegepas is om die inhoud van die ARW te bepaal, is die beginsels van die algemene leksikografiëorie, wat hoofsaaklik aandag gee aan die woordeboekstrukture vir die aanlyn medium, asook die beginsels van die bruikbaarheidsbenadering, ondersoek. Die aspekte wat van toepassing is op aanlyn woordeboeke is soos volg in afdeling 3.2 op die prototipe-ARW toegepas om die aanbieding van die data te bepaal.

3.2 Data-aanbieding of uitleg van die prototipe-ARW

3.2.1 Woordeboekstrukture vir die prototipe-ARW

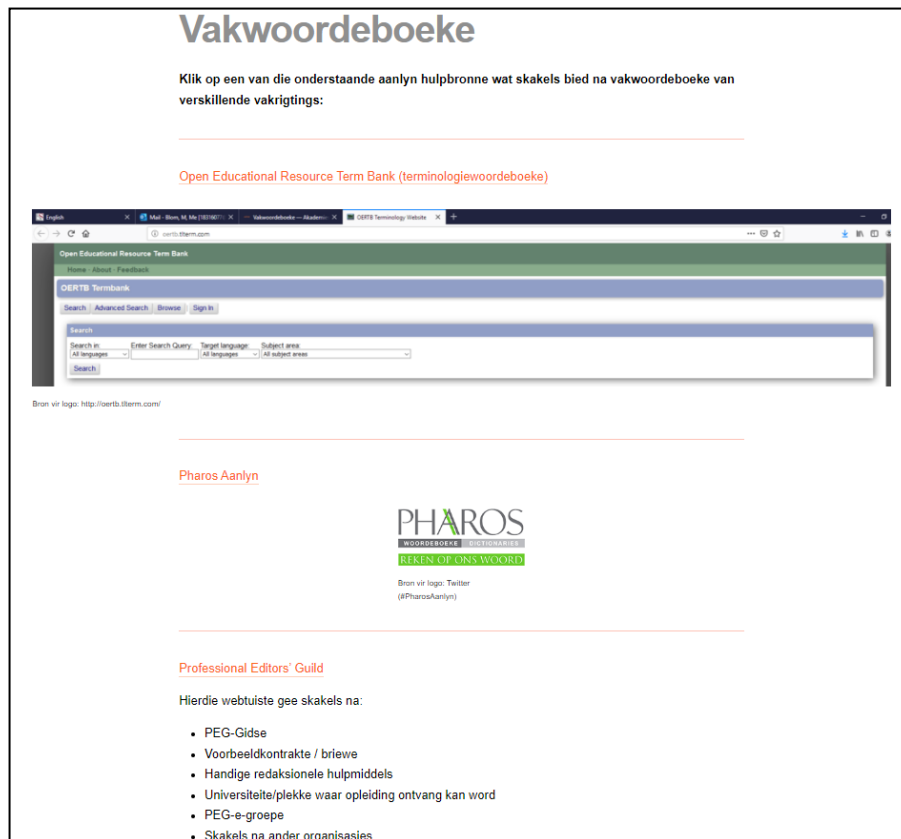
Die strukture wat van toepassing is op die aanbiedingsvlak van die prototipe-ARW, insluitend die makro-, artikel-, mikro-, medio-, toegang- en skermkootstruktuur, word meer noukeurig in hierdie afdeling bespreek en met behulp van skermgrepe geïllustreer. Voordat hierdie strukturele eenhede in die prototipe-ARW saamgestel is, moes die dataverspreiding van die datatipes eers uitgevoer word. Die dataverspreidingsstruktuur word dus eerste bespreek, aangesien dit die basiese leksikografiese struktuur is wat ander strukture bepaal en moontlik maak (Tarp 2015: 223).

Dataverspreiding

Die verspreiding van leksikografiese data in die ARW is 'n sirkulêre proses wat van nuuts af begin ná elke versoek van die akademiese redigeerder, en plaasvind vanaf die databasisvlak (met ander woorde vanaf die leksikografiese databasis) na die aanbiedingsvlak, oftewel gebruikerskoppelvlak (Blom 2021: 49).

Eerstens kan die verspreiding van data in die ARW tot die posisies in die artikels of buitekomponente plaasvind. Indien die akademiese redigeerder byvoorbeeld 'n behoefte aan die tegniese aspekte in akademiese tekste; taal-, spel- en interpunksiereëls of vakterminologie het, kan hy/sy die term of frase in die soekblok intik, of op die relevante afdeling op die tuisbladsy klik. Die versoek word dan na die databasis gestuur en deur middel van tegniese prosesse³ in die ARW se databasis word daar 'n verbinding gemaak dat 'n spesifieke woordeboekartikel waarin die term of inligting voorkom op die redigeerder se rekenaarskerm moet verskyn. In die betrokke woordeboekartikel kan die redigeerder dan verder op die "voorbeeld"-, "nota"-, of "ekstra bronne"-ikoon klik indien hy/sy 'n behoefte het aan voorbeeldsinne in die akademiese konteks, ekstra inligting oor die kwessie of vakterm benodig, of skakels na eksterne webblaaie met meer inligting of demonstrasies oor die kwessie wil volg. Hierdie internetskakels vul dus die sentrale teks van die ARW met buitekomponente aan.

Naas die woordeboekartikels en buitekomponente, is daar ook 'n woordeboekportaal, naamlik 'n woordeboekversameling (soos gedefinieer in Engelberg en Müller-Spitzer 2013: 1027) in die prototipe-ARW aanwesig, waar akademiese redigeerders skakels na ander aanlyn woordeboeke kry. Die afdeling "Vakwoordeboeke" gee vir die redigeerders toegang tot onder andere aanlyn bronne of webtuistes soos *Open Educational Resource Term Bank*, *Pharos Aanlyn*, *Professional Editors Guild*, ensovoorts. Sien skermgreep 1 vir 'n voorbeeld van die prototipe-ARW se woordeboekversameling. Volgens Gouws (2018c: 185) vereis 'n woordeboekportaal dan 'n omvattende dataverspreidingsstruktuur waar daar voorsiening gemaak word vir die tradisionele dataverspreidingsstruktuur van die individuele aanlyn woordeboek (byvoorbeeld die ARW), asook vir die verspreiding van die data in die woordeboekportaal (Gouws 2018c: 185).



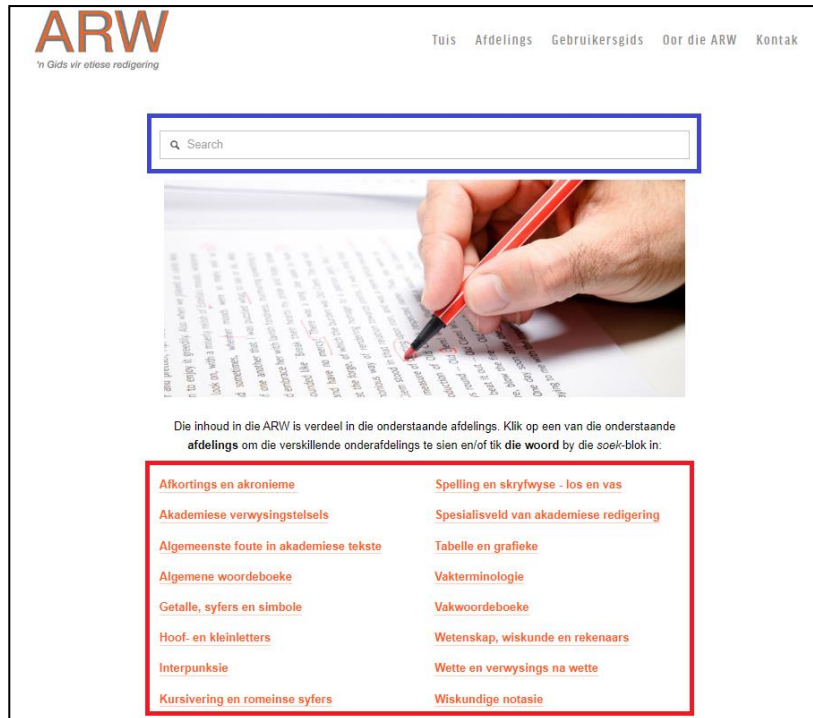
Skermgreep 1: Woordeboekversameling in die prototipe-ARW se afdeling "Vakwoordeboeke"

Makrostrukturele eenhede

Die ARW het 'n sentrale tuisblad en soekenjin wat 'n kitssoekproses moontlik maak, en daarom verval die behoefte om die makrostrukturele ordening te handhaaf wat in gedrukte woordeboeke gebruik word. Volgens Gouws (2018c: 182) bly die makrostruktuur 'n ordeningstruktuur in aanlyn woordeboeke, en die ordeningstruktuur in die prototipe-ARW vertoon die volgende makrostrukturele kenmerke:

'n Tematiese ordening word gebruik om die inhoud in die prototipe-ARW volgens afdelings te organiseer (sien die rooi blok in skermgreep 2). Indien die redigeerder op een van hierdie afdelings op die tuisblad klik, verskyn die lemmas in die spesifieke afdeling in alfabetiese volgorde. Verder kan die soekwoord in die soekblok op die openingskerm van die prototipe-ARW ingevul word (sien die blou blok in skermgreep 2). 'n Subtematiese ordening sal dan verskyn wat die soekwoord as komponent binne die spesifieke afdelings in die

prototipe-ARW bevat. Die redigeerder kan net op die verlangde afdeling klik om toegang te kry tot die inligting wat betrekking het op die soekwoord in die afdeling (sien skermgreep 3 vir 'n illustrasie van die afdelings, in rooi gemerk, waarin die soekwoord "kursivering" verskyn).



Skermgreep 2: Prototipe-ARW se tuisblad met soekenjin en verskillende afdelings



Skermgreep 3: Soekblok uit die prototipe-ARW

Artikel- en mikrostrukturele eenhede

Die artikelstruktuur in die prototipe-ARW bied, soos Gouws (2014a: 164) voorstel, 'n dinamiese data-aanbod wat uit veelvoudige lae bestaan, en drie strukturele komponente, naamlik teksgedeeltes, kommentaartipes en soeksones (Gouws 2014a: 160) bevat. Die **teksgedeeltes** in die prototipe-ARW se woordeboekartikels kan verder tussen aanduiders en struktuurmerkers onderskei word (Gouws 2014a: 161). Die aanduiders in die ARW is datadraers waaruit die akademiese redigeerder spesifieke inligting oor die woordeboekonderwerp kan onttrek. Die "antwoord" vir die akroniem PKR in skermgreep 4 is byvoorbeeld 'n aanduiding, aangesien die redigeerder inligting oor die akroniem uit hierdie inskrywing kan onttrek.

ARW
'n Gids vir etiese redigering

Tuis Afdelings Gebruikersgids Oor die ARW Kontak

Afkortings en Akronieme

PKR

Hoe word die akroniem in Afrikaans geskryf?
Wat is PKR?

+ Antwoord

Die akroniem PKR, staan vir **polimerasekettingreaksie** en verwys na 'n laboratoriumtegniek wat gebruik word om miljoene kopieë van 'n spesifieke teikenarea van DNA te maak. Die teikenarea kan byvoorbeeld 'n geen wees waarvan die navorser die funksie wil weet. Sien voorbeeld. Sien ook DNA/DNS voorbeeld 2.

Bron: Khan Academy. 2019. Polymerase chain reaction (PCR). [Intyds]. Besikbaar: <https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/polymerase-chain-reaction-pcr0> [13, Mei 2019].

+ Voorbeeld

"Die metode is gebaseer op **polimerasekettingreaksie (PKR)** wat akkurate en onbevangende opsporing van virusse binne ure moontlik maak."

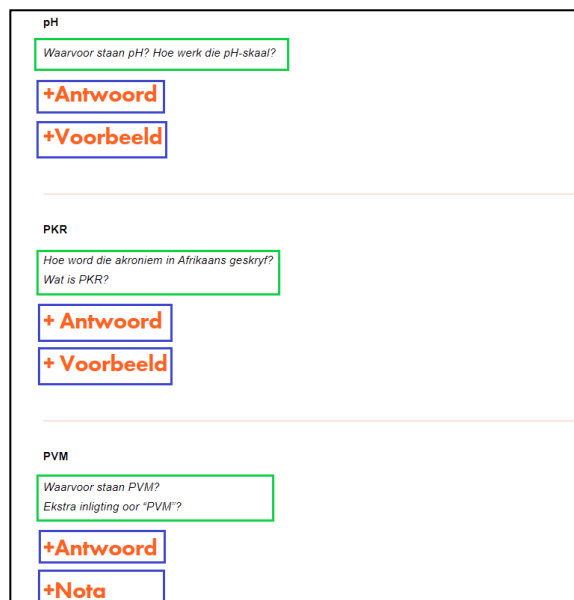
Bron: InnovUS. 2017. Innovus-Lisensieooreenkoms tot voordeel van plaaslike wingerdbedryf. [Intyds]. Besikbaar: <https://www.innovus.co.za/afrikaans-archived-news/afrikaans-innovus-licence-agreement-to-benefit-local-grapevine-industry.html1> [12, Mei 2019].

Skermgreep 4: Artikel van "PKR" uit die afdeling "Afkortings en akronieme" in die prototipe-ARW

Die struktuurmerkers identifiseer sekere aanduiders of artikelgleuwe (Gouws 2014a: 161), en die struktuurmerkers in die prototipe-ARW bestaan hoofsaaklik uit tipografiese struktuurmerkers soos kursivering, vetdruk en letterkleur om sekere aanduiders in die prototipe uit te lig (soos geïllustreer in die groen, rooi en blou blokke in skermgreep 4). Die enigste nietipografiese struktuurmerker in die prototipe-ARW is die "sien meer"-ikoon (+). Hierdie struktuurmerker is eie aan die elektroniese sfeer en gebruikers wat vertrou is met die rekenaar- en webruimte sal weet dat die inligting vir 'n spesifieke datakategorie sal verskyn wanneer hulle daarop klik (Blom 2021: 56).

Verder word die **kommentaartipes** in die prototipe-ARW gebruik om verskillende toegangsroetes tot verskillende leksikografiese datatipes te verskaf. Die opskrifte wat elke datatipe bekendstel, byvoorbeeld "Antwoord" en "Voorbeeld" (in blou gemerk in skermgreep 4 en 5), en die vroe onder die lemma (in groen gemerk in skermgreep 4 en 5) help met eksplisiete en vinnige toegang tot die relevante datatipes, en vervang dus die tradisionele vorm- en semantiese kommentaar wat in gedrukte woordeboeke voorkom.

Die derde kategorie van strukturele komponente wat Gouws (2014a: 169) onderskei, is **soeksones**, wat die onderafdelings van woordeboekartikels is (elke datatipe word in 'n afsonderlike soeksone aangebied). Die datatipe in elke afsonderlike soeksone van die prototipe-ARW word aangebied deur middel van data-identifiserende inskrywings soos "Antwoord", "Nota", "Voorbeeld", ens. (in blou gemerk in skermgreep 5), en dit gee 'n duidelike afbakening van die gleuwe vir die verskillende datatipes.



Skermgreep 5: Artikeldeeltrajekte uit die afdeling "Afkortings en akronieme" in die prototipe-ARW

Gouws (2014a: 175) meen dat verskillende soorte mikrostrukture oorweeg moet word om die beste struktuur te kry vir die ordening van die aanduiders in die spesifieke woordeboekartikels. In die prototipe-ARW word 'n geïntegreerde mikrostruktuur gebruik om die verspreiding van data in die woordeboekartikels te reguleer, wat beteken dat die betekenisparafrase (die "antwoord" in skermgreep 4) en die koteksinskrywing (die voorbeeldsin in skermgreep 4) in 'n enkel soekveld van die artikel "PKR" verskyn. Volgens Gouws en Prinsloo (2005: 139) veroorsaak hierdie nabyheid tussen die betekenisparafrase en koteksinskrywing dat die gebruiker vinnig inligting oor die lemma, soos aangebied in die betekenisparafrase en gepaardgaande koteksinskrywing, kan onttrek. Voordat die redigeerder hierdie inligting kan onttrek moet hy/sy eers 'n bepaalde soekroete volg, en moet daar vervolgens gepaste toegangstrukture aangewend word om optimale toegang tot die data in die ARW te verseker.

Toegangstruktuur

In die prototipe-ARW word daar onderskei tussen eksterne en interne toegangstrukture, wat lei tot eksterne en interne soekroetes om by spesifieke inskrywings in die prototipe-ARW en buitetekste uit te kom. Die prototipe-ARW steun hoofsaaklik op die woordeboek se soekenjin as eksterne toegangstruktuur, en hierdie toegangstruktuur bied hoofsaaklik twee eksterne soekroetes na 'n spesifieke woordeboekartikel. Die redigeerder kan eerstens 'n woord/frase in die soekblok op die tuisbladsy (sien skermgreep 2) intik, waarna 'n lys resultate met die bewerking van die soekopsie in verskillende afdelings van die prototipe-ARW verskyn (sien skermgreep 3). Die redigeerder kan dan op 'n spesifieke afdeling klik om hom/haar na die verlangde woordeboekartikel te lei. Vervolgens kan die redigeerder ook op een van die afdelings op die tuisbladsy klik (sien skermgreep 2) en deur 'n alfabetiese lys lemmas vir daardie spesifieke afdeling blaai om by die verlangde woord uit te kom. Die akademiese redigeerder wat vertrou is met die ARW sal, soos Gouws (2018d:45) noem, sommige fases in die eksterne toegangproses kan oorslaan, en dadelik op die relevante afdeling klik, terwyl die redigeerder wat nie die ARW ken nie, deur al die stappe in die eerste eksterne soekroete moet gaan voordat hy/sy by die verlangde artikel uitkom.

Wiegand en Gouws (2013: 113) definieer die soekroete na buitetekste ook as deel van die eksterne toegangstruktuur. In die prototipe-ARW word verskeie buitetekssoorte gebruik, byvoorbeeld 'n eksterne skakel (sien die rooi blok in skermgreep 6) na 'n Google-soektog oor hoe om 'n kommentaarblokkie in 'n Microsoft Word-dokument te voeg.

WELKOM BY DIE AKADEMIESE REDIGEERWOORDEBOEK

*Die ARW is ontwerp om redigeerders van Afrikaanse akademiese tekste te help om kopie-,
strukturele, inhoudelike en stilistiese redigering op 'n etiese wyse aan te bring.*

Dit is baie belangrik dat u op die onderstaande **etiese-blok** klik en seker maak of 'n redigeerkwessie
met of sonder 'n **kommentaarblokkie** hanteer moet word.

Skermgreep 6: Eksterne skakel na 'n Google-soektog op die tuisbladsy van die prototipe-ARW

Daar word ook in die prototipe-ARW se afdeling "Vakwoordeboeke" toegang tot eksterne naslaanbronne en webwerwe gegee, deur skakels en logo's van die betrokke webwerwe te verskaf (sien voorbeelde van skakels en logo's in skermgreep 1).

Die interne toegangstruktuur bepaal hierna die interne soekroete wat die redigeerder binne die relevante artikel oriënteer om by 'n bepaalde aanduiding of datakategorie uit te kom (Gouws en Prinsloo 2005: 170; Wiegand en Gouws 2013: 139). Hier help die tipografiese en nietipografiese struktuurmerkers, soos Gouws en Prinsloo (2005: 170) voorstel, om vinnige toegang te verseker, en sommige stappe in die interne soekroete oor te slaan, sodat die verlangde soeksone vinniger bereik word. In die onderstaande skermgreep 7 is die vraag onder die lemma (omkring in blou) in kursief om die redigeerders se aandag daarop te vestig en vir hulle 'n idee te gee van die inligting wat hulle kan verwag indien hulle op die antwoord- of voorbeeld-ikoon klik. Verder is die sleutelaspekte onderaan die artikel (omkring in rooi) in vetdruk om die redigeerders se aandag dadelik op hierdie aspekte te vestig wanneer hulle op die data-identifiserende inskrywing "Voorbeeld" klik. Die data-identifiserende inskrywings, "Antwoord", "Nota" en "Voorbeeld" is in oranje, sodat die redigeerders duidelik tussen die verskillende datakategorieë kan onderskei. Die data-identifiserende inskrywings dien ook elkeen as aanwysers in die prototipe-ARW se onderskeie woordeboekartikels aangesien hulle die redigeerders na die relevante soeksone lei en 'n eksplisiete identifikasie gee van die tipe data wat in die betrokke soeksone opgeneem word (Blom 2021: 69).

ARW
In Gids vir etiese redigering

Tuis Afdelings Gebruikersgids Oor die ARW Kontak

Getalle, syfers & simbole

Chemiese bindings

Wat is chemiese bindings?
Wanneer word subskrifte en Romeinse syfers in chemiese formules gebruik?

+ Antwoord

+ Nota

+ Voorbeeld

1. "Termodinamiese oorwegings stel 'n streng raamwerk voor vir die interpretasie van chemiese reaksies, alhoewel daar egter min eksperimentele data openlik beskikbaar is vir die gepaardgaande oplossing spesies [sic] in suur **yster(III)** sulfaat stelsels [sic]."
2. "Enkellading en kontinue **yster(III)** verminderingskinetika is gemeet en die gevolge van die aanvanklike **Fe₂(SO₄)₃- en H₂SO₄-konsentrasies**, temperatuur en in-situ neutralisasie [sic] is gekwantifiseer."

Skermgreep 7: Artikel van "Chemiese bindings" uit die afdeling "Getalle, syfers en simbole" in die prototipe-ARW

Mediostrukturele eenhede

Die medio- of kruisverwysingstruktuur word in die prototipe-ARW gebruik om, soos Gouws en Prinsloo (2005: 177) voorstel, die data toegankliker vir die akademiese redigeerders te maak en 'n verhouding tussen die woordeboek-komponente in die prototipe-ARW te vestig. Die kruisverwysings in die prototipe-ARW neem die redigeerders meestal direk deur middel van hiperskakels (in swart en oranje) na die interne of eksterne kruisverwysingsadres. In skermgreep 8 neem die kruisverwysingsinskrywing, "Sien ekstra bronne" (omkring in rooi), die redigeerder vanaf die kruisverwysingsposisie, "Antwoord", na die interne kruisverwysingsadres, "Ekstra bronne". Die redigeerder word ook deur middel van die kruisverwysingsinskrywing, "Sien ook Wetenskap, wiskunde en rekenaars" (omkring in blou), vanaf die kruisverwysingsposisie, "Antwoord", na die eksterne kruisverwysingsadres, naamlik 'n ander afdeling in die woordeboek (in hierdie geval "Wetenskap, wiskunde en rekenaars") geneem. Die kruisverwysingsinskrywings, "skakel" en "lys" (omkring in groen), neem die redigeerder na woordeboekeksterne kruisverwysingsadresse, wat in hierdie geval 'n webwerf met 'n interaktiewe weergawe van die periodieke tabel en 'n meertalige lys van die elemente is. Hierdie kruisverwysingstruktuur in die proto-

tipe-ARW maak dit dus moontlik om verskillende datatipes met mekaar te skakel, en die skakels na woordeboekeksterne webwerwe of naslaanbronne verwys die redigeerders na die groot naslaanruimte van die internet (Blom 2021: 72).

Periodieke tabel:

+ Antwoord

Die periodieke tabel van elemente is 'n **metode waarvolgens die chemiese elemente in 'n tabel gerangskik is volgens toenemende atoomgetalle**. Hierdie elemente is ook in periodes en groepe verdeel. 'n Groep is 'n vertikale kolom op die periodieke tabel wat elemente met soortgelyke eienskappe bevat en daar word na elke ry verwys as 'n periode. **Sien ekstra bronne** vir 'n skakel na 'n interaktiewe weergawe van die periodieke tabel.

Sien ook Wetenskap, wiskunde en rekenaars vir 'n meer in-diepte bespreking van die ontstaan en uitleg van die periodieke tabel.

Bron: Siyavula Education. 2019. Hoofstuk 5: Die periodieke tabel. In: *Everything science: graad 10 Fisiese Wetenskappe*. [Intyds]. Beskikbaar: <https://www.siyavula.com/read/science/grade-10/the-periodic-table/05-the-periodic-table-01> [2019, 16 Mei].

+ Ekstra bronne

Klik op hierdie **skakel** om 'n interaktiewe weergawe van die periodieke tabel in Afrikaans te kan sien met dinamiese uitleg van name, elektrone, skuifskale, visualisasietendense, orbitale en isotope. Sien hierdie **lys** wat 'n Afrikaans/Engelse vertaling van die elemente, asook elkeen se atoomgetal en simbool aandui.

Skermgreep 8: Artikel van "Periodieke tabel" uit die afdeling "Vakterminologie" in die prototipe-ARW

Skermkootstruktuur

Die prototipe-ARW is saamgestel om hoofsaaklik op 'n rekenaarskerm te verskyn en die skermkootstruktuur bied ruimte op die rekenaarskerm aan dele van die struktuurtypes wat reeds bespreek is, naamlik die makro-, mikro-, artikel- en toegangstruktuur (Blom 2021: 73). Hierdie struktuurtypes is, soos geïllustreer in skermgrepe 1 tot 8, as volg georganiseer dat die redigeerders nie oorlaai word met data of deur te veel vlakke moet klik voordat hulle by die verlangde data uitkom nie.

Die 16 verskillende afdelings word as deel van die makrostruktuur in twee blokke op die prototipe-ARW se tuisblad gerangskik sodat die akademiese redigeerder al die afdelings in een skermgreep kan sien en nie deur 'n lys hoef te blaai nie. As deel van die mikro-, artikel- en toegangstruktuur is die lettertipe eenvoudig genoeg sodat die redigeerder vinnig deur die inskrywings

kan lees, en help die oranje kleur wat vir die aanduiders en hiperskakels gebruik word, asook die tipografiese merkers (vraagsin in kursief, lemma en verkorte antwoord in vetdruk en data-identifiserende inskrywings in oranje) en nietipografiese merkers ("sien meer"-ikoon (+) en data-identifiserende inskrywings soos "Antwoord") die redigeerders om tussen die data-inskrywings op die skerm te onderskei. Die oorgang van een skermgreep na 'n ander is vinnig, en die redigeerder kan kies watter data in die artikel op die skerm moet verskyn deur eers op die (+)-ikoon te klik voordat die data vir 'n spesifieke inskrywing vertoon word. Soos reeds genoem, wil die navorser die akademiese redigeerders se ervaring met die tegnologiese aspekte van die ARW verder verbeter deur ook aandag aan bruikbaarheidsbeginsels te gee. In die volgende afdeling word die teoretiese beginsels van die bruikbaarheidsbenadering, wat op aanlyn woordeboeke van toepassing is, ondersoek en gebruik (in kombinasie met die funksie- en algemene leksikografieteorie) om die samestelling van die prototipe-ARW uiteen te sit.

3.2.2 Bruikbaarheidsbeginsels

Die bruikbaarheid van die prototipe-ARW verwys, soos Du Plessis (2015: 81) noem, na die effektiwiteit waarmee die akademiese redigeerder inligting kan waarneem en onttrek om spesifieke redigeertake te voltooi, en dit is belangrik dat die ARW se bruikbaarheid op sekere regulasies en riglyne gebaseer is (Du Plessis 2017: 4). Du Plessis (2015: 80) het drie hoofstukke uit die ISO-standaard 9241 (sien die ISO-standaard 9241-110:2020 (2006); die ISO-standaard 9241-11:2018 (1998a) en die ISO-standaard 9241-12:2017 (1998b)⁴), wat regulasies en beginsels vir die ontwerp van gebruikersgerigte rekenaartoestelle en -programmatuur bevat, in een oorkoepelende stel beginsels vir mens-rekenaar-interaksie geïntegreer. Die onderstaande stel beginsels van Du Plessis (2015; 2017), soos uiteengesit in Blom (2021: 75), is ook in ag geneem tydens die samestelling van die prototipe-ARW.


1. *Taakgepastheid*: Die gebruikerskoppelvlak van die ARW moet geskik wees om 'n reeks take (kommunikatiewe, kognitiewe en operatiewe take) te prosesseer en die data op so 'n manier aan te bied dat die akademiese redigeerder dit kan interpreteer.
2. *Selfbeskrywings*: Die gebruikerskoppelvlak van die ARW moet vir die akademiese redigeerder duidelike terugvoer oor die program kan lewer, byvoorbeeld terugvoer oor onsuksesvolle soektogte.
3. *Duidelikheid en netheid*: Die data in die ARW moet in gepaste kleure, lettertipes en lettergroottes verskyn, sodat die akademiese redigeerders dit maklik kan raaksien en lees.

4. *Beheerbaarheid en diskrimineerbaarheid*: Die akademiese redigeerders moet in beheer van die interaktiewe elemente van die ARW wees. 'n Hoofspyskaart kan byvoorbeeld help om na bepaalde woordeboekdele te beweeg.
5. *Ooreenstemming met gebruikersverwagtinge (konsekwentheid)*: Die aanbieding van die data in die ARW moet ooreenstem met die akademiese redigeerders se verwagtinge van hoe data in aanlyn woordeboeke moet vertoon. Hierdie data moet ook konsekwent aangebied word deur deurgaans gebruik te maak van dieselfde artikelstruktuur en vertoningsformaat.
6. *Bondigheid en opspoorbaarheid*: 'n Woordeboekartikel in die ARW moet die akademiese redigeerder se aandag slegs na die relevante inligting trek, sodat die redigeerder nie oorlaai word met data of sukkel om die relevante data op te spoor nie.
7. *Foutkontrole*: Die ARW moet die akademiese redigeerders in die regte rigting lei wanneer hulle foute maak, byvoorbeeld wanneer 'n term verkeerd in die soekblok gespel is.
8. *Individugepastheid*: Die gebruikerskoppelvlak van die ARW moet gemanipuleer kan word sodat elke akademiese redigeerder sy/haar eie profiel kan skep.
9. *Leergepastheid*: Die ARW moet eenvoudig genoeg wees dat alle akademiese redigeerders dit kan verstaan. Hulle moet ook in staat wees om byvoorbeeld vorige soektogte en gunstelingsoektogte te stoor.

Hierdie beginsels beklemtoon die aanbieding van die data, die akademiese redigeerder se ervaring met die programmatuur, en hoe effektief, doeltreffend en tevrede die redigeerder met die ARW omgaan (Du Plessis 2017: 6). Die beginsels is as volg tydens die besluitneming oor die uitleg en aanbieding in die prototipe-ARW se tuisbladsy, woordeboekartikels en gebruikersgids toegepas:


Tuisbladsy

Die uitleg van die tuisbladsy is eerstens in ooreenstemming met die beginsel oor *duidelikheid en netheid* ontwerp, aangesien die wit agtergrond in skermgreep 9 dit vir die redigeerders makliker maak om die swart en oranje letters te lees, terwyl die hoof- en kleinletters, en vetdruk en kursief die belangrike data vir die gebruikers uitlig (Blom 2021: 83). Verder bevat die tuisbladsy in skermgreep 9 'n hoofspyskaart (omkring in blou) met skakels na "tuis", "afdelings", "gebruikersgids", "oor die ARW" en "kontak", asook interaktiewe blokke na verskillende gebruikersgidse in die prototipe-ARW (omkring in rooi) en 16 interaktiewe afdelings (omkring in groen), wat die redigeerders kan help met *beheerbaarheid en diskrimineerbaarheid* sodat hulle na bepaalde woordeboekdele van die prototipe-ARW kan beweeg.



Tuis Afdelings Gebruikersgids Oor die ARW Kontak

Soek



WELKOM BY DIE AKADEMIESE REDIGERWOORDEBOEK

Die ARW is ontwerp om redigeerders van Afrikaanse akademiese tekste te help om **kopie-, strukturele, inhoudelike en stilistiese redigering** op 'n etiese wyse aan te bring.

Dit is baie belangrik dat u op die onderstaande **etik-blok** klik en seker maak of 'n redigeerkwessie met of sonder 'n **kommentaarblok** hanteer moet word.

Die inhoud in die ARW is verdeel in verskillende afdelings. Klik op die onderstaande blokke vir 'n verduideliking van die woordeboek se **soekroetes, uitleg, ikone, inhoud** en **etik**:

SOEKROETES	IKONE
UITLEG	INHOUD
ETIEK	

Klik op een van die onderstaande **afdelings** om die verskillende onderafdelings te sien en/of tik die **woord** by die soek-blok, bo-aan die bladsy, in:

Afkortings en akronieme	Spelling en skryfwyse - los en vas
Akademiese verwysingstelsels	Spesialisveld van akademiese redigering
Algemeenste foute in akademiese tekste	Tabelle en grafieke
Algemene woordeboeke	Vakterminologie
Getalle, syfers en simbole	Vakwoordeboeke
Hoof- en kleinletters	Wetenskap, wiskunde en rekenaars
Interpunksie	Wette en verwysings na wette
Kursivering en Romeinse syfers	Wiskundige notasie

Skermgreep 9: Tuisbladsy in die prototipe-ARW

Woordeboekartikels

Die data-uitleg van die woordeboekartikels in die prototipe-ARW is ook volgens Du Plessis (2015; 2017) se beginsels saamgestel. Die *taakgepastheid* van die woordeboek word verseker, aangesien die gebruikerskoppelvlak van die prototipe-ARW geskik is vir die verwerking van kommunikatiewe, kognitiewe en operatiewe take (deur toegang te gee tot 'n "antwoord"/definisie, "notas", "voorbeelde" en "ekstra bronne" (in groen gemerk in skermgreep 10). Verder vertoon die woordeboekartikels *bondigheid en opspoorbaarheid* deur die data onder elke ikoon in skermgreep 10 eers te versteek, en indien die redigeerder byvoorbeeld besluit om op die voorbeeld-ikoon te klik, word die belangrikste data in vetdruk aangedui (soos geïllustreer in skermgreep 10) sodat die redigeerder nie oorlaai word met data nie (Blom 2021: 86). Die data in die prototipe-ARW se woordeboekartikels word ook *konsekwent* aangebied deur deurgaans gebruik te maak van dieselfde artikelstruktuur en vertoningsformaat.

Gene

Hoe word kursivering gebruik om tussen 'n geen en sy ooreenstemmende proteïen te onderskei?

+ Antwoord

+ Nota

+ Voorbeeld

1. **EPO-, GH1- en TNF-** mRNA-uitdrukking is geanaliseer deur gebruik te maak van RT-PCR. **VEGF-, TGFβ1- en P53-** proteïenuitdrukking is ondersoek deur gebruik te maak van immunohistochemie.
2. "Vloeisitometrie-analise is uitgevoer deur gebruik te maak van 'n angiogenese-merker, naamlik **vaskulêre-endoteelgroefaktor (VEGF)**, en die invloed van ESE-15-ol en ESE-16 op angiogenese is bestudeer. SEM- en vloeisitometrie-resultate het geen noemenswaardige verandering in VEGF-vlakke van die bloedselle wat aan ESE-15-ol en ESE-16 blootgestel is, getoon nie."

Bron: Mqoco, T.V, Repsold, L., Wolmarans, E., Nkandeu, S., Theron, A.E., Stander, B.A. 2014. Ex vivo effekte van estradiol analoë op bloed: 'n Loodsstudie. *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie*. 33(1).

+ Ekstra bronne

Skermgreep 10: Uitleg van die woordeboekartikel vir "gene" in die prototipe-ARW

Gebruikersgids

Die uitleg van die gebruikersgids oor soekroetes, ikone, inhoud, uitleg en etiek in die prototipe-ARW is saamgestel om met die *leergepastheid* van die ARW te

help en onder andere nuwe gebruikers van die woordeboek vinniger vertrouwd te maak met die verskillende soekroetes en ikone wat in die ARW voorkom. Die riglyne in die etiek-gebruikersgids (sien skermgreep 11) kan ook nuweling-redigeerders se vaardighede en kennis oor etiese redigering opskerp deurdat daar riglyne gegee word vir die soort veranderinge wat in verskillende situasies aangebring moet word, byvoorbeeld om die fout te korrigeer, of om eerder 'n kommentaarblokkie met 'n voorstel in te voeg.

Etiese redigering

Sien die onderstaande uiteensetting van watter veranderinge in 'n spesifieke situasie aangebring mag word:

Volgens Lourens (2014: 266) mag redigeerders binne die grense van etiese redigering, **nie groot veranderinge aan akademiese tekste** aanbring nie, veral nie indien hierdie veranderinge nie deur die redigeerder aangedui word nie.

*Indien daar onsekerheid oor enige redigeeraksie ontstaan, is die goue reël om eerder 'n **kommentaarblokkie** in te voeg sodat die skrywer steeds die finale besluit oor die moontlike wysiging het.*

AFKORTINGS EN AKRONIEME

- Indien 'n afkorting of akroniem verkeerd in 'n akademiese teks gespél is, kan die **spelfout gekorrigeer** word.
- Maak seker dat die afkorting of akroniem in die res van die teks korrek gespél is en indien die spelfout **op meer as een plek** voorkom, kan 'n **kommentaarblokkie** ingevoeg word om die skrywer se aandag hierop te vestig.
- Indien 'n **omskrywing (m.a.w inhoudelike fout)** van 'n afkorting of akroniem gegee word wat vermoedelik verkeerd is, **MOET 'n kommentaarblokkie** ingevoeg word.

AKADEMIESE VERWYSINGSTELSLS

- **Ontbrekende of verkeerde inligting** in verwysings kan **gekorrigeer** word.
- Indien die teks en bronnelys nagegaan moet word, maak seker dat die inteksverwysings en bronnelys ooreenstem. Maak hier gebruik van die "Ctrl-F" opsie om die verwysings vinniger op te spoor.

Skermgreep 11: 'n Gedeeltelike skermgreep van die etiek-gebruikersgids in die prototipe-ARW

4. Slot en opvolgstudie

In hierdie artikel is 'n uiteensetting gegee oor die eerste taak in Blom (2021) se studie oor "die samestelling van 'n prototipe-ARW met die doel om die bruikbaarheid daarvan te toets", deurdat hierdie samestelling op grond van teoretiese beginsels uit die leksikografie- en bruikbaarheidsveld gedemonstreer is. Fuertes-Olivera en Tarp (2014) se funksieteorie vir gespesialiseerde aanlyn woordeboeke was 'n gepaste beginpunt vir die samestelling van die prototipe-ARW, aangesien die teikengebruikers, gebruikssituasies en funksies van hierdie woordeboek in 'n presamestellingsfase in Blom (2020) bepaal is, en die relevante datatipes, inhoud en woordeboekstrukture vir die prototipe-ARW slegs op grond van die akademiese redigeerders se behoeftes en die ARW se kommunikatiewe, kognitiewe en operatiewe funksies geselekteer kon word. Die inhoud in die prototipe-ARW is, soos uiteengesit in tabel 1, in 16 verskillende afdelings verdeel om aan die akademiese redigeerders se gebruikersbehoefte te voldoen. Hierna bied Gouws (2014a; 2014b; 2018a; 2018b; 2018c; 2018d) se aanpassings van die algemene leksikografieteorie en Du Plessis (2015; 2017) se oorkoepelende stel beginsels vir mens-rekenaarinteraksie (ISO-standaard 9241 (1998a) (1998b)(2006)) 'n voldoende grondslag om te besluit oor 'n aanlyn woordeboek (soos die prototipe-ARW) se woordeboekstrukture en tegnologiese kenmerke te neem. Die aanbieding en uitleg van die data in die prototipe-ARW is gekies sodat die akademiese redigeerders hulself kan oriënteer en vinnig in die prototipe-ARW en buitetekste soos 'n webblad op die internet kan rondbeweeg; nie oorlaai word met data in die onderskeie woordeboekartikels nie; en weet waar om vir hulp te soek indien enige van die eersgenoemde onduidelik is.

In 'n opvolgstudie word daar verslag gedoen oor die tweede en derde take in Blom (2021) se studie, waar die volgende stap in die ontwikkelingsproses van die aanlyn woordeboek, naamlik die toetsing van die prototipe-ARW, uitgevoer is. Soos reeds in Blom (2020: 25-26) genoem, is die bruikbaarheid van die databeskrywing/-inhoud en die data-aanbieding/-uitleg in die prototipe-ARW empiries getoets deurdat studenteredigeerders die prototipe-ARW moes gebruik tydens akademiese redigeertake, en na afloop van die take 'n posttoetsvraelys moes voltooi om hulle gebruikertevredenheid aan te dui. Rubin en Chisnell (2008) se boek *Handbook of Usability Testing* is as riglyn gebruik om kwantitatiewe prestasiedata oor fout- en suksesevaluering, en kwalitatiewe voorkeurdatabe oor die gebruikers se ervaring tydens die gebruik van die prototipe, met die voltooiing van die redigeertoets in te samel. Hierdie bruikbaarheidstoetsing kon help om te bepaal of die tegnologieë suksesvol gebruik is vir die aanbieding van die prototipe-ARW se inhoud, of die studenteredigeerders die prototipe-ARW kon gebruik om die redigeertake effektief te voltooi, en watter aanpassings aan die ARW 'n beter gebruikerservaring vir die redigeerders sal bied. Aanbevelings- en verbeteringsriglyne vir die verfyning en samestelling van die finale ARW is na aanleiding van hierdie resultate opgestel.

Eindnote

1. Mossop (2014) onderskei tussen vier tipes redigering, naamlik inhoudelike, strukturele, stilistiese, en kopieredigering. Die laasgenoemde twee tipes sluit die volgende redigeertake in: die korreksie van spel-, tik- en basiese grammatikale foute, die nagaan van interpunksie, asook die uitwys en aanpassing van ontoepaslike woordgebruik en sinstruktuur vir spesifieke lesers van 'n teks.
2. Die woordeboekmodel bestaan uit 'n tuisbladsy met verskillende afdelings wat die akademiese redigeerder kan gebruik, vyf artikels van voorbeelde uit verskillende vakgebiede, asook 'n minihulp-gebruikersleiding (sien Blom 2020 vir 'n gedetailleerde illustrasie van Blom se 2018-woordeboekmodel).
3. Die data in 'n leksikografiese databasis word meestal gestruktureer in 'n tabel en volgens waardes (soos getalle) gestoor. Hierdie waardes kan slegs deur 'n databasisbestuurstelsel ("database management system" of DBMS) opgesoek en gesorteer word, sodat IT-spesialiste 'n programmatiese onttrekking van die waardes kan gebruik om data te skep en op te dateer wat dan aan die leksikograaf of gebruiker in 'n gebruikerskoppelvlak verskyn (Bergenholtz en Nielsen 2013: 81).
4. Sedert Du Plessis se studie in 2015, is nuwe weergawes van hierdie drie ISO-standaarde gepubliseer. Vandaar die twee publikasiejare by elk van dié standaarde.

Bronnelys

- Bergenholtz, H. en J.S. Nielsen.** 2013. What is a Lexicographical Database? *Lexikos* 23: 77-87.
- Blom, M.** 2020. 'n Teoretiese model vir die ontwerp van 'n Afrikaanse aanlyn Akademiese Redigeerwoordeboek. *Tydskrif vir Nederlands en Afrikaans* 27(1): 5-29.
- Blom, M.** 2021. 'n Aanlyn Afrikaanse akademiese redigeerwoordeboek: Die bruikbaarheid van 'n prototipe. M.A.-tesis. Stellenbosch: Universiteit Stellenbosch.
- Carstens, W.A.M. en K. van de Poel.** 2012. *Teksredaksie*. Hersiene uitgawe. Stellenbosch: African Sun Media.
- Du Plessis, A.H.** 2015. 'n Analise van die selfoon-WAT: 'n Grondslag vir die verbetering van selfoonwoordeboeke. M.A.-tesis. Stellenbosch: Universiteit Stellenbosch.
- Du Plessis, A.H.** 2017. Die rol van die bruikbaarheidsbenadering binne die e-leksikografie. *Literator* 38(2): 1-11.
- Engelberg, S. en C. Müller-Spitzer.** 2013. Dictionary Portals. Gouws, R.H., U. Heid, W. Schweickard en H.E. Wiegand (Reds.). 2013: 1023-1035.
- Fuertes-Olivera, P.A. en S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlyn/Boston: De Gruyter.
- Gouws, R.H.** 2014a. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Gouws, R.H.** 2014b. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3): 481-504.
- Gouws, R.H.** 2018a. Accessibility, Access Structures and Access Procedures. Jesenšek, V. en M. Enčeva (Reds.). 2018. *Wörterbuchstrukturen zwischen Theorie und Praxis*: 35-56. Berlyn: De Gruyter.

- Gouws, R.H.** 2018b. Internet Lexicography in the 21st Century. Engelberg, S., H. Kämper en P. Storjohann (Reds.). 2018. *Wortschatz: Theorie, Empirie, Dokumentation*: 215-236. Berlyn: De Gruyter.
- Gouws, R.H.** 2018c. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.
- Gouws, R.H.** 2018d. Dictionaries and Access. Fuertes-Olivera, P.A. (Red.). 2018. *The Routledge Handbook of Lexicography*: 43-58. Londen: Routledge.
- Gouws, R.H., U. Heid, W. Schweickard en H.E. Wiegand (Reds.)**. 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlyn/New York: De Gruyter.
- Gouws, R.H. en D.J. Prinsloo**. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- International Organisation of Standardisation**. 2017 (1998b). ISO 9241-12. [Internet]. Beskikbaar: <https://www.iso.org/standard/64839.html> [20 Oktober 2020].
- International Organisation of Standardisation**. 2018 (1998a). ISO 9241-11. [Internet]. Beskikbaar: <https://www.iso.org/standard/63500.html> [20 Oktober 2020].
- International Organisation of Standardisation**. 2020 (2006). ISO 9241-110. [Internet]. Beskikbaar: <https://www.iso.org/standard/75258.html> [20 Oktober 2020].
- Law, M.A.** 2011. The Development of Core Standards for Editing in South Africa. *Southern African Linguistics and Applied Language Studies* 29(3): 275-292.
- Lourens, A.** 2014. Produk teenoor proses tydens akademiese redigering: Opmerkings as aanduiders van redigeergerigtheid. *Stellenbosch Papers in Linguistics Plus* 43: 261-296.
- Mossop, B.** 2014. *Revising and Editing for Translators*. Derde uitgawe. Manchester: St. Jerome.
- Rubin, J. en D. Chisnell**. 2008. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Second edition. Indianapolis, IN: John Wiley.
- Universiteit Stellenbosch Taalsentrum**. 2019. *Redigering van jou tesis of navorsingsverslag*. [Internet]. Beskikbaar: http://www0.sun.ac.za/taalsentrum/assets/files/TaaldiensDokumente/Tesisredigering_2019_Okt.pdf [28 Maart 2020].
- Tarp, S.** 2015. Structures in the Communication between Lexicographer and Programmer: Database and Interface. *Lexicographica* 31: 217-246.
- Van Aswegen, E.S.** 2007. Postgraduate Supervision: The Role of the (Language) Editor: Sed quis custodiet ipsos custodes? *South African Journal of Higher Education* 21(8): 1139-1151.
- Wiegand, H.E. en R.H. Gouws**. 2013. Macrostructures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard en H.E. Wiegand (Reds.). 2013: 73-109.

A Comparative Analysis of Piotr Borkowski's (1963) and Roman Gajda's (1970) English– Polish Phraseological Dictionaries: Practice vs. Theory

Mirosława Podhajecka, *Institute of Linguistics, University of Opole,
Poland* (mpodhajecka@uni.opole.pl)

Abstract: This paper offers an insight into the short and largely unexplored history of English–Polish and Polish–English phraseological lexicography. It aims to analyse two post-war English–Polish phraseological dictionaries, *An English–Polish Dictionary of Idioms and Phrases* (1963) by Piotr Borkowski and *Wybór idiomów angielskich [A Selection of English Idioms]* (1970) by Roman Gajda, from a qualitative and quantitative perspective. At first sight, they seem to share several features, insofar as both are monoscopal English–Polish volumes of a similar size; both were addressed to Polish learners of English; and both drew on *The Kosciuszko Foundation Dictionary: English–Polish* (1959), an exhaustive reference work available at that time. Closer scrutiny, however, reveals marked differences not only in the contents of the dictionaries, but also in their authors' ideas of phraseology. This case study is preceded by theoretical considerations concerning the nature of comparative analyses in (meta)lexicographical discourse. The conclusions draw partly on the practical and partly on the theoretical findings.

Keywords: DICTIONARY, POLISH, ENGLISH, PHRASEOLOGY, MULTI-WORD EXPRESSION, IDIOM, EQUIVALENT, COMPARATIVE (META)LEXICOGRAPHY, SOCIOCULTURAL (META)-LEXICOGRAPHY

Opsomming: 'n Vergelykende analise van die Engels—Poolse fraseologiese woordeboeke van Piotr Borkowski (1963) en Roman Gajda (1970): Praktyk vs. teorie. Hierdie artikel bied 'n perspektief op die kort en grootliks onbekende geskiedenis van die Engels–Poolse en Pools–Engelse fraseologiese leksikografie. Daar word gestreef om vanuit 'n kwalitatiewe en kwantitatiewe beskouing twee ná-oorlogse Engels–Poolse fraseologiese woordeboeke, *An English–Polish Dictionary of Idioms and Phrases* (1963) deur Piotr Borkowski en *Wybór idiomów angielskich [A Selection of English Idioms]* (1970) deur Roman Gajda, te analiseer. By die eerste aanblik kom dit voor asof hulle, as beide eenrigting- Engels–Poolse woordeboeke van soortgelyke grootte, sekere kenmerke deel; albei is gerig op Poolse leerders van Engels; en albei maak (sterk) gebruik van *The Kosciuszko Foundation Dictionary: English–Polish* (1959), 'n uitgebreide naslaanwerk wat ten tyde van die samestelling van hierdie woordeboeke beskikbaar was. By nadere ondersoek blyk dit egter duidelik dat daar merkbare verskille tussen die twee bestaan, nie net in die inhoud van die woordeboeke nie, maar ook in die outeurs se sienings van die fraseologie.

Hierdie gevallestudie word voorafgegaan deur teoretiese oorwegings rakende die aard van vergelykende analises in die (meta)leksikografiese diskoers. Die gevolgtrekkings waartoe gekom word, word deels op die praktiese en deels op die teoretiese bevindings gebaseer.

Slutelwoorde: WOORDEBOEK, POOLS, ENGELS, FRASEOLOGIE, MEERWOORDIGE UITDRUKKING, IDIOOM, EKWIVALENT, VERGELYKENDE (META)LEKSIKOGRAFIE, SOSIO-KULTURELE (META)LEKSIKOGRAFIE

1. Introduction

Given the European dictionary-making traditions, the history of English–Polish and Polish–English lexicography is not extensive, inasmuch as its beginnings may be traced back to the end of the eighteenth century (Antonowicz 1788), the first major work of reference being published in the mid-nineteenth century (Rykaczewski 1849–1851). Nor has it been particularly rich. Although a number of bilingual dictionaries appeared during that period, most were pocket-size books produced by amateurs rather than specialists. The history of English–Polish phraseological dictionaries, which has attracted very little attention, is even humbler. The first booklet that may be considered a phraseological dictionary, Adam Richter's *Polish Dictionary of English Idioms, Proverbs and Slangs*, came out in Tel-Aviv in 1945 and never found its way into the hands of Polish learners of English, at least in Poland.

Until the early 1950s, only one high-quality dictionary, Mieczysław Kobylański's *Wybór idiomów angielskich [A Selection of English Idioms]* (1951), had been launched onto the domestic market. This was by no means a quantum leap, but the volume was compiled by an expert. A graduate in English from the Jagiellonian University in Kraków, Kobylański knew several foreign languages, had both interest in and practical experience of foreign-language teaching, and was well aware of the complexity and fuzzy boundaries of English phraseology. Regrettably, for ideological reasons, the dictionary never went into a second edition (Podhajecka 2020: 134).

The situation improved with the publication of two further phraseological dictionaries: Piotr Borkowski's *An English–Polish Dictionary of Idioms and Phrases* (1963) and Roman Gajda's *Wybór idiomów angielskich [A Selection of English Idioms]* (1970). The former was originally issued in London but came to be widely republished on the Polish market. The latter, too, enjoyed unflagging popularity, which required Warsaw's state-owned publisher Wiedza Powszechna continuously to issue subsequent editions well into the 1990s. Needless to say, the demand for English handbooks and bilingual dictionaries in Poland stemmed from the growing significance of English as a foreign language.

This article looks at the two competing English–Polish phraseological dictionaries from a comparative perspective. By exploring the ways in which reference works are contrasted, the case study is also expected to contribute to what may be termed comparative (meta)lexicography and, by researching the

sociocultural contexts of dictionary production, to what may be termed sociocultural (meta)lexicography.

2. Theoretical considerations

A study of this kind begins with a number of research questions. Some concern the purely comparative perspective: What techniques should be applied in evaluating dictionaries? How different are the techniques used for contrasting general and specialised dictionaries? What are the peculiarities of comparative analyses of phraseological dictionaries? Others pertain to the sociocultural context: What is the relationship between dictionary compilers and their environments? Do dictionaries communicate with existing lexicographical traditions? What motivates compilers to undertake their projects? Let us try to address the theoretical framework first.

Hartmann and James (2001: 24) tell us that *comparison of dictionaries* denotes 'the contrastive evaluation of two or more dictionaries [...] for the purpose of dictionary criticism or the study of dictionary history'. The definition of *evaluation* is equally concise: 'the process of assessing a dictionary or other reference work, often in comparison with others. A systematic framework for formulating criteria with respect to coverage, format, scope, size, title etc. has yet to be developed' (Hartmann and James 2001: 53). A review of the literature indicates that this issue has indeed been neglected. Studies dedicated to a variety of dictionaries in both lines of inquiry, ranging from general to fine-grained, abound, however (e.g. Starnes 1963; Kerling 1979; Atkins 1985; Kister 1992; Masuda et al. 1999; Cormier 2003; Miyoshi 2007; Chen 2010; Considine 2014; Lew and Szarowska 2017; McConchie 2019).

Yong and Peng (2021: 222) adumbrate the comparative approach as follows:

Comparison can be made of different versions and editions of the same dictionary,¹ of different dictionaries of identical and similar types, of different types of dictionaries and of dictionaries across cultures, languages and nations. In doing so, lexicographers can discover attributes dictionaries share; reveal similarities, associations and differences that exist in dictionaries of the same language and culture or different languages and cultures and eventually develop and establish more scientific notions, principles and methodology of lexicography through reference, exemplification and inspiration.

This offers a useful bird's-eye view, but the approach to the task is left unexplained. Faced with methodological obstacles, I decided to treat this issue by rule of thumb, focusing mainly on studies of dictionary history. This results from two factors. Firstly, evaluation criteria proposed in user-oriented studies (e.g. Nakamoto 1995; Nielsen 2009; Swanepoel 2013) clearly favour the synchronic perspective. Secondly, such studies aim "to contribute towards improving the quality of a dictionary [...] to help to further progress [in] lexicography per se" (Akasu 2022: 48), whereas historical research tends to have little or no relevance to future lexicography.²

The extent of the subtlety in the differences between the synchronic and diachronic approaches may well be exemplified. Jackson (2002: 176-177) has it that two types of criteria have been distinguished in dictionary criticism: internal and external. The former derives "from what a dictionary says about itself, or what the editors claim", helping to test whether the claims are supported by the working practice. The latter should cover two sets: one relating to "the reference function and the user's perspective" (i.e. presentation and accessibility) and the second "to the recording function of dictionaries" (i.e. content). Irrespective of the universality of this framework, referring to the user's perspective is often futile in the historical context, as it may be hard to establish who the users of, say, a sixteenth-century dictionary were and whether the lexicographical material addressed their needs. Considine (2022: 1) argues that "every educated adult in England, and in much of Wales and Scotland [...] had been a careful dictionary user in her or his formative years, because dictionaries and wordlists were the keys to even the most modest knowledge of Latin", but he wisely stops short of conjecture, which would inevitably be speculative.

Comparative analyses take into account the content and structure of two or more dictionaries, especially the megastructure, e.g. the front and middle matter (e.g. Nkomo 2016; Vişan 2018); the macrostructure, including the coverage and types of headwords (e.g. Stein 1985; Ogilvie 2013); and the microstructure with its wide repertoire of information categories (e.g. Cowie 1999; Farina and Durman 2009). For bilingual dictionaries, as Tomaszczyk (1988: 289) notes, the choice of equivalents is an additional parameter (e.g. Frączek 1999; Bately 2009). Analyses are usually done dictionary by dictionary (e.g. Steiner 1970) or feature by feature (e.g. Bogaards 1996), but other configurations also come into play (e.g. Reddick 2009). As might be expected, selected elements are depicted visually, while others are described. To arrive at sound conclusions, scholars are encouraged to combine quantitative and qualitative methods (cf. Coleman and Ogilvie 2009).³

Contrary to popular belief, dictionaries are far from impartial repositories of information (e.g. Benson 2002: 4-5; Chen 2015: 312). They are rather socially and culturally rooted in an environment that changes in time and space, which is why they need to be compared in the relevant context (e.g. Fishman 1995: 34). Ever since the first Sumerian word-lists, dictionaries have evolved and the evolution has been prompted by innovation, inspiration, and imitation (e.g. Landau 2001: 46-47; Stein 2014: 397-398). As they are in various degrees influenced by, or indebted to, others' works, tracing them to their sources is essential (e.g. Cormier and Fernández 2004; Podhajecka 2013). Dictionary production has also experienced complex interactive relationships with socio-cultural advances, driven by compilers' aspirations and vision. These are also worthy of investigation. All in all, integrating the textual data with biographical and sociocultural information helps discover the stories lying behind "the neatly printed pages of the finished text" (Mugglestone 2011: xii). The more insightful the analysis, the more discoveries will come to light.

3. Idioms and lexicography

Idioms constitute a central category of phraseology, an interdisciplinary domain which has become "pervasive in all language fields" (Granger and Meunier 2008: xix), but which is in itself a challenge to define consistently (see, e.g., Cowie 2001: 210; Mel'čuk 2012: 31; Miller 2013: 275; Espinal and Mateu 2019). One rudimentary problem in idiom research is that "they are found at one extreme of a continuum ranging from totally free combinations of words to completely frozen, fixed multiword units", exhibiting degrees of fixedness (Fontenelle 2001: 191). Another, as Grant and Bauer (2004: 42) point out, is the use of different sets of criteria and classifications, which has resulted in "mismatches" between the terminologies of a variety of theoretical approaches (Moon 2015: 319), including those of generative linguistics, cognitive linguistics, psycholinguistics, and corpus and computational linguistics.

These complexities lead to practical problems of how one might present idioms in dictionaries for foreign language learners. Several factors are responsible for this state of affairs. Firstly, idioms are characterised by different degrees of opacity and this influences their lemmatization and position within the entry (Harras and Proost 2005: 280-285). It also influences the choice of the form of citation, such as the conventional canonical form (e.g. *to kill two birds with one stone*), a genuine citation (e.g. *fortune favours the brave*), or a contextual use (e.g. *how right you are*). Even if we assume that "dictionary users do seem to expect all multi-word units to have one element that is more important than the others" (Béjoint qtd. in Szczepaniak 2012: 49), finding idioms is fraught with difficulty. Secondly, idioms are "inherently complicated: they have their own internal grammars, their own connotations and pragmatic functions, and they very often have fluid, contextually-dependent meaning" (Moon 1999: 265), all of which may be confusing for the dictionary user. Around 40% of English idioms, moreover, have no fixed forms (Moon 2001: 92); many accept lexical and grammatical variation (Pinnavaia 2002: 57); and some are manipulated creatively despite being no "open-slot idioms" (cf. Solano 2013: 153). For example, of eleven instances of *a bird in the hand* in the BNC, only four end with *is worth two in the bush*; among the innovative creations we find *a bird in the hand is worth two votes for Bush* and *I wanted a bird in the hand, but this one was practically in Shepherd's Bush!* Lastly, idioms represent a distinct linguo-cultural heritage, so seeking equivalence between them and foreign language counterparts may be a daunting task (cf. Piirainen 2008: 248, 252).

From the point of view of language learners, the equivalent is a key element of a bilingual dictionary. Dobrovolskij (2013) distinguishes four classes, treating the image component as an important criterion:

- (1) "full equivalents" ("absolute equivalents"), i.e. idioms of L1 and L2 that are identical with regard to meaning, syntactic and lexical structure, and imagery basis;

- (2) "partial equivalents", i.e. idioms of L1 and L2 that have identical or near-identical meanings, but correspond imperfectly in syntactic and lexical structure, or imagery basis;
- (3) "phraseological parallels", i.e. different idioms of L1 and L2 that correspond to each other in the core meaning, but not with regard to the image component;
- (4) "non-equivalents", i.e. a given L1 idiom that has no idiomatic correspondences in L2.

Classes (1)–(3) may be safely considered functional equivalents, i.e. lexical items of both L1 and L2 that may be used "in the same situations" (Dobrowol'skij 2000: 169). As idiomatic expressions, they have the same status in both languages, even though the image evoked may be strikingly different.⁴ It should be noted that closeness of imagery alone "does not guarantee identity with respect to all parameters of comparison (especially pragmatics)", as Szczepaniak and Adamska-Sałaciak (2010: 91) suggest in their comparative analysis.

4. The lexicographical context

As has been mentioned, until the mid-20th century, most of the English–Polish and Polish–English dictionaries were small and unsophisticated. It was only in the late 1950s and early 1960s that two comprehensive dictionaries appeared. *The Kościuszko Foundation Dictionary: English–Polish* (1959) (henceforth, KFD), compiled by Kazimierz Bulas and Francis Whitfield, was published in the Hague mainly to serve the Western market. In 1961, in collaboration with Lawrence L. Thomas, they produced the Polish–English volume. The first volume had 1029 pages and the second, with 758 pages, was considerably smaller. Bulas, a Polish archaeologist, emigrated to the United States, where he worked as a librarian at Rice University in Houston, whilst Whitfield and Thomas were American Slavists affiliated with the University of California in Berkeley (Adamska-Sałaciak 2016: 84–85).

The second comprehensive dictionary was undertaken by Jan Stanisławski, a Pole born in Siberia and educated in Britain and France, who had already made his name in Poland as a bilingual lexicographer. His monumental work, *The Great English–Polish Dictionary* (1964), was issued by Wiedza Powszechna, a major Polish publisher. The English–Polish volume consisted of 1103 pages and the Polish–English version, published in 1969, as many as 1502 (Piotrowski 2001: 203). Both were furnished with supplements compiled by Stanisławski and his daughter Małgorzata Szercha. Although the KFD was later republished in Poland, Stanisławski's endeavour and its abridgments practically monopolised the Polish dictionary market until the end of the twentieth century.

5. Borkowski's and Gajda's biographies

It is customary for a study of an important dictionary to begin with essential

biographical facts about its author that may "help to explain some of his lexicographic policies" (Landau 2009: 182-184; cf. Considine 2008; Cormier 2009; Rennie 2012). Neither Borkowski nor Gajda was an accomplished lexicographer, but sketching their biographical backgrounds may nonetheless be of some value.

Piotr Borkowski (1907–1985) was born in Pokrów (in what today is Russia). In 1918, after Poland regained independence, his family moved to Poland. In 1933, he graduated from the Szkoła Główna Handlowa [Principal School of Commerce] in Warsaw and later worked as editor of one of Warsaw's weeklies. He left for France and, when the war broke out, headed to Coëtquidan, in Brittany, where the formation of the Polish Armed Forces in the West began in November 1939 (cf. Grodziska 2001: 346). When France fell, he made his way, through Spain, to Britain (*Pamiętnik literacki* 1986: 78). After the war, he remained in London as a journalist with the Polish-language newspaper *Wiadomości*.

Borkowski had a keen interest in phraseology, which became his hobby-horse. His first publication, *Praktyczny poradnik podatkowy* [*A Practical Tax Guidebook*] (1953), was followed by *An English–Polish Dictionary of Idioms and Phrases* (1963), *The Great Russian–English Dictionary of Idioms and Set Expressions* (1973), and an entertaining book on less serious aspects of dictionary-making, *O językach i słownikach na wesoło* [*A Humorous Look at Languages and Dictionaries*] (1974) (cf. Laks 1978: 14).

The English–Polish dictionary was republished in Britain (London: Odnova 1970, 1972 (2nd ed.), Orbis Books 1982) and the United States (New York: Hippocrene Books 1982, 1983) for the benefit of the 'hyphenated' Polish–British and Polish–American immigrants. It also appeared in Poland (Poznań: Adam Mickiewicz University 1987, 1988, 1989, 1991, 1994). The Polish editions were possible thanks to Jerzy Kulczycki (1931–2013), a Polish publisher and bookseller in England, to whom Borkowski transferred copyright to the dictionary (Fisiak 1989: 3).

The biography of Roman Gajda (1908–?) has received less attention. He was born in Pabianice, in the Łódź province, where his parents worked in the textile industry. Financial difficulties in the family meant that he was unable to pursue a secondary education. Being self-taught, he did odd jobs before settling down as a newspaper editor. Around 1926, the local press first printed samples of his poetry and prose. During the war, he worked for *Głos Ludu* [*The People's Voice*] and afterwards for the Polska Agencja Wydawnicza [Polish Press Agency]. He wrote two novels: *Miasto mojej młodości* [*The Town of My Youth*], an unpublished autobiography, and *Ludzie ery atomowej* [*People of the Atomic Era*] (1957), the first Polish science-fiction novel. He also produced the radio play *Tahiti* and wrote memoirs, which remained in manuscript form (*Fantastyka* 1983: 80).⁵ His English–Polish dictionary of idioms, first published in 1970, met with wide acclaim and went through several editions (1972, 1975, 1985, 1988, 1990, 1991, 1992, 1993, 1995). It is unknown when and where Gajda learnt English to the level that enabled him to compile the dictionary and why it was a

phraseological dictionary.

As is clear from the two sketches above, both compilers were language enthusiasts rather than fully-fledged Anglicists. Borkowski had an advantage over Gajda insofar as he lived in an English-speaking country, although his working for a Polish newspaper suggests that he belonged to a close-knit ethnic community in London and may not have been fully bilingual. Gajda's biography is patchy, but there is every indication that he never received a formal education in English, so his linguistic knowledge seems to have been a product of his own efforts. This should not be treated as a curiosity, however, because some of the best English–Polish and Polish–English dictionaries were made by amateurs with a flair for languages.⁶ Erazm Rykaczewski (1803–1873), a graduate in law and philosophy from Wilno University; Władysław Kierst (1868–1945), a would-be medical doctor expelled from Warsaw University; and Kazimierz Bulas (1903–1970), an archeologist affiliated with the Jagiellonian University before the Second World War, are among the most conspicuous examples of this.

Borkowski's and Gajda's dictionaries are still mentioned in specialist literature (e.g. Zakrzewski 2002; Szerszunowicz 2006, 2016), which shows that they filled an important niche on the dictionary market in the second half of the twentieth century.

6. The dictionaries

6.1 Megastructure

At first sight, the two dictionaries look very much alike: both are monoscopal English–Polish volumes of a similar size and length; both were compiled by journalists; and both authors drew on the KFD. Upon closer scrutiny, however, one discovers marked differences not only in the contents of the dictionaries, but also in their authors' ideas of phraseology. This subsection looks at the front and back matters.

Borkowski is articulate regarding the motivations that prompted him to undertake the task, one of which was "to refresh the speech of Polish immigrants" [M.P.]. His agenda is complex. On eleven pages of the Polish preface and three pages of the English foreword, he tackles various issues: MWEs typical of British and American English, the Bible and national literatures as sources of idioms, and different types of phraseological units. He pays special attention to explaining his arrangement of idioms, albeit, as my research indicates, he failed to treat them with absolute consistency. This is what he writes about his envisaged readership:

it has been written primarily for Poles who settled in English-speaking countries during and after the last war ... I fervently hope that the dictionary may prove to be of considerable help to all students of Polish, both Polish children born in English-speaking countries, and Polish pre-war emigrants who are trying to

brush up and enrich their idiomatic Polish, and, last but not least, all other students of Polish.

Much of the front matter is, however, purple prose. This explains why, in the Polish editions, both components were replaced by a brief preface penned by Jacek Fisiak.⁷ The back matter in Borkowski's dictionary is composed of a single blank page for the user's handwritten notes and a list of abbreviations, including *hist* 'historyczne' [historical], *pol* 'polityczne' [political], *pot* 'potoczne' [colloquial], and *sl* 'slang' [slang]. We know that financial restrictions in the production of the dictionary made it impossible to incorporate an index. This apparent shortcoming was compensated for in later editions.

Gajda's three-page introduction is better suited for foreign language learners, the dictionary's target users. He explains what word combinations he regards as idioms and how they are arranged, acknowledging openly, or so it seems, the sources used in his compilation. These include four monolingual dictionaries: A.S. Hornby's *The Advanced Learner's Dictionary of Current English* (1948?), B.L.K. Henderson's *A Dictionary of English Idioms* (1937?), Thomas L. Cromwell's *A Glossary of Phrases with Prepositions* (1960?), and A.J. Worrall's *English Idioms for Foreign Students* (1932?), even though their actual impact is hard to ascertain with any degree of credibility.⁸ The KFD is the only bilingual English–Polish reference work mentioned.

On the whole, Gajda's introduction is more coherent than Borkowski's, but his discussion of *mieć węża w kieszeni* and *his money comes from him like drops of blood*, a Polish idiom and an English proverb respectively,⁹ appears somewhat out of place. He claims to have obtained the authors' permission for the use of illustrative material taken from the above-mentioned dictionaries, demonstrating a sensitivity to ethical issues.¹⁰ With a key to exercises and an index of all the headwords the idioms contain, the back matter is more extensive than Borkowski's.

6.2 Macrostructure

The two dictionaries are dissimilar in terms of their macrostructures. Borkowski's dictionary includes 4,341 word combinations: idioms (e.g. *make the dust fly*), proverbs (e.g. *no garden without its weeds*), colloquial or humorous sayings (e.g. *pigs might fly*), binomials (e.g. *by and large*), collocations (e.g. *top secret*), and formulaic expressions (e.g. *at first sight*), but also a host of single words (e.g. *bobby, dare-devil, irons, labour, nobody, snow-man, Yankee*) and clauses (e.g. *I have not seen him for years, he is young for his age*).¹¹ He selected them on the basis of monolingual dictionaries, dozens of which he perused during the preparatory phase (Borkowski 1963: iv).

By contrast, Gajda registers merely 662 idioms, i.e. less than one sixth of Borkowski's coverage. At the same time, however, he treats them more consistently, so the number of MWEs whose phraseological status is unclear is negli-

gible. It should be emphasised that he records as many as 263 phrasal verbs (e.g. *to knock off*, *to point out*, and *to tell on*), which are few and far between in Borkowski's dictionary. The lexicographical material is complemented by different types of exercises.

Both dictionaries are arranged alphabetically. Borkowski listed his MWEs by content words according to a system devised by himself, which he neglected to explain in his long preface. His working principle, i.e. "each idiom is based on a content word which we call a lemma. How the idiom is arranged alphabetically depends on the lemma" [M.P.], is rather uninformative. Consequently, the ordering of the MWEs is haphazard: *meet half way* is entered under *meet*, *set tongues wagging* under *tongues*, and *to drive a coach and six through the law* under *law*, which is the last of the four content words.¹² This kind of arrangement clearly forced the user, as it did me, to search for the idioms by trial and error. To make things worse, the headwords remain undetermined typographically. This is a far cry from what we would consider user-friendliness today.¹³

Gajda's headwords in bold capital letters are followed by headphrases (i.e. MWEs) in bold. He explains that, in arranging the idioms, he paid attention both to the first word and the words it went with (e.g. *grasp* in *beyond one's grasp*). This is also uninformative. One might wonder why *a bone of contention* is entered under *bone*, but *to pay in kind* under *kind*. Fortunately, the index in the back matter is transparent and helps identify the idioms instantaneously, the more so because each idiom is entered under all the content words it includes (e.g. *to make one's flesh creep* is entered under *make*, *flesh*, and *creep*).

6.3 Microstructure

6.3.1 Borkowski's dictionary

Borkowski's entry covers only the English MWE, sometimes labelled, and its Polish counterpart. The use of small capitals for the headword,¹⁴ contrasted with unmarked typeface for the equivalent, is the only typographical device. The use of small capitals is unsatisfactory, because the target user is given no information as to which words are spelled in upper-case and which in lower-case (e.g. *April Fool-Day*, *Jolly Roger*, *keep up with the Joneses*, *to catch a Tartar*, or *Tom Thumb*). The small capitals were wisely replaced by bold in the Polish editions.

Whenever possible, Borkowski proposes variant forms of the MWEs or their elements, usually introduced after a comma or in round brackets:

blind drunk, drunk as a fiddler, as a lord, as an owl
a long dozen, devil's dozen, baker's dozen, printer's dozen
dressed up to the nines (dressed to kill)
go and have your head (brains) examined

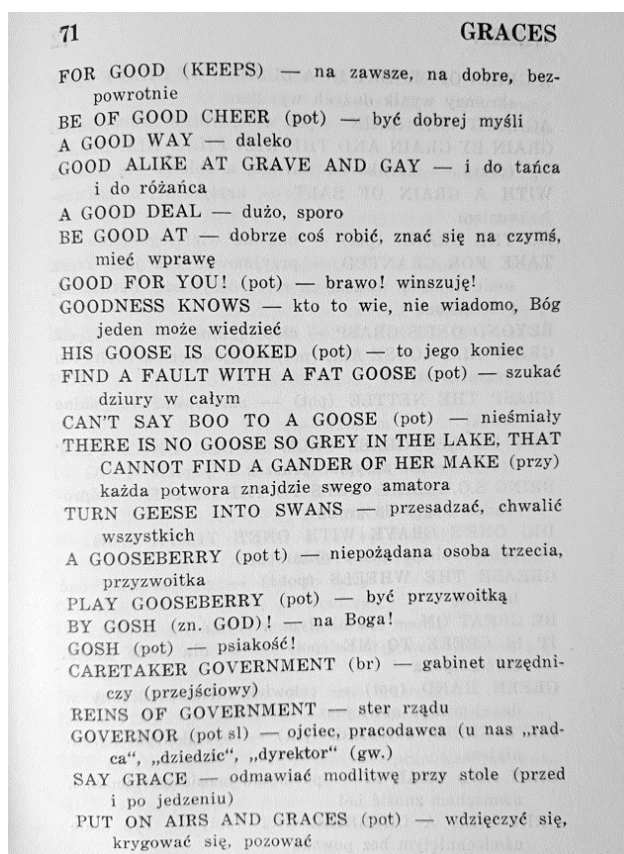


Figure 1: A sample page from Borkowski's dictionary

The value of phraseological dictionaries consists in a prudent selection of L1 or L2 word combinations and appropriate equivalents.¹⁵ Despite speaking different idiolects, lexicographers need to make sure that both are correct from the linguistic and stylistic points of view. Let us look at a few entries exhibiting the weaknesses of Borkowski's translations:

- beat s.o. black and blue* 'zrobić komuś siniec' > 'zbić kogoś na kwaśne jabłko'
- clock-wise* 'w kierunku jak zegar chodzi' > 'zgodnie z ruchem wskazówek zegara'
- pigs might fly* 'dzieją się cuda' > 'prędyj mi kaktus na dłoni wyrośnie'
- think poor of s.o.* 'być ujemnego zdania o kimś' > 'mieć o kimś złe zdanie'

On closer examination, the range of quirks turns out to be far wider. For example, some equivalents are ambiguous (e.g. *gammon the hind leg of a donkey* '(pot)

wziąć na kawał');¹⁶ calqued and, hence, culturally inadequate (e.g. *handle-bar moustache* 'wąsy w kształcie kierownicy roweru'); unidiomatic even though Polish idioms were readily available (*the darkest place is under the candlestick* 'osoba zainteresowana w czymś nic o tym nie wie' > 'najciemniej jest pod latarnią'); or simply wrong (e.g. *can-opener* 'tani samochód'). Apart from that, numerous spelling mistakes are also in evidence throughout (e.g. *come of the grass* > *come off the grass*; *Mondaish* > *Mondayish*; *mine concern* > *my concern*). The mistakes were corrected in the Polish editions.

There are also cases when the English headphrases are syntactically incongruent with their Polish counterparts. This may lead to misunderstandings and misuse, particularly by a less competent user. A few examples are given below.

born out of wedlock 'nieślubne dziecko' > *a child born out of wedlock*
s.o. is unable to see a hole in a ladder 'załany w pestkę' > *unable to see a hole in a ladder*
you must take the pot luck 'czym chata bogata, tym rada' > *to take pot luck*
'zadowolić się tym, co jest'
sell like hot cakes 'szybko sprzedawać towar, cieszący się dużym popytem' >
to sell like hot cakes 'sprzedawać się jak świeże bułeczki'

It is worth looking at the dictionary material through the prism of Dobrovolskij's classification. Analysis suggests that full equivalents in Borkowski's dictionary are relatively rare. Here are a few examples of category (1):

like a red rag to the bull 'jak czerwona płachta na byka'
over my dead body 'po moim trupie'
a wolf in likeness of a lamb 'wilk w owczej skórze'
back a wrong horse 'stawiać na złego konia'

Most idiomatic equivalents fall into the two subsequent categories, (2) partial equivalents and (3) phraseological parallels. The former are characterised by semantic affinity and differences in structure or mental imagery.

burn one's boats 'spalić za sobą mosty'
cry for the moon 'chcieć gwiazdki z nieba'
the pot calling the kettle black 'przyganiał kocioł garnkowi'
work one's fingers to the bone 'zapracować się po łokcie'

The basic feature of the latter is a significantly different image component. This category may be exemplified by the following:

as hungry as a hawk 'głodny jak wilk'
half a loaf is better than no bread 'lepszy rydz niż nic'
that is where the shoe pinches 'w tym sęk'
talk of the devil and he will appear 'o wilku mowa, a wilk tuż'

Category (4) of Dobrovolskij's classification, which may for the sake of convenience be referred to as paraphrases, covers examples of non-equivalents:

milk the bull 'zajmować się sprawą z góry skazaną na niepowodzenie'
dressed up to the nines 'elegancko ubrany'
a penny for your thoughts 'ciekaw jestem, nad czym się zamyśliłeś'
a stitch in time saves nine 'zeszyj dziurkę póki mała'

Some of Borkowski's entries contain more than one idiomatic equivalent (e.g. *not to mince matter* (*without mincing matters*) 'nie owijać w bawełnę, wykładać kawę na ławę'), an idiomatic equivalent and an unidiomatic paraphrase (e.g. *wallflower* 'panna siejąca pietruszkę (gdy inni tańczą, ona siedzi pod ścianą)'), or just a paraphrase (e.g. *sugar daddy* 'starszy pan, wydający sporo pieniędzy na młodą kobietę').

6.3.2 Gajda's dictionary

Gajda's coverage is much better than Borkowski's. He includes MWEs from the peripheries of phraseology, such as *close to sb/sth* 'blisko kogoś, czegoś', *made by hand* 'wykonany ręcznie', and *to wait for* 'czekać na kogoś, na coś', but these are infrequent. Idioms and phrasal verbs make up the majority of the headphrases, but every now and then we also come across proverbs:

Ill weeds grow apace 'Złe ziele najlepiej się krzewi'
There's no place like home 'Wszędzie dobrze, ale w domu najlepiej'
Make hay while the sun shines 'Kuj żelazo, póki gorące'
Any stick to beat a dog 'Kto chce psa uderzyć, zawsze kij znajdzie'

The typical entry structure in Gajda's dictionary includes the headword, which is not only a content word, but also a function word (e.g. *across*, *off*, *through*), and one or more headphrases (for *back*, these will be *to back out* (*of sth*) and *with one's back to the wall*). Headphrases are paired with one or more Polish equivalents and are followed by one or more contextual uses with Polish translations. Each headword section is then followed by elements of the middle matter, i.e. translation exercises from or into English, gap-fill exercises, or sentence-formation exercises. As well as unmarked type, Gajda employs bold and italics to distinguish specific types of lexicographical information.

Whenever possible, the entry includes optional elements or words with which the headword collocates:

in [due, good] time
have [keep] sth in view
to be [beyond, out of] sb's reach [within easy reach]
to be [in the] right [wrong]; to lose [put on] weight

It is hardly surprising that the quality of Gajda's bilingual material is almost

perfect. Not only was it based on a handful of monolingual dictionaries and the KFD, but it was also reviewed by competent Anglicists.¹⁷ Tiny inconsistencies might be found in, among other things, the illustration of the literal rather than the transferred sense (e.g. *to turn the corner* 'skręcić na rogu ulicy; minąć punkt krytyczny (w chorobie itp.)') or no fully suitable equivalents (e.g. 'ujść płazem; ujść na sucho' for *to get away with*). In a few cases, Gajda clarifies the meanings of the Polish equivalents (e.g. *to mind one's own business* 'pilnować swego własnego nosa (swoich własnych spraw)'), albeit native speakers of Polish were unlikely to need such explanations.

help	108	hold
to help sb [oneself] to sth		podać komuś coś; poczęstować kogoś [się] czymś; nałożyć komuś [sobie] na talerz itp.
May I help you to some more meat?		Czy mogę podać panu jeszcze trochę mięsa?
Help yourself to the fruit.		Proszę poczęstować się owocami.
Ćwiczenia		
I. Ułożyć zdania w języku angielskim stosując podane zwroty:		
— „it can't be helped ”;		
— „to help sb (oneself) to sth ”.		
II. Przetłumaczyć na język angielski podane niżej zdania:		
1. Proszę poczęstować się czekoladkami.		
2. Pomogliśmy rannemu mężczyźnie zdjąć ubranie.		
3. Pozwolisz, że podam ci płaszcz?		
4. Nic na to nie poradzę, że mój mąż ma aż tylu nieciekawych krewnych.		
5. Proszę pomóc tej kobiecie założyć płaszcz.		
6. Czy mogę poczęstować się papierosami?		
HOLD		
to catch [get, take] hold of sb/sth		chwycić kogoś, coś; złapać za coś; chwycić się czegoś
Let's move the table. You catch hold of that end.		Przestawmy stół. Ty <i>chwyc</i> za ten koniec.
The policeman tried to get hold of the boy, but he failed.		Policjant usiłował <i>chwycić chłopca</i> , ale mu się nie udało.
to hold sb [oneself] back (from sth)		powstrzymywać kogoś [się] od czegoś; opóźniać coś; zataić
He was so impatient that I could not hold him back.		Był tak niecierpliwy, że nie mogłem <i>go powstrzymać</i> .
I held myself back from saying a word.		<i>Powstrzymałem się</i> i nie powiedziałem ani słowa.

Figure 2: A sample page from Gajda's dictionary

The main difference between the two dictionaries is that only Gajda provides illustrative examples to show how the idioms are used in context. This was a step in the right direction; Lubensky and McShane (2007: 927) stress that "no amount of information and no number of examples are superfluous to the motivated L2 learner — the more, the better". The quotations, simple and succinct, were borrowed from the monolingual dictionaries, but some, we are told, were also coined by the author. Table 1 exhibits the treatment of a few phrasal verbs.

Idiom	Equivalentents	Quotations
to burn out (of <i>sth</i>)	wykurzyć ogniem; zmusić do ucieczki (przez podpalenie); spalić, wypalić	They burnt the enemy out of their fox-holes. <i>Wykurzyli</i> ogniem nieprzyjaciela z jego (ich) nor. That light bulb has burned out . Ta żarówka <i>wypaliła się</i> .
to lead off	rozpocząć (coś); iść na czele; odwieść kogoś (od czegoś)	Who is going to lead off ? Kto <i>rozpocznie</i> (kto <i>pójdzie na czele</i>)? I tried to lead him off from that. Próbowałem <i>odwieść</i> go od tego.
to make out	sporządzać (listę, dokument); wystawiać (czek); wypełniać (formularz); zrozumieć, pojąć; przedstawić; udawać	When will you make out a list of those books? Kiedy <i>sporządzisz</i> spis tych książek? Shall I make you out a receipt? Czy mam panu <i>wypisać</i> pokwitowanie? I can't make out what he wants. Nie mogę <i>zrozumieć</i> , o co mu chodzi. I daresay he's not so badly off as he makes out . Uważam, że nie powodzi mu się tak źle, jak on to <i>przedstawia</i> . I hope your affairs are making out well. Spodziewam się, że twoje sprawy <i>układają się</i> (przedstawiają się) dobrze.

Table 1: The treatment of phrasal verbs in Gajda's dictionary

Not every sense was illustrated. This would have taken too much space, but Gajda made a sensible selection of contextual uses.¹⁸ Typography played an important role here: the idioms in the citations were given in bold, while the equivalentents were italicised. This was a handy strategy by which to expose contrastive differences between the two languages, thanks to which semantic and grammatical explanations could be kept to an absolute minimum. The translation exercises that followed allowed the user to apply the idioms actively.

The fact that nearly 40% of the headphrases are phrasal verbs makes it difficult to apply Dobrovolskij's classification, but a number of Gajda's equivalents are shown below.

6.4 Searching for influences

Samples from both dictionaries show discrepancies, but there is also a degree of overlap. It is evident that both Borkowski and Gajda drew on the KFD. Borkowski, in particular, was under its influence, because he even copied the abbreviation *s.o.* for *someone*, which was never to become a lexicographical standard.

Table 2 below presents the treatment of several MWEs in both dictionaries. One has to admit that the Polish equivalents are similar, which might imply that Gajda had Borkowski's dictionary at his disposal. Gajda offers a slightly more extensive selection of English items, some of which are functional equivalents and some paraphrases. In three cases, the idioms were entered under different headwords.

MWE	Borkowski	Gajda
<i>take the air</i>	'wyjść na przechadzkę'	'wyjść na świeże powietrze'
<i>it beats everything! / it beats me</i>	'Coś podobnego! Coś niebywałego!'	'ja tego nie rozumiem; to przechodzi moje pojęcie'
<i>for the best</i>	'w najlepszej intencji'	'w najlepszej intencji; (obrócić się na dobre)'
<i>bone of contention</i>	'kość niezgody'	'przedmiot sporu; kość niezgody'
<i>to have money to burn</i>	'mieć pieniędzy jak lodu' (<i>money</i>)	'mieć dosyć pieniędzy (na zaspokajanie swoich kaprysów); mieć pieniądze jak lodu, w bród' (<i>burn</i>)
<i>face the music</i>	(pot) 'wyjść odważnie na spotkanie krytyce, trudnościom'	'stawić czoło trudnościom; ponieść konsekwencje czegoś'
<i>to break the news to sb</i>	'ostrożnie podać komuś złą wiadomość'	'zakomunikować komuś przykrą wiadomość (w sposób oględny)'
<i>to make one's flesh creep</i>	'wywołać ciarki na skórze' (<i>flesh</i>)	'wywoływać u kogoś ciarki, gęsia skórę' (<i>make</i>)
<i>live from hand to mouth</i>	'żyć z dnia na dzień, pchać biedę, klepać biedę' (<i>live</i>)	'żyć z dnia na dzień (bez widoków na przyszłość)' (<i>hand</i>)

Table 2: Examples of MWEs and Polish equivalents in Borkowski's and Gajda's dictionaries

We cannot date Gajda's interest in cross-linguistic phraseology to any specific point in time, but he was surely aware of Kobylański's *Wybór idiomów angielskich* (1951), the first dictionary of English idioms in post-war Poland. In fact, Gajda owes a huge debt to his predecessor; firstly, for the wording of the title; secondly, for the choice of L1 equivalents; and, thirdly, for the general lexicographical model of description, in which headwords lead to headphrases supported by quotations and juxtaposed with Polish translations. A sample of both dictionaries is shown in Table 3.

MWEs	Gajda	Kobylański
<i>with one's back to the wall</i>	'w sytuacji bez wyjścia; przyparty do muru'	'Przyparty do muru; w sytuacji bez wyjścia'
<i>to call up</i>	'telefonować (do kogoś); przypominać (coś); przywołać (coś) na pamięć'	'1. Zatelefonować; zadzwonić do kogo; 2. Wezwać do wojska; powołać; 3. Przywołać; wywołać (np. wspomnienia)'
<i>to make a fuss about [over] sth</i>	'robić wiele hałasu (zamieszania) o coś; robić o coś kwestię; robić ceremonie z czymś, kimś'	'1. Narzekać; wyrzekać; uskarżać się; 2. Robić co ostentacyjnie; robić co na pokaz; robić ceremonie'
<i>to let alone</i>	'zostawić w spokoju; nie ruszać, nie dotykać; pomijać'	'1. Zostawić kogo w spokoju; dać komu spokój; 2. Nie ruszać czego; nie dotykać czego; zostawić co w spokoju; 3. Nie mówiąc (już o), abstrahując od'
<i>to take on</i>	'przyjąć; zabrać; denerwować się; robić scenę'	'1. Przyjąć (kogo lub co); podjąć się (czego); 2. Zgodzić się być (czym) przeciwnikiem; 3. Brać sobie (co) do serca; przejmować się (czym); denerwować się (czym); robić scenę'

Table 3: Examples of MWEs and Polish equivalents in Gajda's and Kobylański's dictionaries

As can be seen, Kobylański was more methodical in that he numbered the senses of the idioms, a technique Gajda chose to ignore.

Gajda mentions the KFD as the only bilingual dictionary consulted. Indeed, there is sufficient evidence of borrowing, as exhibited in Table 4. Whenever Bulas's translations (e.g. *have a finger in every pie* 'p. umaczać w tym palec') and

his use of punctuation (e.g. *Make hay while the sun shines* 'kuj żelazo póki gorące') left much to be desired, Gajda corrected them in his own work.

MWEs	Gajda	KFD
<i>to burst one's sides with laughing</i>	'zrywać sobie boki ze śmiechu; śmiać się do rozpuku'	'zrywać sobie boki ze śmiechu; śmiać się do rozpuku'
<i>clear away</i>	'usuwać; sprzątać (ze stołu); ustępować (o mgle, chmurach itp.)'	'usuwać; sprzątać (ze stołu), (o mgle, dymie itp.) ustępować, rozchodzić się, t. usuwać się, odchodzić'
<i>turn the corner</i>	'skrócić na rogu ulicy; minąć punkt krytyczny (w chorobie itp.)'	'skrócić na rogu ulicy, p. minąć punkt krytyczny (w chorobie itp.)'
<i>Do by others as you would be done by</i>	'Nie czyń drugiemu, co tobie nie miło' ¹⁹	'nie czyń drugiemu, co tobie nie miło'
<i>have a finger in every pie</i>	'maczać we wszystkim palce; wtrącać się do cudzych spraw'	'p. umaczać w tym palec'
<i>Make hay while the sun shines</i>	'Kuj żelazo, póki gorące' (<i>make</i>)	'kuj żelazo póki gorące' (<i>hay</i>)
<i>show off</i>	'popisywać się; podkreślać; uwydatniać'	'podkreślać, uwydatniać, popisywać się (czymś)'
<i>be through with sb/sth</i>	'skończyć z czymś; mieć (już) czegoś dosyć; zerwać (znajomość) z kimś'	'skończyć z czymś, p. z kimś'

Table 4: Examples of MWEs and Polish equivalents in Gajda's dictionary and the KFD

Gajda also drew on Stanisławski's English–Polish dictionary newly launched onto the market (Table 5). The monolingual works of reference which he claims to have used explained the meanings of English idioms, but offered no Polish equivalents. Research reveals that even an outdated bilingual dictionary is more helpful to the lexicographer than a monolingual work, because "establishing equivalents on the basis of the latter involves guesswork and is prone to error" (Podhajecka 2016: 556). A bulky dictionary compiled by an experienced lexicographer may not, therefore, be overestimated in a project on bilingual phraseology. Let us look at the following examples:

MWEs	Gajda	Stanisławski
<i>to be on [off] one's guard (against sb/sth)</i>	'mieć [nie mieć] się na baczności (przed kimś, czymś); być [nie być] przygotowanym na coś'	mieć [nie mieć] się na baczności (<i>against sth</i> przed czymś); być [nie być] przygotowanym (<i>against sth</i> na coś)'
<i>head over heels</i>	'na łeb, na szyję; bez opamiętania; po uszy; bez pamięci'	'a) do góry nogami, b) na łeb na szyję; na gwałt; ~ <i>over heels in love</i> zakochany bez pamięci;
<i>to have sth at one's finger-ends [finger-tips]</i>	'znać coś na wylot; mieć coś w małym palcu'	'mieć coś w małym palcu; znać coś na wylot'
<i>to take back</i>	'odbierać (przyjmować) coś z powrotem; cofnąć (dane słowo); odwołać; przenieść wstecz; odwieść; zanieść z powrotem'	'1. od-ebrać/bierać (coś komuś (od kogoś)) z powrotem; wycofać/ywać (<i>sth from sth</i> coś z czegoś); cofnąć/ać (dane słowo (to, co się powiedziało)) 2. odprowadz-ić/ać (człowieka (zwierzę)) dokądś. 3. zan-ieść/osić z powrotem (<i>sth to sb</i> coś komuś)
<i>to be through with sb/sth</i>	'skończyć z czymś; mieć (już) czegoś dosyć; zerwać (znajomość) z kimś'	'a) skończyć coś b) nie potrzebować już czegoś; <i>I am ~ with him (her)</i> skończyłem (znajomość) z nim (nią)'

Table 5: Examples of MWEs and Polish equivalents in Gajda's and Stanisławski's dictionaries

In this case, Gajda made use not only of Stanisławski's equivalents, but also other types of lexicographical information, such as translations of English citations. Nonetheless, identifying the scope of borrowing is more difficult than it seems. There are reasons for thinking that Stanisławski used the KFD not only as a comparator, but also as a direct source of equivalents, even though he never admitted it openly.

Occasionally, Gajda's choice of Polish items also reflects his own linguistic preferences. This manifests itself, for instance, in the entries for *Handsome is what handsome does* 'O wartości człowieka świadczą jego uczynki, a nie jego wygląd' ('ten jest ładny kto ładnie postępuje' (KFD) vs. 'nie urodzenie stanowi o szlachetności' (Stanisławski)) and *to put one's foot in it* '... zrobić błąd; wywołać zakłopotanie' ('wpaść, zrobić gafę' (KFD) vs. 'popęlnić gafę' (Stanisławski)).

6.5 Summary of the case study

The two different structural patterns discussed here are indicative of the search for an effective model for presenting idioms in dictionary form.

Borkowski's methodology consisted in collecting as many MWEs as he could in the hope of providing the target user with a fully exhaustive set, however precarious it might have been. As an 'enlightened amateur', he gave little consideration to the consistency of his selection; his dictionary included a mixture of single words and word combinations of different status, only a proportion of which are idiomatic in character. Despite this phraseological mishmash, *An English–Polish Dictionary of Idioms and Phrases* represented a contribution to a network of specialised materials for Polish learners of English, specifically immigrants wishing "to advance in the social hierarchy" [M.P.] of their adopted countries. Still, the fact that the MWEs are hard to find was a serious disadvantage that compromised the usefulness of the volume. One might assume that, with no examples of usage, the dictionary was intended primarily for receptive tasks. Since the pool of modern English–Polish works was limited, Borkowski consulted the KFD, a brand-new endeavour, to ensure that his dictionary would be up-to-date.

While Gajda's dictionary is inferior to Borkowski's in terms of quantity, it is evidently superior in quality, in particular, by being better tuned to the needs of foreign language learners. It was salient, because Gajda focused on the learning process entirely disregarded by his competitor. He, therefore, provided Polish equivalents for the idioms, Polish translations for the citations, and exercises aimed at consolidating the user's knowledge of English phraseology. Clear emphasis on the learning process through translation from and into English, as well as an ample choice of Polish functional equivalents and paraphrases, suggests that the dictionary was to be applied both for decoding and encoding purposes. Turning to monolingual sources and the KFD was a guarantee that the bilingual material would be free from gross errors. What remains odd in this otherwise laudatory overview of Gajda's dictionary is his unacknowledged debt to Kobyłański (1951) and Stanisławski (1964).

Borkowski and Gajda opted for a different type of dictionary, but their reasons for doing so remain undisclosed. I will attempt to explain their decisions by taking a broader view. Although both of them pursued similar professional careers, their experiences were dissimilar. The Second World War saw Borkowski move to Britain, while Gajda spent all his life in Poland. An enthusiast of English–Polish phraseology, Borkowski had, however, no expert knowledge to help him create an acceptable bilingual dictionary and had no one to turn to for advice. This comes as little surprise. At the turn of the twentieth century, Europe became the battleground for new ideas in foreign language teaching known as the Reform Movement (1880–1920). In Britain, during the so-called Scientific Period (1920–1970), all the methods were confined to the teaching of English in English without resorting to the learners' mother tongues

(e.g. Howatt and Smith 2014). Linguists, therefore, frowned upon both bilingual instruction and bilingual aids. Gajda was an autodidact, which presupposes his learning English with bilingual materials, including English–Polish and Polish–English dictionaries, that dominated the Polish educational market during the nineteenth and twentieth centuries. It is conceivable that Gajda contacted a specialist in English, who might be credited with the idea of incorporating the exercises, for guidance in the case of problematic issues²⁰ and he also had Kobyłański's volume to hand, whose methodology he followed closely. The two different sociocultural contexts thus appear to have been at play.

7. Conclusions

This paper reports on the findings of an analysis of two English–Polish phraseological dictionaries by Piotr Borkowski (1963) and Roman Gajda (1970) carried out in the fields of comparative (meta)lexicography and sociocultural (meta)lexicography, of which the latter is fundamental for studies of dictionary history.

It is time we answered the questions posed in Section 2. Let us begin with those concerning comparative (meta)lexicography. Dictionaries may be evaluated with different sampling methods, but analysis of the entire material provides the most comprehensive results. The techniques for contrasting general and special-purpose dictionaries are superficially similar, with the proviso that only selected information categories (e.g. etymology, pronunciation, or slang) are considered in the latter. Special-purpose dictionaries cover, among other things, phraseological dictionaries. The monolingual subtypes are compared to show the MWEs they record and the way the meanings and usage are described for the sake of the target users. In the case of the bilingual subtypes, the cross-linguistic perspective additionally allows for assessments of the quality of target language equivalents. Since both subtypes record word combinations instead of single-word lemmas, their access structures also qualify for comparison (e.g. Buendía Castro and Faber 2014).

Moving to questions related to sociocultural (meta)lexicography, the compilers' familiarity with the environments is often taken for granted, and rightly so. Not only does it ensure a knowledge of the dictionary market, major competitors, and lexicographical trends in vogue, but it is also instrumental in providing access to experts, informants, supporters, sponsors, and user groups. In any case, compiling a bilingual dictionary involves a complex decision-making process and compiling a bilingual phraseological dictionary is no less laborious (e.g. Lubensky and McShane 2007: 920). The range of problems faced induces the compilers to look for inspiration in their predecessors' works, some of which are found in prefaces, while others remain hidden from the public, of which Gajda's dictionary is a case in point. Once the lexicographical context has been examined, potential sources need to be singled out for comparative scrutiny. It may reveal both striking affinities and marked contrasts, if only because

dictionaries communicate with lexicographical traditions in two ways: either by following similar models of description or by rejecting them.²¹ The compilers' motives, real or assumed, are difficult to pinpoint and must therefore be subjected to thorough biographical research.

The choice of materials and methods in dictionary criticism admittedly involves a great deal of flexibility. What may be compared are editions of the same dictionary (e.g. Kamiński 2013); opposite genres, such as monolingual and bilingual dictionaries (e.g. Piotrowski 1989); or dictionaries belonging to lexicographical traditions of different languages (e.g. Zgusta 1986). No wonder the comparative perspective often aims at capturing developmental trends in lexicography (e.g. Liberman 1998; Hüllen 1999; Cormier and Francoeur 2004). Dictionaries also undergo comparison in pursuit of common features, such as social values (e.g. Landau 1985) and user-friendliness (e.g. Dziemińko 2012), or to expose variation, such as differences in the treatment of gender (e.g. Norri 2019), metonymy (e.g. Wojciechowska 2012), and idioms (e.g. Pinnavaia 2010). Such studies tend to combine the comparative approach of dictionary analysis with approaches typical of other disciplines (e.g. cognitive linguistics or corpus linguistics). It is important that all evaluations follow a clear and unbiased methodology and be "drawn from an expert knowledge of dictionaries, dictionary making and dictionary use" (Jackson 2002: 183), a fairly reasonable requirement.

Sociocultural (meta)lexicography is based on the assumption that analyses should not be restricted to the dictionary text. As Yagello (qtd. in Mackintosh 2006: 60) cogently puts it, "the dictionary is an ideological creation. It is a mirror of society and of the dominant ideology. As an indisputable authority and a cultural tool, the dictionary plays a part in establishing and preserving language, but also attitudes and ideology [...] every revolution should be accompanied by dictionary reform". This points to the dynamic and multifaceted interplay between dictionaries, their compilers, and their sociocultural settings. Scholarly dictionaries, including those "repeatedly presented as heroic works" (Considine 2008: 3), have been associated with noble purposes, but dictionaries are also profit-oriented commodities (e.g. Whitcut 1989). The 'war of the dictionaries' between the American lexicographers Noah Webster and Joseph Worcester, for instance, was "fought over who could best represent the soul and identity of American culture", as the blurb of Martin's (2020) book tells us, but money, too, was a crucial factor.

Recent decades have witnessed an upsurge of interest in comparative (meta)lexicography and sociocultural (meta)lexicography, although the latter term remains infrequent.²² Theoretically speaking, applying a tried and tested research framework and combining it with a relevant methodology is all that is necessary. Practice, however, is more complicated. It shows that there are virtually no limits to what aspects of dictionaries may be contrasted, in what formats, along what criteria, and for what purposes, depending on the (meta)lexicographers' conceptions. Consequently, the diversity of approaches and variables

means that providing any guidelines, let alone explicit guidelines, is a sheer impossibility.

Acknowledgements

I would like to thank two anonymous reviewers and the editor of this volume for their constructive comments and suggestions on the first draft of this paper.

Endnotes

1. Ilson (1986) names this type of comparison *lexicographic archeology*.
2. It is worthy of note that the terminology is somewhat vague, because neither *evaluation* nor *dictionary criticism* implies the comparative approach.
3. Generalisations are recommended, but Steiner (1984: 179) also allows more impressionistic observations, such as supplying "entries which should have been included, choices of words and equivalents which are better than those of the compilers, typos which the editors let slide, and lapses by the compilers themselves".
4. Dobrowol'skij's classification closely parallels that introduced by Baker (1992: 68-71) in the translational context.
5. Attempts to identify them have been unsuccessful, which suggests that they were discarded.
6. Collison's argues (1982: 17) that "throughout the whole history of the making of language dictionaries the part played by amateurs is very striking".
7. He was the long-time head of the Department of English of Adam Mickiewicz University in Poznań (1965–2005) and a renowned lexicographer himself.
8. Generally speaking, there is relatively little resemblance between Gajda's and Worrall's dictionaries, the latter being organised thematically, or between Gajda's and Crowell's, the latter focusing specifically on prepositions.
9. The latter comes from Apperson's *Dictionary of Proverbs* (1993). It first appeared in John Ray's *Handbook of Proverbs* (1670) and its status seems to have been unchallenged since then.
10. This is definitely true of the KFD. Gajda contacted the Kosciuszko Foundation in March 1963. In his reply of 22 April 1963, Stephen Mizwa, the Foundation's President, granted him consent to reproduce illustrative material from the KFD (Archives of the Kosciuszko Foundation, XX.13 Kazimierz Bulas).
11. The last two categories are by no means idiomatic. Borkowski's Russian–English dictionary was also criticised for his haphazard selection of MWEs (Arsenteva and Ayupova 2013: 67).
12. Cowie, Mackin, and McCaig's *Oxford Dictionary of English Idioms* (1993) entered the idioms by the first content word, but the principle may have been somewhat hindered by lexical variation (e.g. *give/lose ground (to sb/sth)*, *(my) goodness (me)!*, *(a matter etc of) life and/or death or (those) who the gods love, die young*).
13. Borkowski was aware of the difficulty of arranging idioms neatly and he tried to do something about it. In his 1974 book, he disclosed a system according to which they should be placed by the first noun, verb, adjective, or, in the absence of these parts of speech, simply by the first word (see Gribble 1976: 780). A similar system was used in the second edition of *Longman Dictionary of the English Language* (Jackson 2002: 100).

14. In this paper, small capitals have been replaced with standard print in italics.
15. Theoretically speaking, phraseological dictionaries pairing L1 idioms with L2 equivalents should be easy to find, but they are a rarity. In the Polish context, "there are very few Polish–English dictionaries of this kind, vastly outnumbered by English–Polish ones" (Szerszunowicz 2014: 2).
16. The translation equivalent was again borrowed from the KFD (cf. Stanislawski's *she would talk the hind leg of a donkey* 'ona gada tyle, że aż głowa puchnie; ona każdego przegada').
17. These were Henryk Kaluża, Janina Smólska, and Oskar Chomicki. Smólska was a well-known author of textbooks and grammars for Polish learners of English.
18. Some of the citations (e.g. *I hope your affairs are making out well*), however, look bookish.
19. Here, Gajda took over even the erroneously spelled Polish adverb 'nie miło' (> 'niemiło').
20. Tadeusz Grzebieniowski, affiliated with the University of Łódź (1945–1953; 1957–1964), seems to be a most suitable candidate. Not only was he an esteemed English scholar, but he also made his name as an author of English textbooks for Poles (e.g. 1947a, 1947b) and a bilingual lexicographer (e.g. 1950).
21. It may be something of a paradox, but compilers experimenting with innovative methods are often aware of the lexicographical tradition they leave behind.
22. Sociocultural (meta)lexicography as a theoretical branch, however, is not reflective of *cultural lexicography*, which Hartmann and James (2001: 33) define as 'a complex of activities concerned with the design, compilation, use and evaluation of cultural dictionaries [...]'.

References

Dictionaries

- Antonowicz, J.** 1788. *Słownik potrzebniejszych wyrazów* [A Dictionary of the Most Necessary Words]. *Grammatyka dla Polaków chcących się uczyć angielskiego języka krótko zebrana przez ... Bazyliańską prowincję litewską za pozwoleniem zwierzchności pierwszy raz pod prasę oddana* [A Grammar for Poles Wishing to Learn the English Language Compiled Briefly by ... a Basilian of the Lithuanian Province, Submitted to the Press for the First Time with the Consent of the Authorities]: 106-133. Warszawa: W Drukarni Nadwornej J. K. Mci i P. K. Edu.
- Apperson, G.L.** 1993. *Dictionary of Proverbs*. Ware: Wordsworth.
- Borkowski, P.** 1963. *An English–Polish Dictionary of Idioms and Phrases / Angielsko–polski słownik idiomów i zwrotów*. London: P. Borkowski.
- Borkowski, P.** 1973. *The Great Russian–English Dictionary of Idioms and Set Expressions*. London: Piotr Borkowski.
- Bulas, K. and F. Whitfield.** 1959. *The Kosciuszko Foundation Dictionary*. Vol. 1: English–Polish. The Hague: Mouton.
- Bulas, K., F. Whitfield and L. Thomas.** 1961. *The Kosciuszko Foundation Dictionary*. Vol. 2: Polish–English. The Hague: Mouton.
- Cowie, A.P., R. Mackin and I.R. McCaig.** 1993. *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.
- Cromwell, T.L.** 1960. *A Glossary of Phrases with Prepositions*. Wrocław: Zakład Narodowy im. Ossolińskich.
- Gajda, R.** 1970. *Wybór idiomów angielskich* [A Selection of English Idioms]. Warszawa: Wiedza Powszechna.
- Grzebieniowski, T.** 1950. *Słownik angielsko–polski* [An English–Polish Dictionary]. Warszawa: Czytelnik.

- Hartmann, R.R.K. and G. James.** 2001. *Dictionary of Lexicography*. London/New York: Routledge.
- Henderson, B.L.K.** 1937. *A Dictionary of English Idioms*. Vols. 1–2. London: James Blackwood.
- Hornby, A.S.** 1948. *The Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- Kobyłański, M.** 1951. *Wybór idiomów angielskich [A Selection of English Idioms]*. Warszawa: Państwowy Zakład Wydawnictw Szkolnych.
- Richter, A.** 1945. *Polish Dictionary of English Idioms, Proverbs and Slangs / Słownik zwrotów angielskich*. Tel-Aviv: Wydawnictwo Książek Polskich "Świt".
- [Rykaczewski, E.]** 1849. *A Complete Dictionary English and Polish, and Polish and English Compiled from the Dictionaries of Johnson, Webster, Walker, Fleming and Tibbins, etc., from the Polish Lexicon of Linde and the Polish–German Dictionary by Mrongovius ...* Vol. 1. Berlin: B. Behr.
- [Rykaczewski, E.]** 1851. *Dokładny słownik polsko–angielski i angielsko–polski, czerpany z najlepszych źródeł krajowych i obcych; a mianowicie ze słowników polskich: Lindego, Mrongoviusa i Ropelewskiego; z angielskich: Johnson, Webster, Walker, Fleming-Tibbins i innych. Ten tom polsko–angielski ...* Vol. 2. [A Complete Dictionary Polish and English and English and Polish, Compiled from the Best Domestic and Foreign Sources, that is, the Polish Lexicons of Linde, Mrongovius and Ropelewski, and the English Dictionaries of Johnson, Webster, Walker, Fleming, Tibbins, etc. Volume [2]: Polish–English ...] Berlin: B. Behr.
- Stanisławski, J.** 1964. *The Great English–Polish Dictionary*. Vol. 1. Warszawa: Wiedza Powszechna.
- Stanisławski, J.** 1969. *The Great Polish–English Dictionary*. Vol. 2. Warszawa: Wiedza Powszechna.
- Stanisławski, J. and M. Szercha.** 1975. *Wielski słownik angielsko–polski A–Z: Suplement [A Great English–Polish Dictionary A–Z: A Supplement]*. Warszawa: Wiedza Powszechna.
- Stanisławski, J. and M. Szercha.** 1977. *Wielski słownik polsko–angielski A–Z: Suplement [A Great Polish–English Dictionary A–Z: A Supplement]*. Warszawa: Wiedza Powszechna.
- Worrall, A.J.** 1932. *English Idioms for Foreign Students*. London/New York: Longmans.

Other references

- Adamska-Sałaciak, A.** 2016. Continuity and Change in *The (New) Kosciuszko Foundation Dictionary*. *Studia Anglica Posnaniensia* 51(1): 83-98.
- Akasu, K.** 2022. Methods in Dictionary Criticism. Jackson, H. (Ed.). 2022. *The Bloomsbury Handbook of Lexicography*: 31-42. Second edition. London/New York: Bloomsbury Academic.
- Arsenteva, E. and R. Ayupova.** 2013. Bilingual and Multilingual Russian–English Phraseological Dictionaries. *Yearbook of Phraseology* 4(1): 65-78.
- Atkins, B.T.** 1985. Monolingual and Bilingual Learners' Dictionaries: A Comparison. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 15-24. Oxford: Pergamon Press/The British Council.
- Baker, M.** 1992. *In Other Words: A Coursebook on Translation*. London/New York: Routledge.
- Bately, J.** 2009. Bilingual and Multilingual Dictionaries of the Renaissance and Early Seventeenth Century. Cowie, A. (Ed.). 2009. *The Oxford History of English Lexicography*. Vol. 1: General-Purpose Dictionaries: 41-64. Oxford: Oxford University Press.
- Benson, P.** 2002. *Ethnocentrism and the English Dictionary*. London/New York: Routledge.
- [Biography of Roman Gajda]. 1983. *Fantastyka [Fantasy]* 11–12: 80.
- BNC = The British National Corpus.** Available at: <https://www.english-corpora.org/bnc/>.

- Bogaards, P.** 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.
- Borkowski, P.** [1953]. *Praktyczny poradnik podatkowy* [A Practical Tax Guidebook]. Londyn: [The Author].
- Borkowski, P.** 1974. *O językach i słownikach na wesoło* [A Humorous Look at Languages and Dictionaries]. Londyn: Nakładem Polskiej Fundacji Kulturalnej.
- Buendía Castro, M. and P. Faber.** 2014. Collocation Dictionaries: A Comparative Analysis. *MonTI* 6: 203-235.
- Chen, W.** 2015. Bilingual Lexicography as Recontextualization: A Case Study of Illustrative Examples in *A New English–Chinese Dictionary*. *Australian Journal of Linguistics* 35(4): 311-333.
- Chen, Y.** 2010. Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.
- Coleman, J. and S. Ogilvie.** 2009. Forensic Dictionary Analysis: Principles and Practice. *International Journal of Lexicography* 22(1): 1-22.
- Collison, R.L.** 1982. *A History of Foreign-Language Dictionaries*. London: Andre Deutsch.
- Considine, J.** 2008. *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage*. Cambridge: Cambridge University Press.
- Considine, J.** 2014. *Academy Dictionaries 1600–1800*. Cambridge: Cambridge University Press.
- Considine, J.** 2022. *Sixteenth-Century English Dictionaries*. Oxford: Oxford University Press.
- Cormier, M.C.** 2003. From the *Dictionnaire de l'Académie Française*, dédié au Roy (1694) to the *Royal Dictionary* (1699) of Abel Boyer: Tracing Inspiration. *International Journal of Lexicography* 16(1): 19-41.
- Cormier, M.C.** 2009. Bilingual Dictionaries of the Late Seventeenth and Eighteenth Centuries. Cowie, A. (Ed.). 2009. *The Oxford History of English Lexicography*. Vol. 1: General-Purpose Dictionaries: 65-85. Oxford: Oxford University Press.
- Cormier, M.C. and H. Fernández.** 2004. Influence in Lexicography: A Case Study. Abel Boyer's *Royal Dictionary* (1699) and Captain John Stevens' *Dictionary English and Spanish* (1705). *International Journal of Lexicography* 17(3): 291-308.
- Cormier, M.C. and A. Francoeur.** 2004. French–English Bilingual Dictionaries in the 16th and 17th Centuries: Achievements and Innovations by French-speaking Protestants. *Dictionaries: Journal of the Dictionary Society of North America* 25: 77-106.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- Cowie, A.P.** 2001. Phraseological Dictionaries: Some East-West Comparisons. Cowie, A.P. (Ed.) 2001. *Phraseology. Theory, Analysis, and Applications*: 209-228. Oxford: Oxford University Press.
- Dobrovol'skij, D.** 2000. Contrastive Idiom Analysis: Russian and German Idioms in Theory and in the Bilingual Dictionary. *International Journal of Lexicography* 13(3): 169-186.
- Dobrovol'skij, D.** 2013. German–Russian Idioms Online: On a New Corpus-Based Dictionary. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue 2013"* 12(19): 210-217.
- Dziemianko, A.** 2012. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Berlin/Boston: Walter de Gruyter.
- Espinal, M.T. and J. Mateu.** 2019. Idioms and Phraseology. Aronoff, M. (Ed.). 2019. *Oxford Research Encyclopedia of Linguistics*. New York, NY: Oxford University Press.
- Farina, D.M. and G. Durman.** 2009. Bilingual Dictionaries of English and Russian in the Eighteenth to the Twentieth Centuries. Cowie, A. (Ed.). 2009. *The Oxford History of English Lexicography*. Vol. 1: General-Purpose Dictionaries: 105-126. Oxford: Oxford University Press.

- Fisiak, J.** 1989. Przedmowa do wydania polskiego [Preface to the Polish Edition]. Borkowski, P. 1989. *Angielsko-polski słownik idiomów i zwrotów / An English-Polish Dictionary of Idioms and Phrases*: 3-5. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.
- Fishman, J.A.** 1995. Dictionaries as Culturally Constructed and as Culture-Constructing Artifacts: The Reciprocity View as Seen from Yiddish Sources. Kachru, B.B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary. Studies in Honor of Ladislav Zgusta*: 29-34. Lexicographica. Series Maior 64. Tübingen: Max Niemeyer.
- Fontenelle, T.** 2001. Discovering Significant Lexical Functions in Dictionary Entries. Cowie, A.P. (Ed.). 2001. *Phraseology. Theory, Analysis, and Applications*: 189-207. Oxford: Oxford University Press.
- Frączek, A.** 1999. *Zur Geschichte der deutschen-polnischen und polnisch-deutschen Lexikographie (1772-1868)*. Lexicographica. Series Maior 93. Tübingen: Max Niemeyer.
- Gajda, R.** 1957. *Ludzie ery atomowej: powieść fantastyczno-naukowa [People of the Atomic Era: A Science-Fiction Novel]*. Łódź: "Iskry".
- Granger, S. and F. Meunier.** 2008. Introduction: The Many Faces of Phraseology. Granger, S. and F. Meunier (Eds.). 2008. *Phraseology: An Interdisciplinary Perspective*: xix-xxviii. Amsterdam/Philadelphia: John Benjamins.
- Grant, L. and L. Bauer.** 2004. Criteria for Re-defining Idioms: Are We Barking up the Wrong Tree? *Applied Linguistics* 25(1): 38-61.
- Gribble, C.E.** 1976. Review of *The Great Russian-English Dictionary of Idioms and Set Expressions ... The Slavic Review* 35(4): 780-781.
- Grodziska, K.** 2001. *Polskie groby na cmentarzach Londynu [Polish Graves in London's Cemeteries]*. Vol. 2. Kraków: Polska Akademia Umiejętności.
- Grzebieniowski, T.** 1947a. *Język angielski na co dzień: podręcznik mowy potocznej do samodzielnej nauki dla zaawansowanych z kluczem [English for Everyday Use: A Self-Study Book of Colloquial Speech for Advanced Learners with a Key]*. Kraków: Wiedza, Zawód, Kultura.
- Grzebieniowski, T.** 1947b. *A Primer of Everyday English: kurs języka angielskiego dla początkujących i samouków [An English Course for Beginners and Self-Study Learners]*. Warszawa: "Czytelnik".
- Harras, G. and K. Proost.** 2005. The Lemmatisation of Idioms. Gottlieb, H., J.E. Mogensen and A. Zettersen (Eds.). 2005. *Symposium on Lexicography XI: Proceedings of the Eleventh International Symposium on Lexicography* May 2-4, 2002 at the University of Copenhagen: 277-291. Tübingen: Max Niemeyer.
- Howatt, A.P.R. and R. Smith.** 2014. The History of Teaching English as a Foreign Language, from a British and European Perspective. *Language & History* 57(1): 75-95.
- Hüllen, W.** 1999. *English Dictionaries, 800-1700. The Topical Tradition*. Oxford: Oxford University Press.
- Ilsion, R.** 1986. Lexicographic Archaeology: Comparing Dictionaries of the Same Family. Hartmann, R.R.K. (Eds.). 1986. *The History of Lexicography. Papers from the Dictionary Research Centre Seminar at Exeter, March 1986*: 127-136. Amsterdam/Philadelphia: John Benjamins.
- Jackson, H.** 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Kamiński, M.P.** 2013. *A History of the Chambers Dictionary*. Lexicographica Series Maior 143. Berlin/Boston: Walter de Gruyter.
- Kerling, J.** 1979. *Chaucer in Early English Dictionaries: The Old-word Tradition in English Lexicography down to 1721 and Speght's Chaucer Glossaries*. Leiden: Leiden University Press.
- Kister, K.F.** 1992. *Kister's Best Dictionaries for Adults and Young People: A Comparative Guide*. Phoenix, AZ: Oryx Press.

- Laks, S.** 1978. *Stowa i antysłowa [Words and Anti-words]*. Londyn: Oficyna Poetów i Malarzy.
- Landau, S.I.** 1985. The Expression of Changing Social Values in Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 7: 261-269.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.
- Landau, S.I.** 2009. Major American Dictionaries. Cowie, A. (Ed.). 2009. *The Oxford History of English Lexicography*. Vol. 1: General-Purpose Dictionaries: 182-229. Oxford: Oxford University Press.
- Lew, R. and A. Szarowska.** 2017. Evaluating Online Bilingual Dictionaries: The Case of Popular Free English–Polish Dictionaries. *ReCALL* 29(2): 138–159.
- Lieberman, A.** 1998. An Annotated Survey of English Etymological Dictionaries and Glossaries. *Dictionaries: Journal of the Dictionary Society of North America* 19: 21-96.
- Lubensky, S. and M. McShane.** 2007. Bilingual Phraseological Dictionaries. Burger, H., D. Dobrovolskij, P. Kühn and N.R. Norrick (Eds.). 2007. *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung*. Vol. 2: 919-928. Berlin/New York: Walter de Gruyter.
- Mackintosh, K.** 2006. Biased Books by Harmless Drudges: How Dictionaries are Influenced by Social Values. Bowker, L. (Ed.). 2006. *Lexicography, Terminology, and Translation: Text-based Studies in Honour of Ingrid Meyer*: 45-63. Ottawa: The University of Ottawa Press.
- Martin, P.** 2020. *The Dictionary Wars: The American Fight over the English Language*. Princeton: Princeton University Press.
- Masuda, H. et al.** 1999. A Comparative Analysis of the Two American Learner's Dictionaries — NHD and RHW. *Lexicon* 29: 1-65.
- McConchie, R.W.I.** 2019. *Discovery in Haste. English Medical Dictionaries and Lexicographers 1547 to 1796*. Berlin/New York: De Gruyter.
- Mel'čuk, I.** 2012. Phraseology in the Language, in the Dictionary, and in the Computer. *Yearbook of Phraseology* 3(1): 31-56.
- Miller, J.** 2013. *Birds of a Feather Don't Always Flock Together: User Problems in Identifying Headwords in Online English Learner's Dictionaries*. *Lexikos* 23: 273-285.
- Miyoshi, K.** 2007. *Johnson's and Webster's Verbal Examples: With Special Reference to Exemplifying Usage in Dictionary Entries*. Lexicographica. Series Maior 132. Tübingen: Max Niemeyer.
- Moon, R.** 1999. Needles and Haystacks, Idioms and Corpora: Gaining Insights into Idioms, Using Corpus Analysis. Herbst, T. and K. Popp (Eds.). 1999. *The Perfect Learner's Dictionary (?)*: 265-281. Tübingen: Max Niemeyer.
- Moon, R.** 2001. Frequencies and Forms of Phrasal Lexemes in English. Cowie, A.P. (Eds.). 2001. *Phraseology. Theory, Analysis, and Applications*: 79-100. Oxford: Oxford University Press.
- Moon, R.** 2015. Idioms: A View from the Web. *International Journal of Lexicography* 28(3): 318-337.
- Mugglestone, L.** 2011. *Dictionaries: A Very Short Introduction*. Oxford: Oxford University Press.
- Nakamoto, K.** 1995. A Checklist for Reviewers of EFL Dictionaries: Checkpoints about the Review. *Lexicon* 25: 1-13.
- Nielsen, S.** 2009. Reviewing Printed and Electronic Dictionaries: A Theoretical and Practical Framework. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century: In Honour of Henning Bergenholtz*: 23-42. Amsterdam/Philadelphia: John Benjamins.
- Nkomo, D.** 2016. An African User-Perspective on English Children's and School Dictionaries. *International Journal of Lexicography* 29(1): 31-54.
- Norri, J.** 2019. Gender in Dictionary Definitions: A Comparison of Five Learner's Dictionaries and Their Different Editions. *English Studies* 100(7): 866-890.

- [Obituary of Piotr Borkowski]. 1986. *Pamiętnik Literacki* [A Literary Memoir] 10: 78.
- Ogilvie, S. 2013. *Words of the World. A Global History of the Oxford English Dictionary*. Cambridge: Cambridge University Press.
- Pinnavaia, L. 2010. *Sugar and Spice ... Exploring Food and Drink Idioms in English*. Monza: Polimetrica.
- Piotrowski, T. 1989. Monolingual and Bilingual Dictionaries: Fundamental Differences. Tickoo, M.L. (Ed.). 1989. *Learners' Dictionaries: State of the Art*: 72-83. Singapore: SEMAEO Regional Language Centre.
- Piotrowski, T. 2001. *Zrozumieć leksykografię* [Understanding Lexicography]. Warszawa: Wydawnictwo Naukowe PWN.
- Piirainen, E. 2008. Phraseology in a European Framework: A Cross-Linguistic and Cross-Cultural Research Project on Widespread Idioms. Granger, S. and F. Meunier (Eds.). 2008. *Phraseology: An Interdisciplinary Perspective*: 243-258. Amsterdam/Philadelphia: John Benjamins.
- Podhajecka, M. 2013. Tracing the Sources of Krystyn Lach-Szyrma's English-Polish Dictionary (1828): Inspired or Borrowed? McConchie, R.W. et al. (Eds.). 2013. *Selected Proceedings of the 2012 Symposium on New Approaches in English Historical Lexis (HEL-LEX 3)*: 141-158. Somerville, MA: Cascadilla Proceedings Project.
- Podhajecka, M. 2016. *A History of Polish-English / English-Polish Bilingual Lexicography (1788-1947)*. Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Podhajecka, M. 2020. The Earliest English-Polish and Polish-English Phraseological Dictionaries: A Preliminary Survey. *Rassegna Italiana di Linguistica Applicata* (RILA) 52(2-3): 121-136.
- Reddick, A. 2009. Johnson and Richardson. Cowie, A. (Ed.). 2009. *The Oxford History of English Lexicography*. Vol. 1: General-Purpose Dictionaries: 155-181. Oxford: Oxford University Press.
- Rennie, S. 2012. *Jamieson's Dictionary of Scots: The Story of the First Historical Dictionary of the Scots Language*. Oxford: Oxford University Press.
- Solano, R.M. 2013. From Idiom Variants to Open-Slot Idioms: Close-Ended and Open-Ended Variational Paradigms. Szerszunowicz, J. et al. (Eds.). 2013. *Research on Phraseology across Continents*. Vol. 2: 149-162. Białystok: University of Białystok Publishing House.
- Starnes, D.T. 1963. *Robert Estienne's Influence on Lexicography*. Austin: University of Texas Press.
- Stein, G. 1985. *The English Dictionary before Cawdrey*. Lexicographica. Series Maior 9. Tübingen: Max Niemeyer.
- Stein, G. 2014. *Sir Thomas Elyot as Lexicographer*. Oxford: Oxford University Press.
- Steiner, R.J. 1970. *Two Centuries of Spanish and English Bilingual Lexicography, 1590-1800*. The Hague/Paris: Mouton.
- Steiner, R.J. 1984. Guidelines for Reviewers of Bilingual Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 6: 166-181.
- Swanepoel, P. 2013. Evaluation of Dictionaries. Gouws, R. et al. (Eds.). 2013. *Dictionaries: An International Encyclopedia of Lexicography*. Supplementary Volume: 587-596. Berlin/Boston: Mouton de Gruyter.
- Szczepaniak, R. 2012. *The Role of Dictionary Use in the Comprehension of Idiom Variants*. Lexicographica. Series Maior 131. Berlin/New York: Walter de Gruyter.
- Szczepaniak, R. and A. Adamska-Salaciak. 2010. Telling it Straight: A Comparison of Selected English and Polish Idioms from the Semantic Field of Speaking. *Studia Anglica Posnaniensia* 46(1): 71-93.
- Szerszunowicz, J. 2006. Pseudo-equivalents in English, Italian and Polish Faunal Phraseology. Corino, E., C. Marelllo and C. Onesti (Eds.). 2006. *Proceedings of 12th EURALEX International Congress, Torino, Italia, September 6-9, 2006*: 1055-1060. Alessandria: Edizioni Dell'Orso.

- Szerszunowicz, J.** 2014. Phraseological Units Containing Archaic Elements in Bilingual Lexicography. G. Zuckermann, J. Miller and J. Morely (Eds.). 2014. *Endangered Words, Signs of Revival*: 1-17. Adelaide: AustraLex. Available at: <https://www.adelaide.edu.au/australlex/conferences/2013/szerszunowicz.pdf>.
- Szerszunowicz, J.** 2016. *Lakunarne jednostki wielowyrzowe w perspektywie międzyjęzykowej* [Multi-Word Lacuna Expressions in a Cross-Linguistic Perspective]. Białystok: Uniwersytet w Białymstoku.
- Tomaszczyk, J.** 1988. The Bilingual Dictionary under Review. Snell-Hornby, M. (Ed.). 1988. *Zürich-LEX '86 Proceedings, Papers Read at the EURALEX International Congress, University of Zürich, 9–14 September 1986*: 289-297. Tübingen: A. Francke Verlag.
- Vişan, R.** 2018. A "Florid" Preface about "a Language that is Very Short, Concise and Sententious". McConchie, R. and J. Tyrkkö (Eds.). 2018. *Historical Dictionaries in Their Paratextual Context*. *Lexicographica. Series Maior* 153: 285-305. Berlin/Boston: De Gruyter.
- Whitcut, J.** 1989. The Dictionary as a Commodity. Hausmann, F.J. et al. (Eds.). 1989. *Dictionaries. An International Encyclopedia of Lexicography*. Vol. 1: 111-122. Berlin: De Gruyter.
- Wojciechowska, S.** 2012. *Conceptual Metonymy and Lexicographic Representation*. Frankfurt am Main: Peter Lang.
- Yong, H. and J. Peng.** 2021. *A Sociolinguistic History of British English Lexicography*. London: Routledge.
- Zakrzewski, P.** 2002. W sprawie definicji idiomu [On the Definition of 'Idiom']. *Acta Universitatis Lodzianensis. Folia Linguistica* 42: 19-23.
- Zgusta, L.** 1986. Grimm, Littré, OED, and Richardson: A Comparison of Their Historicity: *Cātuškoşyam. Dictionaries: Journal of the Dictionary Society of North America* 8: 74-93.

Appendix 1: The number of MWEs across Borkowski's and Gajda's dictionaries

Letter	Borkowski's headphrases	%	Letter	Gajda's headphrases	%
A	168	3.8	A	22	3.30
B	292	6.73	B	60	9.06
C	282	6.49	C	61	9.21
D	245	5.64	D	66	9.97
E	140	3.22	E	24	3.62
F	249	5.74	F	78	11.78
G	146	3.36	G	50	7.55
H	251	5.78	H	48	7.25
I	64	1.47	I	–	–
J	66	1.52	J	4	0.60
K	51	1.17	K	18	2.72
L	279	6.43	L	13	1.96
M	229	5.27	M	18	2.72
N	131	3.02	N	1	0.15
O	108	2.49	O	16	2.42
P	170	3.92	P	23	3.47
Q	43	0.99	Q	–	–
R	253	5.83	R	16	2.42
S	415	9.56	S	46	6.95
T	256	5.90	T	68	10.27
U	62	1.43	U	2	0.30
V	53	1.22	V	3	0.45
W	337	7.76	W	25	3.78
X	–	–	X	–	–
Y	49	1.13	Y	–	–
Z	2	0.05	Z	–	–
Total	4,341	100	Total	662	100

Popma's Monolingual Dictionary of Latin Synonyms

Nikita Murzintcev, *New Uzbekistan University, Tashkent, 100007,
Uzbekistan (n.murzincev@newuzbekistanuniversity.uz)*

Abstract: This report presents the results of our project of digitizing a dictionary of Latin synonyms which was popular for three centuries from XVII to XIX; describes its history and the contributions of authors and editors, explains its pedagogical value for modern students of classical languages; and discusses technical questions of coding the transcript and preparing it for end-users.

Keywords: LATIN, SYNONYMS, LEXICON, MEDIEVAL LATIN, MONOLINGUAL DICTIONARY, *DE DIFFERENTIIS VERBORUM*

Opsomming: Popma se eentalige woordeboek van Latynse sinonieme. In hierdie verslag word die resultate van ons projek rakende die digitalisering van 'n woordeboek van Latynse sinonieme wat vir drie eeue van XVII tot XIX gewild was, weergegee; die geskiedenis daarvan en die bydraes van outeurs en uitgewers word beskryf; die pedagogiese waarde vir die moderne student van klassieke tale word uiteengesit; en tegniese vrae rondom die kodering van die transkripsie en die voorbereiding daarvan vir die eindgebruikers word bespreek.

Sleutelwoorde: LATYN, SINONIEME, LEKSIKON, MIDDELEEUSE LATYN, EENTALIGE WOORDEBOEK, *DE DIFFERENTIIS VERBORUM*

Introduction

The book titled *De Differentiis Verborum* was composed by a Dutch jurisconsult Ausonius Popma, first published in 1606 and reprinted 22 times in the following years. It became a basis for commentators of Latin texts and a source of citations. This dictionary was highly praised by scholars, so almost two centuries later Hill (1794) states that "remarks 'de differentiis verborum,' are often both ingenious and solid".

Popma's book is a monolingual Latin dictionary — also discussing few Ancient Greek words — explaining groups of words similar in meaning, pronunciation, or spelling. Among such pairs are homonyms (*acus* 'needle' and 'husks of grain'), homophones ("Arundo crescit, Hirundo cantat, Hirudo fugit"), homographs (*buccinus, buccina, buccinum*), etymologically related words (*aestas, aestivare, aestivum*), terms of familiarity (see *affinitas*), or words having similar meaning (*amator* and *amicus*). Hereafter we will call them "synonyms" in the wide sense of the word.

The dictionary represents also a class of words having separate meaning in modern English, but not distinguished by Romans. For example, Popma explains noun *altum* as "quod sursum est, et quod deorsum". Other examples: "Hospes est, qui recipit et qui recipitur"; "Vector est ille, qui portat, et qui portatur."

Taking into account the direct usefulness of such a book, as any other dictionary, we would like to note several additional situations when this, relatively old dictionary, can be beneficial to a modern student of Latin language.

The first good point is that the language of Popma's book is rather simple, compared to authentic Roman texts, so intermediate students should be able to understand it without difficulty. We could estimate this level of proficiency, though speculatively, as an active vocabulary of about 2000 words or the completion of the first part of the Ørberg's course "Lingua Latina per se Illustrata" (Ørberg 2010). Regular use of the monolingual dictionary suits the idea of natural language learning method, or it can be considered as kind of extensive reading.

All articles of the dictionary can be separated into two big topics, a group of legal terms and synonyms from common literature. The first group represents the professional interests of the author who was a jurisconsult, so that the articles about *ampliatio* 'trial postponement' and *comperendinatio* 'adjournment of a trial for two days' are two of the lengthiest in the book.

The second group is based significantly on the texts of Cicero, Virgil, and Pliny, following the medieval tradition of annotating and explaining the most popular documents. From this point of view, Popma's dictionary will be helpful to modern students, because the above-mentioned Roman authors are a basis of every Classic language course.

Spaced repetition and flashcards are methods often used to memorize new words. From our own experience, dictionaries of synonyms are obligatory companions to this method. For example, the first book of the Ørberg's Latin language course contains three words meaning 'to shed tears': *plorare* (ch. 3), *lacrimare* (ch. 7), and *flere* (ch. 24). Opening a flashcard with this English word (or similar 'to cry', 'to weep'), a student can answer any of its Latin counterparts, so we recommend adding all suitable synonyms to the answers of the flashcard. These additions can be made by students themselves during the course of education along with self-studying of differences between these words.

Evolution of the book

Our transcript is based on the most recent edition by Tommaso Vallauri in 1865 (and reprinted in 1870), but from the first publication in 1606 the text of the book underwent so many corrections and amplifications that it will be better to call it a collaborative work (see Figure 1).

The first edition, produced by Ausonius van Popma (1563–1613), included lists from several ancient authors as separate chapters: Marcus Cornelius Fronto (fl. 1st century AD), Flavius Caper (fl. 2nd century AD), Agroecius (or Agraetius; fl. 5th century AD), Flavius Sosipater Charisius (fl. 4th century AD), Servius Honoratus (fl. 4th century AD), Aelius Donatus (fl. 4th century AD), and Nonius Marcellius (fl. 4th or 5th century AD).

The third printing in 1618, curated by Bartholomaeus Meuschen (1578–1629), added a chapter from Sex Pompeius Festus (fl. 2nd century AD).

After a series of reprints in the following years, curated first by Bartholomaeus Meuschen and later by Thomas Hieronymus (1642–1716), the first significant revision of the text was made in 1694 by Johann Friedrich Heckel (1640–1715), who rewrote many articles, added new groups of synonymous words, incorporated separate lists into the main text and ordered all words alphabetically.

The next key editor was Adam Daniel Richter (1709–1782) who also significantly increased and improved content of the book in 1741. In 1750, he published an essay "Differentias quae in Ausonii Popmae De differentiis verborum libris amissae sunt" containing 62 articles. However, it was not noticed by later editors and was not included in subsequent editions of the Popma's dictionary. A digitized text of the Richter's *Differentias* is included in the supplementary files.

Johann Christoph Strodtmann (1717–1756), for his part, printed in 1743 "Spicilegium differentiarum in Ausonii Popmae De differentiis verborum libris amissarum". It was followed in 1753–1754 by two big articles "Centuria differentium apud latinos vocum ab Ausonio Popma vel omissarum vel curatius explicandarum" in *Acta Societatis Latinae Ienensis* (Strodtmann 1753; Strodtmann 1754). As it follows from the title, he added new articles, supplemented and elaborated Popma's text.

In 1769, Johann Christian Messerschmid (1720–1794) included most of the Strodtmann's commentaries into his edition of the Popma's dictionary (1769). Collation of the texts revealed only 17 articles missed by Messerschmid and

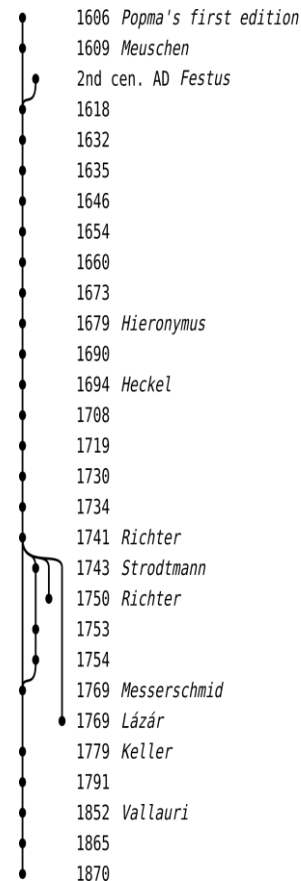


Figure 1: Editions of the book and the derived works

several more articles with an essentially different text. We plan to transcribe them in the future.

Another book published in 1769 was written by János Lázár (1703–1772). Image scans of this book were recently made available by librarians of Székely Nemzeti Múzeum (Sfântu Gheorghe, Romania) and still await a thorough analysis. The book generally follows Popma's dictionary, but János Lázár rewrote articles in a concise manner and added several new groups of words. The approximate number of articles in the book is about 600. It seems that Lázár's work was not known outside Romania and Hungary.

The next editions of Popma's dictionary, printed in 1779 and 1791 and curated by Johannes Christophorus Keller (1737–1796), fully correspond to the text of Messerschmid. Comparison of the word indices has found no differences.

Finally, the most recent version of the text was made by Tommaso Valauri (1805–1897), who continued to add new groups of synonyms and to polish text of the articles. A distinctive property of this edition is more and more active use of vernacular languages, such as French or Italian. For example: "Crustae sunt laminae inauratae, quae poculis, aut vasis inferuntur, ut vix refelli possint; italice: *riporti di basso rilievo*". His second edition (1865) contains even more commentaries in Italian. German words can be found in earlier editions however, especially in Strodtmann's version: "Hi enim flagella *Spitzruthen*, virgam *Ruthe*, scuticam *Peistche*, fustes *Prugel* dicunt."

Transcript format

Choosing a final transcript format, our primary candidate was the Text Encoding Initiative (TEI), and particularly its dictionaries encoding Guidelines (TEI Consortium 2007). TEI is a rich markup language based on XML and designed for coding a wide range of texts from lexicons to prose and poetry. It covers every case of semantic information we wanted to mark up in the transcript of Popma's book.

The Text Encoding Initiative was widely accepted in big Latin research projects, such as Corpus Corporum, Digital Library of Late-Antique Latin Texts (digilibLT), Glossarium mediae et infimae latinitatis, Perseus Digital Library and Corpus Automatum Manhemense Electorum Neolatinitatis Auctorum (CAMENA) (Schibel and Rydberg-Cox 2006; Glorieux and Thuillier 2010; Tabacco and Lana 2010; Roelli 2014; Crane 2021). We plan to submit our transcript to the Corpus Corporum so the choice of the TEI format is reasonable.

Another notable project publishing Latin texts is Lexicons of Early Modern English (LEME) (Lancashire 2018). Its editors adopted their own format based on XML but using a different set of tags. However, since 2019, this project also publishes documents in the TEI format.

Other formats, such as XML Dictionary Exchange Format (XDXF) or formats promoted by SIL International (MDF and LIFT), are more rigid and only suitable for bilingual dictionaries with a strict structure, so are not applicable to our dictionary of synonyms.

The TEI format provides tags to mark up fragments of the text semantically and visually. In our transcript we use the following semantic tags:

- `<entryFree>` is the parent node of every article; in terms of TEI standard it marks articles with loose structure; attribute "id" connects every entry with the corresponding record in the lexicon file (see next section);
- `<quote>` and `<bibl>` mark a quotation and its author (together with the title of the work); these tags are not joined into `<cit>` block;
- `<foreign xml:lang="grc">` is used for non-Latin fragments (Ancient Greek, French, etc.);
- `<abbr>` marks common abbreviations (v.g., h.e., v.c.).

Corrections made by the transcribers:

- `<corr>` tag contains both original and corrected reading of the fragment; in most cases typographic errors were checked against the 1852 edition;
- `<add>` is used when one letter was missing or a fragment was added by the transcriber.

Visual formatting tags:

- `<hi rend="bold">` highlights key (index) words;
- `<hi rend="italic">` is used for visual formatting;
- `<hi rend="term">` highlights key words inside of quotes;
- `<label>` marks labels of list items;
- `<lg>` is a parent node for lines `<l>`; it is used to format poetry.

End-user format

TEI has become a standard format of book transcripts. It is suitable for machine processing or producing a paper book layout. However, end-users are not able to read it directly and need a computer application to load the dictionary and look up words.

There are several desktop and mobile application, such as GoldenDict, AARD 2, StarDict, MDict, ABBYY Lingvo and many other. These dictionary shells support one or several file formats.

XML based formats, and TEI in particular, can be easily transformed into HTML (see Figure 2). Among the popular dictionary formats, Slob, StarDict and MDict can store articles with HTML codes, so we routinely encode Popma's dictionary in these formats. Files in DSL format were compiled by an anonymous volunteer.

amo, amare
diligo, diligere
colo, colere
adamo, adamare
deamo, deamare

Amo et **Diligo** hoc differunt, quod **amo** maiorem vim atque affectionem denotet, seu longe amplius efficacius sit, quam **diligo** (1). Sicut apud Graecos plus est **στέργειν**, quam **φιλεῖν**, et quam **ἀγαπάειν** (2). Ter. Eun. I. 2. 16: *Non quo quemquam plus amem aut plus diligam*. Cic. Fam. IX. 14: *Tantum tamen accessit, ut nunc demum amare videar, antea dilexisse*. Id. Fragm. ap. Non. IV. 78: *Me aut amabis, aut, quo contentus sum, diliges*. Praeterea **diligimus** ex virtute; **amamus** vero interdum utilitatis aut voluptatis causa. Id. Amic. IX: *Ego admiratione quadam virtutis eius, ille vicissim opinione fortasse nonnulla, quam de meis moribus habebat, me dilexit*. Valla V. 37. **Amare** est efficacius quam **Diligere**; **Colere** modo est venerari seu cultum tribuere maioribus, modo est prosequi: sic **colimus** pares amore et officio, minores humanitate et beneficiis; id est, prosequimur amore, officio. R.

(1) Videtur Scaliger dilige dici a diligendo, quod tamquam initium sit amandi.

(2) **Στέργειν** autem utraque hac voce est **ἐμφρατικώτερον**. Notat enim naturali quadam caritate complecti, qualis est amor atque affectus parentum erga liberos, liberorum vicissim erga parentes et sanguine iunctos, Graecis **στοργή** dictus. Sed quomodo dii (addamus et deae) amare nos dicantur, Leg. Gasp. Sciopp. Franci, Suspectar. Lection. lib. II. epist. XVII: *Venus nos amare dicitur, quando venustatem nobis et potentiam in amore largitur*. Sic **amamus** nos invicem, **amamus** nos mutuo, **amamus** nos pariter, **amamus** inter nos, idem significat; sed in postremo non geminabitur accusativus, ut: **amamus** nos inter nos; sed **amamus** inter nos. **Adamare** est amatorie amare, sive amare incipere et multo minus quam **deamare**, quod idem est, ac valde amare. V. g. Si quem sincero amore adamasti, constanter deama, amantemque te redama sicut solet amasius amasiam. Sic a latinis amplecti et amplexari pro valde amare ponuntur.

See article on [Archive.org](#), [Google Books](#)

Figure 2: An example of an article with semantic markup rendered in HTML format

In this way, our dictionary is available for use with at least 5 desktop applications, and 23 mobile dictionary shells. It is included into our collection of Latin dictionaries (<https://latin-dict.github.io>) and was downloaded more than 130 times (400 times counting pre-release versions).

Another situation in which computer devices can assist users is so-called "morphological search". When a user types in a word in declined or conjugated form, such as *amavisti* '(you) loved', a dictionary shell should find the correct article — *amo* 'to love'. Hunspell spell checking library is one of the lemmatizers the most widely used in dictionary shells. Keywords in Popma's dictionary are also written in a non-standard form. For example, in "*Acini* densius nascuntur; *Baccae* autem rarius", both headwords are in plural forms. To harmonize them with Hunspell, a list of corresponding keywords and their normal forms was compiled (file *lexicon.json*) and used in producing of files in end-user formats. In such a way, users can look up not only derived forms of the words, but also their orthographical or medieval variants: *epistola* or *epistula*, *coelum* or *caelum*.

The text of the transcript, and all derived and supplementary materials are distributed as Public Domain on the web-site <https://latin-dict.github.io/dictionaries/Popma1865.html>.

Conclusion

Popma's *De Differentiis Verborum* was a prominent dictionary for three centu-

ries before it was substituted with dictionaries written in vernacular languages. However, it can gain demand in new conditions following the modern methods of teaching Latin languages, such as the Direct Approach (Natural Method) and the promotion of extensive reading.

Recently digitized into a machine-readable form, the text will increase the corpus of Latin literature and find application in future linguistic research. Students of Classic courses can use it on their mobile devices along with tens of other Latin dictionaries provided on our website.

Acknowledgement

The author would like to acknowledge contribution of Daniel Elhacham who participated in the transcription of the book and prepared the layout of the book for printing, and several other anonymous internet users who participated in typing and proofing, or making the dictionary available for different computer applications.

Bibliography

- Crane, G.R.** 2021. Perseus Digital Library. URL: <http://www.perseus.tufts.edu>.
- Glorieux, F. and S. Thuillier.** 2010. Grec ancien, latin médiéval, balisage comparé de deux dictionnaires, vers des ressources linguistiques. *Archivum Latinitatis Medii Aevi* 68: 161-181. URL: <http://hdl.handle.net/2042/56776>.
- Hill, J.** 1794. An Essay upon the Utility of Defining Synonymous Terms in All Languages. *Transactions of the Royal Society of Edinburgh* 3:93-130. URL: <https://books.google.com/books?id=CqRbAAAACAAJ>.
- Lancashire, I.** 2018. Lexicons of Early Modern English. URL: <http://leme.library.utoronto.ca/>.
- Roelli, P.** 2014. The Corpus Corporum, a New Open Latin Text Repository and Tool. *Archivum Latinitatis Medii Aevi* 72: 289-304. URL: <http://hdl.handle.net/2042/69525>.
- Schibel, W. and J.A. Rydberg-Cox.** 2006. Early Modern Culture in a Comprehensive Digital Library. *D-Lib Magazine* 12(3): 1082-9873. URL: <http://www.dlib.org/dlib/march06/schibel/03schibel.html>.
- Strodtmann, J.C.** 1753. Centuria differentium apud latinos vocum ab Ausonio Popma vel omissarum vel curatius explicandarum. *Acta Societatis Latinae Jenensis* 2. URL: https://zs.thulb.uni-jena.de/receive/jportal_jparticle_00204932.
- Strodtmann, J.C.** 1754. Centuria secunda differentium apud latinos vocum ab Ausonio Popma vel omissarum vel curatius explicandarum. *Acta Societatis Latinae Jenensis* 3. URL: https://zs.thulb.uni-jena.de/receive/jportal_jparticle_00205207.
- Tabacco, R. and M. Lana.** 2010. Digital Library of Late-Antique Latin Texts. URL: <https://digiliblt.uniupo.it>.
- TEI Consortium.** 2007. 9 Dictionaries. URL: <https://tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>.
- Ørberg, H.H.** 2010. *Lingua Latina per se Illustrata, Pars I: Familia Romana*. Newburyport, MA: Focus.

Howard Jackson (Ed.). *The Bloomsbury Handbook of Lexicography*. Second edition. 2022, xxiv + 473 pp. ISBN: 978-1-3501-8170-0 (Hardback); 978-1-3501-8171-7 (ePDF); 978-1-3501-8172-4 (eBook). London: Bloomsbury. Price £130.00 (Hardback), £117.00 (eBook/ePDF).

The Bloomsbury Handbook of Lexicography (2022) is the updated edition of *The Bloomsbury Companion to Lexicography*, which was published in 2013.

This has been updated and expanded, not completely revised.

The second edition of *The Bloomsbury Handbook of Lexicography* is an edited collection of chapters covering a comprehensive range of topics within the field of lexicography. The list of contributors includes well-known practical and theoretical lexicographers, linguists, translators, and teachers. Editor Howard Jackson describes the aim of the book as providing a broad overview of the discipline, "dealing with the main trends and issues in the contemporary study of lexicography". According to the introduction, all of the original chapters, except for Chapters 2 and 7 have been updated by their original authors.

Part 1, which covers research methods and problems, opens with an excellent overview of researching lexicographic practice, by Lars Trap-Jensen. This chapter explores the different stages of dictionary-making, from conceptualisation to data access and presentation. While very general, this chapter sets the stage for the other, more focused chapters in the book.

Part 1 also includes a chapter on dictionary criticism, by Kaoru Akasu and one on researching dictionary use and dictionary users by Hilary Nesi. The first part of the *Handbook* concludes with a new chapter by Howard Jackson, Methods in (meta)lexicography. This chapter presents "an overview of the variety of methods that are used in the pursuit of lexicography". While the Trap-Jensen chapter presents an overview of the methods of dictionary-making, the Jackson chapter concludes this section with an excellent overview of the research of lexicography, while also providing a summary of the other chapters in this section.

Part 2 of the *Handbook* covers current research and issues. This section contains more focused chapters on aspects of lexicography, such as Researching the use of electronic dictionaries (Verónica Pastor and Amparo Alcina), which leads on from the more general Nesi chapter in Part 1. The Pastor and Alcina chapter in Part 2 presents a very detailed discussion on the different searches offered in electronic dictionaries, and how they are used.

Part 2 also includes the chapter by Adam Kilgarriff, Using corpora as data sources for dictionaries. This chapter discusses the use of corpora in dictionary-making, and their applications in different parts of dictionary-making.

Also in Part 2 is D.J. Prinsloo's Aspects of African language lexicography, which presents the challenges posed to lexicographers who compile dictionaries in African languages and want them to be as user-friendly and useful as possible. (The chapter was Issues in compiling dictionaries for African languages in the previous edition of this book.) Most of the challenges discussed by Prinsloo

are related to the complex grammar of many African languages, and lexicographers need to find ways to present these in an accessible and logical manner.

Other chapters in Part 2 of the *Handbook* cover historical lexicography (John Considine), pedagogical lexicography (Amy Chi), and bilingual lexicography (Arleta Adamska-Salaciak). A new chapter in this section is Compiling dictionaries for minority and endangered languages, by Verna Stutzman and Kevin Warfel. This chapter offers reasons for the necessity of minority language dictionaries, and discusses how the compilation of such dictionaries is very different from the compilation of dictionaries from languages in wider use. The chapter describes the collection of data (as well as the development of a writing system, if necessary) and the development of entries, specifically focusing on bilingual and multilingual dictionaries that include the minority language in question.

Another new chapter in Part 2 is Aspects of Multi-word expressions in Asian lexicography, by Vincent B.Y. Ooi, Ai Inoue, Kilim Nam, and Cuilian Zhou. This chapter documents the issue with multi-word expressions in Chinese, Japanese, Korean, and Malay lexicography. This chapter presents a fascinating insight into the treatment of multi-word expressions that are present in these languages.

The third new chapter in this section is Issues in onomasiological lexicography, by Gerardo Sierra. The chapter begins with an introduction to onomasiological lexicography and where it fits in lexicographic theory. It presents different types of onomasiological dictionaries and discusses how they fit into onomasiological lexicography. Finally, the chapter presents drawbacks of onomasiological dictionaries, and concludes with reassurance that complete and efficient onomasiological dictionaries that are very useful to their users are possible.

The final new chapter in Part 2 is Issues in collaborative and crowdsourced lexicography, by Franck Sajous and Amélie Josselin-Leray. This chapter lists the areas in which progress has been made in lexicography in the past few decades, and then details the rise of 'DIY dictionaries', which are dictionaries such as *Wiktionary* and *Urban Dictionary* that are compiled entirely by collaboration or crowdsourcing. The other form of collaboration or crowdsourcing is where lexicographers working on a dictionary delegate small parts of the lexicographic process, such as annotation, to volunteers. This chapter asks whether an analysis of these dictionaries gives an accurate picture of what amateurs can produce.

The authors begin to answer that question with the distinction between crowdsourcing and collaboration, and show how the two different techniques are used in dictionary making. The chapter also touches on the ethics of such techniques as well as quality control. The chapter concludes with a question on whether crowdsourcing is the new innovation in lexicography or whether it will simply lead to further innovation.

This final chapter of Part 2 segues neatly into Part 3, which covers New directions in lexicography. This section contains four chapters as well as a

chapter on Resources, by Reinhard Hartmann. The first chapter in Part 3 is an updated chapter, by Pedro A. Fuertes-Olivera, Theoretical, technological and financial challenges: Some reflections for making online dictionaries. This chapter is updated with discussions on the use of the terms *e-lexicography* and *online dictionary* to reflect more current thinking and the state of lexicography today. The chapter also discusses the relationships between the technology, the users, the makers of dictionaries, and the economic costs of making and using dictionaries. This chapter also offers guidance to how dictionaries can be produced with a profitable business model for publishers.

The new chapter in Part 3 is The design of internet dictionaries, by Annette Klosa-Kückelhaus and Frank Michaelis. This chapter presents an overview of the role of design in the form of dictionaries and their usability. This chapter presents different traditions in the designs of both print and electronic dictionaries as well as usage studies which consider user needs in dictionary design. The chapter considers design fundamentals and presents online dictionaries that adhere to these fundamentals. The authors show how the design of electronic dictionaries has progressed beyond electronic versions of print dictionaries, to what we see today, and offers a challenge to see what comes next in electronic dictionary design.

Chapter 25 is included as a chapter of Part 3, but is a list of resources, set out according to the types of resources and who needs them. The sections are: Academics, Associations, Corpora/Databases, Journals, Networks, Online dictionaries, Publishers, and University Research Centres. This section, while very useful, has the potential to fall out of date very quickly. It can already be seen that the associations listed have not been updated since 2019, as instead of just providing the foundation date of each association, the end date of many is given as 2019. However, despite this minor drawback, the resources section provides a very comprehensive list of lexicographic resources to suit any reader of this book.

The end pages include a clear and accessible glossary of lexicographic terms, with concise descriptions of the terms included. This is followed by an incredibly useful annotated bibliography, which has entries arranged by topic. In all but the first topic, the entries are listed in reverse chronological order. It is useful to see that the most recent works are dated 2020.

The bibliography is followed by a names index, with a note indicating that names are only included if "their work has been extensively or significantly cited". This is followed by a general index.

Strengths: all the chapters in this *Handbook* are well written and accessible. They have been written by experts in their fields, and are current and coherently written. The research is relevant and clearly contextualised. The book is systematically organised in a logical order, and the chapters can be read in order, or independently of other chapters.

Weakness: the only shortcoming I could find is that many of the links in the book are dead links that no longer work. This would be mostly apparent in

the e-book where the links are hyperlinked. For example, in Chapter 6, Methods in (Meta)lexicography, which is one of the new chapters, the first link on page 58 leads one to a "page not found" error message.

The Bloomsbury Handbook of Lexicography is an excellent resource that I would strongly recommend for students of lexicography and practitioners alike.

Is it a necessary update? As to whether *The Bloomsbury Companion to Lexicography* (published 2013) needed an update, yes. The developments in lexicography, especially electronic lexicography, have happened very rapidly in the last decade, and as the new and updated chapters of *The Bloomsbury Handbook of Lexicography* can confirm, this new edition is a valid and valuable contribution to the library of lexicographic research.

Lorna Hiles Morris
Cape Town
South Africa
(lorna@lemma.co.za)

Who is Sharpest at Looking Up *sharp*? Comparing Two Parallel Groups of Dictionary Users

Marjeta Vrbinc, *Faculty of Arts, University of Ljubljana, Slovenia*
(marjeta.vrbinc@ff.uni-lj.si),

Donna M.T.Cr. Farina, *Department of Multicultural Education,
New Jersey City University, New Jersey, USA*
(dfarina@njcu.edu)

and

Alenka Vrbinc, *School of Economics and Business,
University of Ljubljana, Slovenia*
(alenka.vrbinc@ef.uni-lj.si)

Abstract: This article is a comparison of two previous research studies (Farina et al. 2019; Vrbinc et al. in press), both of which examined the dictionary look-up behaviors of two very different cohorts of undergraduate students from the University of Ljubljana, Slovenia. At the time the two original and parallel look-up studies were carried out, one cohort was majoring in business and economics in the School of Economics and Business and the other was majoring in English in the Faculty of Arts. The previously published work reports on how, in both groups, participants were given nine contexts containing a clearly marked common English word used in an infrequent and often unknown sense; they had to locate the relevant sense related to a given context in an unknown-to-them learner's dictionary, *The Britannica Dictionary*. The participants were asked to think aloud as they looked up words; the researchers observed and recorded their approaches and problems. Prior to, during, and after the look-up process, the members of these two cohorts responded to fourteen questions about their habits of dictionary use and their perceptions of the utility and quality of definitions and illustrative examples that they encountered. This article contrasting the two studies indicates that the look-up proficiency of the two groups differed significantly.

Keywords: DICTIONARY USER, ADVANCED ENGLISH LEARNER, LEARNER'S DICTIONARY, DICTIONARY AWARENESS, LOOK-UP BEHAVIOR, QUALITY OF DEFINITION, QUALITY OF EXAMPLES

Opsomming: Wie is die skerpsinnigste in die naslaan van *skerp*? Die vergelyking van twee parallelle groepe woordeboekgebruikers. Hierdie artikel is 'n vergelyking van twee vorige navorsingstudies (Farina et al. 2019; Vrbinc et al. ter perse) waarin die woordeboeknaslaangedrag van twee heeltemal verskillende groepe voorgraadse studente van die

Universiteit van Ljubljana, Slowenië ondersoek is. Ten tyde van die uitvoer van die twee oorspronklike en parallelle woordeboeknaslaanstudies het een groep studente in besigheidstudies en ekonomie aan die Ekonomie- en Sakeskool gespesialiseer en die ander groep in Engels aan die Fakulteit Lettere. Die voorheen gepubliseerde studies doen verslag oor hoe, in beide groepe, aan deelnemers nege kontekste gegee is wat 'n duidelik gemerkte algemene Engelse woord wat in 'n ongewone en dikwels onbekende betekenis gebruik is, bevat; hulle moes die relevante betekenis binne 'n gegewe konteks in 'n aanleerderswoordeboek wat onbekend aan hulle was, *The Britannica Dictionary*, bepaal. Die deelnemers is gevra om hardop te dink terwyl hulle woorde naslaan; die navorsers het hul benaderings en probleme waargeneem en aangeteken. Voor, tydens en ná die naslaanproses het die lede van die twee groepe veertien vrae oor hul woordeboekgebruiksgewoontes en waarnemings van die bruikbaarheid en kwaliteit van definisies en illustratiewe voorbeelde wat hulle teëgekomp het, beantwoord. Hierdie artikel waarin die twee studies gekontrasteer word, dui daarop dat die naslaanvaardighede van die twee groepe beduidend van mekaar verskil.

Sleutelwoorde: WOORDEBOEKGEBRUIKER, GEVORDERDE AANLEERDER VAN ENGELS, AANLEERDERSWOORDEBOEK, WOORDEBOEKBEWUSTHEID, NASLAANGEDRAG, KWALITEIT VAN DEFINISIE, KWALITEIT VAN VOORBEELE

1. Introduction

Lexicographers usually cater to an audience with education, but one that does not think extensively about the finer points of a language. Lexicographic works are normally consulted by such an audience for utilitarian purposes and no more (see Béjoint 2010). Most educated users want what they want from a dictionary; they want to get what they want quickly and move forward. Possibly the greatest challenge for a lexicographer is to produce an online reference work that is easy to use for the general educated public; using this reference work should not demand an extensive understanding of language or linguistics. But then we must ask: Just how different is the general educated user from a user who is closer in training to an actual lexicographer? How much better would a language specialist perform in dictionary look-up tasks as compared to other educated users who do not specialize in language or linguistics?

This article is an effort to discover answers to such questions. It compares the results of two parallel studies (Farina et al. 2019; Vrbinc et al. in press), both of which used an identical methodology to examine the look-up process of online learner's dictionary users. These qualitative studies, carried out in March 2018 at the University of Ljubljana, were the first to directly investigate the behaviors of advanced learners of English/dictionary users in Slovenia, and were a joint project between the United States and Slovenia (see Acknowledgements). The studies had two different sets of undergraduate participants. The first cohort, nine third-year students from the School of Economics and Business (FE¹), were majoring in different areas of business and economics and were not specialists in English. The second cohort, eight third-year undergraduate majors from the

Department of English Philology in the Faculty of Arts (FA), were preparing as specialists in the English language. While the business/economics majors were competent English users, they were not inordinately interested in the English language per se; on the other hand, the English majors had both a deep interest in English and extensive experience in the use of dictionaries. The look-up process of both studies refers to the numerous simultaneous actions taken by the users as they scanned a dictionary (in this case, online) for specific information. In the present work, the FE and the FA students are being compared, to discern how their practices in using an online learner's dictionary differ.

The two cohorts of the parallel studies were asked to read nine contexts, each containing a clearly marked, well-known English word used in an infrequent sense; they had to search for the correct, relevant meaning among all senses of the word in a learner's dictionary, *The Britannica Dictionary* (BD). When the two studies were conducted, this dictionary was titled the *Merriam-Webster Learner's Dictionary*. At the time of writing, it has been rebranded by Encyclopaedia Britannica, Inc. as the BD.² Typologically this dictionary is a work appearing in both print (2008 and 2016) and online forms, the second category in Rundell (2015: 305), with the digital dictionary derived from the print versions. The BD was completely unknown to the participants.

As was outlined in Farina et al. (2019) and Vrbinc et al. (in press), the two parallel studies being compared here do not share characteristics with many previous studies gaging the habits of dictionary users. We did not investigate *how* users access dictionaries as did Lew (2011) and Lorentzen and Theilgaard (2012); we presented students immediately with the online learner's dictionary, the BD, that we expected them to use. We did not compare students' use of several different dictionaries, as in De Schryver and Prinsloo (2011) or Herbst (1996). While neither study sought to investigate design or presentation of information in an online learner's dictionary (cf. Gouws 2014; Lew and De Schryver 2014), the results contrasted here indicate that for only one of the two cohorts did format turn out to be significant. Two older studies have elements in common with the parallel studies. Mitchell (1983) discussed how children use contexts and locate related information in monolingual dictionaries; while the age of the informants is different (younger) and the (print) dictionaries used were for native speakers, there were some commonalities with the findings of the parallel studies. Tono (1984) looked at how long users were willing to read down through an entry before they would essentially give up and accept the closest information found. The investigation presented here contrasts differences in how the two parallel cohorts responded when information was located far down in an entry and required prolonged online scrolling.³

2. Methodology

This section describes two separate methodologies. First, the uniform methodology of the two original, parallel studies is described, albeit more briefly than

in Farina et al. (2019) and Vrbinc et al. (in press). Second, the methodology followed for the present comparison of the two parallel studies is provided.

2.1 Set-up of the two parallel studies

Ahead of dictionary consultation in both parallel studies, the researchers selected nine contexts for the participants to read, to be used as the drivers of their dictionary look-up of targeted words. Seven contexts were taken from an American newspaper and two from texts of American non-fiction and American fiction (see Appendix A: List of Contexts and Their References). Within each context appears an underlined, boldfaced word of standard English, not used in its most common sense but in an infrequent sense. During both studies, the participants were first asked general questions about their dictionary-use habits, as well as a few additional questions. Next, each participant began to read the contexts, one at a time. After reading a context, they began to look up the target word's meaning in the online BD. We observed the participants while they were looking up the meaning in both studies; the participants also discussed aloud what they were doing during look-up and answered both scripted and other questions. Our roles in both parallel studies were to ask questions, closely observe participant behaviors throughout, and take notes.

2.1.1 Dictionary used in the two studies

As noted above, the dictionary used was *The Britannica Dictionary* (BD), rebranded from the online *Merriam-Webster Learner's Dictionary*. While Encyclopaedia Britannica, Inc. has removed the mention of "learner's" from the current title, during the parallel studies participants were quite aware that the dictionary is intended for English learners. This is not a dictionary that is familiar to university students in Slovenia; no participant in the parallel studies had ever used it.

2.1.2 Study participants, their English proficiency, and their expertise in English

For the parallel studies, eighteen volunteers were recruited from among students in the School of Economics and Business (FE) and the Faculty of Arts (FA) at the University of Ljubljana; one student from the FA dropped out, leaving a total of 17 participants, with nine persons volunteering from the FE and eight from the FA. All persons in both groups were advanced English users. We estimate that at the time of the study, the FE students were at the C1 level and some may have been at the C2 level in the Common European Framework of Reference for Languages (CEFR), or at the Advanced High or Superior level in the proficiency scale of the American Council on the Teaching of Foreign Lan-

guages (ACTFL) (*Učni načrt Angleščina 2008; Assigning CEFR Ratings to ACTFL Assessments* n.d.). We estimate that at the time of the study, the FA students were very near to or had already achieved the C2 level in the CEFR or the Superior level in ACTFL (*Učni načrt Angleščina 2008; Assigning CEFR Ratings to ACTFL Assessments* n.d.).⁴

While the respective valid and reliable proficiency scales of the U.S. and Europe paint part of the picture about knowledge of a language, they do not fully capture the linguistic knowledge of grammar or experience with nuances of language. When such factors are taken into consideration, it is clear that the FA study participants have skills and experience with English that, even if they cannot yet be called "professional-level," they are moving in that direction. As almost-English-language professionals, the FA participants had a skill set that distinguished them from the general users of English in the FE. As will become clear here, this almost-professional skill set impacted positively the FA performance in dictionary look-up. As is discussed below, while all participants from both studies were proficient speakers and readers of English, the FAs (and not the FEs) stood out as atypical dictionary users with an avid interest in lexicographic products.

2.1.3 Target words and target contexts for the parallel studies

In the two parallel studies, the targeted words, in context, in an infrequent sense, were selected to push the students to demonstrate the full gamut of their dictionary look-up skills. These highly-proficient English learners certainly knew all the target words, but this knowledge could be deceptive because the senses chosen were different from the most common meaning — and these infrequent senses were often unknown or at least unfamiliar to the students. To determine whether the targeted senses of these words were infrequent, we examined the sense ordering in four learner's dictionaries.⁵

The following nine words were used in both studies: *tax* (verb), *fix* (verb), *score* (verb), *pitch* (noun), *plug* (noun), *ticket* (noun), *sharp* (adjective), *mean* (adjective), *rich* (adjective). Note that only content-word classes and not function-word classes were chosen. The number of words chosen was low, to allow the participants enough time to read the contexts provided, look up the words online, and answer pre- and post-look-up questions, all within sixty minutes.

Criteria for contexts were established in advance of the parallel studies. They had to reflect language in the US between 2000 and 2018; they should not be academic but contemporary, standard texts that most educated advanced learners would comprehend. Our understanding of standard language corresponds to Finegan (2011: 13): a language variety for activities such as radio broadcasts, university lectures, or political speech. We tried to select senses whose examples in the dictionary were not identical to the language of the contexts; the only exception was the adjective *rich* as in "Oh, now that's rich" (Context 9, Appendix A). The contexts used in both studies contain between 37 and 85 words: the

mean is 62 and the median is 60. In the time allotted, the students were able to read through each context, answer questions about it and then proceed to looking up the target sense in the dictionary; as will be seen here, the FA group demonstrated more alacrity in these tasks than the FE students.

As an example of a typical context used in the parallel studies, *ticket* at 66 words is just above the mean and the median:

Context 6, *ticket* (noun)

The state's Republican chairman, Edward F. Cox, offered a respectful, if cautious statement about Mr. Kolb's candidacy. 'We're glad he has formally announced his intentions,' Mr. Cox said, before saying he was excited by the possibility of other candidates, too. [...]

He added that the party would be 'working collaboratively with our county chairs over the coming weeks and months to put together the state-wide **ticket**.'

In each context, the target word was in boldface and underlined, to avoid ambiguity about what to look for in the dictionary. Contexts are generous, far more than a single sentence; they afforded students the chance to understand the infrequent sense, even if they had never encountered it before. At the same time, the contexts were not overlong.

2.1.4 Procedures of the two parallel studies

These qualitative studies with anonymous volunteers intended to obtain detailed information about dictionary use, using semi-structured interviews and observation of dictionary look-up tasks (Hatherall 1984; Merriam 2009; Qu and Dumay 2011; Rubin and Rubin 1995). Our interview scripts for both studies corresponded to pre-selected topics, namely:

- 1: Habits of Dictionary Use
- 2: Look-up Ability of Participants
- 3: Perceptions of Utility and Quality of Definitions
- 4: Perceptions of Utility and Quality of Illustrative Examples

Despite having the same pre-selected topics for both studies, the interview process was still intended to discover unexpected information and not be limited to the investigation of "ready-made ... categories" (Qu and Dumay 2011: 243). The scripted interview questions were supplemented when necessary; we tried to be sensitive to events as they happened during the interviews and ask appropriate follow-up questions (Hannabuss 1996). In addition, the interviewees were encouraged to speak about what they were doing (think aloud) as they looked up words (Qu and Dumay 2011; Wingate 2002); the participants were quite willing to engage in this process. We also relied on direct observation,

made possible by the face-to-face environment used (Newton 2010). Overall, this set-up provided many possibilities to gain insight into the chosen topics and beyond (Gillham 2000; Merriam 2009). The results obtained from both studies validate the qualitative format choice. It is unlikely that (for example) a written questionnaire would have yielded the type of detailed information that these two in-person studies allowed.

The interview script used in the parallel studies contained 14 questions. The first six were general questions (related to Topic 1: Habits of Dictionary Use), asked before the student was given the nine contexts to read: Three covered dictionary usage habits, two asked about the students' satisfaction with what they find in dictionaries most of the time and how quickly they find information, and one question asked what they dislike or miss in the dictionaries they use. Next in the study, we observed as the students read a context. Then the participants were asked whether they know the meaning of the word in the context and whether they can tell us what it is. Then, under our direct observation, the informants searched online in the BD for the sense that corresponded to their context (Topic 2: Look-up Ability of Participants).

The interviewees were asked whether their initial definition was correct, whether their first ideas about the meaning were comparable to what was in the dictionary (Topic 3: Perceptions of Utility and Quality of Definitions). Subsequently, students were asked about the usefulness of the illustrative examples, what they liked about them, and how they could be improved (Topic 4: Perceptions of Utility and Quality of Illustrative Examples). Another question concerned information found in square brackets "[]" within the examples, and whether a given example was comprehensible without this information (Topic 4). The final question asked what part of the entry was the most helpful in understanding the meaning of the word as used in the context (Topics 3 and 4). After this, the participants were asked unscripted general questions: about the dictionary of the studies, about their perceptions of the study participation experience and other (all four Topics).

2.2 Methodology for comparing the two parallel studies

After both user studies were completed and the results had been analyzed, we began the present contrastive study — without specific themes in mind. This approach differs from that used in the two original studies, where themes were pre-identified and then later supplemented with other discoveries after the parallel studies had been carried out. For this contrastive study, we began by examining the interview notes connected with the parallel FE and the FA studies. This examination allowed repeated patterns and recurring themes to emerge within the qualitative data gathered (Caulfield 2022). From a fuller list of patterns and themes, similar ones were combined to create a more concise list (Caulfield 2022), what Ryan and Bernard (2003: 85) call "winnowing themes". As Maguire and Delahunt (2017: 3353) indicate: "This means that, unlike many qualitative meth-

odologies, it is not tied to a particular epistemological or theoretical perspective".

The process of data examination proceeded as follows: We took the data previously collected from each word looked up by the participants and juxtaposed the data of the FA participants with those of the FEs. The recurring themes that emerged reflect the researcher's prior experiences with lexicographic analysis ("understanding of the phenomenon under study," Ryan and Bernard (2003: 88)) as well as their values concerning what information could prove useful in the production of future online dictionary work (Ryan and Bernard 2003). From examination of both the FA and the FE data, the following six themes or areas of discussion emerged as the most productive and significant, and will be addressed below:

- Dictionary awareness
- Frequency of dictionary use
- Satisfaction with dictionaries used
- Navigation of contexts
- Navigation of the dictionary entry
- Evaluation of dictionary components

3. Comparison

3.1 Comparing dictionary awareness

Starting with the first questions asked in the two parallel studies, the responses of the FE and FA cohorts differed. At the outset, the deceptively simple question, "Do you use dictionaries, and if yes, which ones?" indicated that our groups were not on the same page. While all except one student in both studies stated that they use dictionaries, it was striking that the FE group only once referred to a specific, named dictionary (and this was Urban Dictionary). The FEs appeared to be unaware or minimally aware of which sources they consulted. In addition, the FE group mentioned Google more often, and not everyone in this group seemed to understand that Google Translate is not a dictionary. The response of the FA group could not have been more different. Here, each member stated that they use not one or two but several dictionaries; what is more, all FAs could name the specific dictionaries they use. Only in a single instance did an FA member state that they did not know the name of a source they used and referred to "the orange thesaurus."

3.2 Comparing frequency of dictionary use

Using dictionaries successfully is a skill like many others; engaging in the practice more often leads to better performance. It is not a surprise that the FA students, who performed better than the FEs overall in identifying the correct infrequent

senses of common words, also use dictionaries much more often. In the FA group, consulting dictionaries was, at a minimum, an activity that took place several times per week. On the other hand, the minimum usage in the FE group was a few times per month. Some of the FEs reported using dictionaries weekly as the maximum, compared to the FAs who at a maximum used dictionaries daily or several times per week. The comment of one FA: "If I don't find the exact thing [being sought] then I explore" is indicative of the full FA cohort's inclination to turn to dictionaries often and not to limit their engagement in dictionary activities.

3.3 "I can't get no satisfaction"? (Comparing satisfaction with dictionaries used)

Frequency of dictionary use is linked to satisfaction with use; a user who explores dictionaries often is one who likes dictionaries. Our questions in the two parallel studies sought to discern to what degree participants enjoyed using dictionaries and what (if anything) they missed in or disliked about dictionaries. Certainly, these themes are related to success in dictionary look-up; if users are able to find what they are looking for easily and quickly, we would guess that they are more likely to express satisfaction.

At first glance, the FE students appeared to express satisfaction with the dictionaries they used, with eight of nine saying they are satisfied and find what they are looking for quickly. (The ninth does not use dictionaries, but Google Translate.) However, a different picture emerged when the informants were probed further, to discover what they missed or disliked in dictionaries: This led to the expression on the part of the FEs of diverse dissatisfactions. While there was some very limited (and thus inconclusive) evidence that a few FEs who use dictionaries more often were more satisfied with them overall, still, depending on the individual, we heard that: Dictionaries were too complex, were poorly organized, contained too much information or too many abbreviations; definitions were unclear, too difficult, or too simple; and there were too few illustrative examples.

The FA students discussed their satisfaction in a different manner. They too claimed to be satisfied most of the time; individual comments underscored a high level of satisfaction or pointed out the usefulness in particular of print dictionaries. Concerning the quickness of information retrieval, on the surface the answers of the FAs were similar to those of the FEs. But what was striking was how the FAs appeared to define "quick." One FA mentioned that they spend no more than five to 10 minutes on searches; another mentioned that it might be necessary to scroll through a long entry if a word is complex. One FA discussed a half-hour search for a word but did not give us the impression that this half hour was problematic. Only one of the FE students mentioned a specific amount of time in response to our question about how "quickly" they retrieve information; this person stated that they take a maximum of one minute to find what they are looking for. This is not enough evidence to state with

certainty whether the FEs as a group would have considered 10 or 30 minutes to be quick or not. However, judging holistically from FE respondents' answers throughout the full interview, we consider it highly unlikely that any FE student would tolerate or engage in dictionary searches of even a full 10 minutes in length. "Quickness" is in the eye of the beholder.

The illustrative examples received a lot of attention during the questioning about satisfaction — but only from the FA students; the FEs made few specific comments apart from one passing remark about there being too few examples, and a second remark about examples not always hitting the spot. The FA students commented that they want more collocations in the dictionary, they want more examples for some words and they want numerous examples for every meaning of a word; they do not find all examples to be clear and they dislike it when examples do not fit with the specific context they are examining. Also telling in the FA discussion of examples was how these informants again named specific dictionaries; two dictionaries by name were said not to have enough illustrative examples. These comments were received before any FA looked at the target dictionary of the study. We consider that these opinions are indicative of the FAs' greater awareness that a dictionary text is not monolithic but consists of numerous components, one of which is the examples. During and following the look-up process, the FEs did home in on the illustrative examples and had much to say about them. However, as they began to think about dictionaries (the FEs perhaps for the first time; one FE student stated directly that they had never thought about our questions) in the initial, pre-look-up stage of the parallel studies, the FAs but not the FEs already demonstrated a more nuanced sense of what a dictionary is.

There is only one area in which the FEs had more specific comments to make about satisfaction, and those pertain to vocabulary rather than to a section of the dictionary. In terms of language for general purposes, the FEs stated that they were satisfied overall; however, they were not satisfied with how dictionaries treated the terminology of business and economics, or language for special purposes. They complained that they could not find many necessary terms in general dictionaries and resorted mostly to using Google or the internet to find them.

3.4 Comparing navigation of contexts

As was noted above, the informants in the parallel studies were asked to read nine relatively generous contexts, with the target word underlined and bold-faced in it. They were given as much time as they wished to read a given context and they usually signaled us (either with body language or words) that they had finished reading and were ready to begin looking up the target word.

It is safe to state that none of the participants had ever experienced a task like the one they were given in these studies. There are multiple ways of assessing how well they handled this task. First, did they understand the con-

text taken as a whole; was their proficiency level in English up to the job of comprehending the context? Second, even if they did grasp the full context eventually, how arduous was it for them to get to the point of understanding? In terms of the meaning of the target word, did they know what it meant in the context or did they just think that they knew it? If they did not know it initially, did the full context help them in figuring out the infrequent meaning of the target word?

For neither of the studies were the participants timed as they read through the contexts provided. However, impressionistically, we did not discern a difference in the read time of the two groups; both groups appeared to navigate the contexts efficiently and not struggle with them for long; we concluded that the informants all had adequate English proficiency to cope with the contexts. After each reading of a context, students were immediately asked if they knew the meaning of the target word. Here, there was a noticeable difference between the two study groups: the eight FAs considerably more frequently stated that they did *not* know the meaning of a word, as compared with the nine FEs. If we consider the noun *plug*, for which the FAs had their poorest look-up performance (five of the eight FA students found the correct sense), seven of the eight admitted to *not* knowing the meaning of the word in the context. If we consider the noun *ticket*, where the end result was an excellent FA look-up performance (all eight FAs found the right meaning), again seven of the eight FAs stated that they did *not* know the meaning when they finished reading the context.

On the other hand, the FEs said much less frequently that they did *not* know the meaning of a word. For only three of the nine target words did a majority of FEs admit that they did not know the meaning after reading the context; for five of the nine words a majority of the FAs admitted lack of knowledge of meaning. (Note that we did not include in the count of those who stated that they did *not* know anyone who said they were *unsure*.) The FEs' poorest performances were on the verb *tax* (only four of nine students selected the right sense in the dictionary) and on the adjectives *sharp* and *mean* (six of nine found the correct meaning for these two words). Despite these performances, immediately after reading the contexts, only two FEs stated that they did not know the meaning of *tax*; likewise, two of the nine FEs said they did not know the meaning of *sharp* or *mean*.

Apart from the numbers indicated above, our impression from direct observation was most certainly that the FEs often *said* that they knew an infrequent sense when they did not. One researcher noted the following while observing an FE student looking up *tax*: "[S/he] ... says [s/he] ... knows the meaning of *tax* but identifies the incorrect sense in the entry. [S/he] ... knows the word in its basic sense ..., but is not really aware of the fact that [s/he] ... doesn't know it in other senses." This occurred several times during the interviews with the FEs. On the other hand, the English-philologist FAs were much more cautious in their self-assessments of their own word knowledge. They were more sensitive to the contexts and more aware when the words they already knew were not

being used in their most common sense. For eight of nine target words of the study, either all FAs or all but one selected the correct infrequent sense in the online BD. However, for four of those eight words, a majority of the FAs said they did *not* know the meaning in the context.

We interpreted these FA behaviors as an alertness to nuances of meaning that the FEs often did not possess. We consider that the FAs were actually helped by their initial lack of familiarity with infrequent senses and their awareness of their own lack of familiarity, because it heightened their attention during the reading of the contexts and during the look-up tasks. Of course, the FEs knew the common meanings of the targeted words in the study, and this little knowledge was a dangerous thing. For only five of the nine target words of the study, either all FEs or all but one selected the correct infrequent sense in the dictionary. When they approached the contexts, the FEs' observed behavior gave us the impression that they knew what they were dealing with, but this often turned out not to be the case and this misconception apparently sometimes contributed to the FE struggles with the look-up process.

While the FEs gave the appearance of navigating the contexts well, they often missed cues about the infrequent word senses — this would hinder them later, during the look-up process. The FEs didn't know what they didn't know. On the other hand, apparently the FAs did not understand the contexts any more or less than did the FEs, but the FAs *did* know what they didn't know, and this sped up their look-up process. As to whether the full context helped the students to decipher the meaning of the target word in infrequent meaning, we consider that the answer depends on the cohort. Most likely, at times the FEs were helped by the context, but at other times, it appeared that the FEs were not sensitive enough to the context they were given. As for the FA cohort, even when the context was not helpful in shedding light on a target word's meaning, it seems the context alerted them that there was something new (to them) going on and this alertness aided the look-up process.

3.5 Comparing navigation of the dictionary entry

Once the participants had read a given context and predicted the target word's meaning, they turned to word look-up in the online BD.

3.5.1 Problems with meaning

As was noted above, a greater sensitivity to the fact that the meanings involved were not the most common ones helped the FAs perform better in the navigation of the dictionary; a lower level of sensitivity hindered the efforts of the FEs. As was previously indicated, for eight of the nine target words of the study, all or all but one of the FAs selected the correct sense in the dictionary. In comparison, for only five words of the study, all or all but one of the FEs selected the correct meaning. We mentioned above that there were three items which

gave the FEs difficulty, the verb *tax* and the adjectives *sharp* and *mean*. For these senses, four out of nine FEs were correct on *tax*, six out of nine were correct on *sharp* and on *mean*. Here we will discuss these three senses and the kinds of problems that arose.

Starting with the FEs and the verb *tax*, five wrongly chose a more common verb sense instead of the correct infrequent sense. Initially (before look-up), these same students defined the word according to its common meaning after reading the provided context. Because they believed that they knew the meaning from the outset, during look-up they were not able to change their minds — despite having read a context containing a different meaning. This was not the case with every FE; three of the four who ended up making the right choice on *tax* initially considered the same incorrect senses as their peers. However, they slowly moved away from these incorrect choices of a common meaning to accept the infrequent, correct sense. One of those who went down this path commented that the meaning was "not the same as what I thought."

For the adjectives *sharp* and *mean*, we observed a very similar process as with *tax*. On *sharp*, four of the nine FE students spent time vacillating between the correct, infrequent sense and a more common incorrect sense; three ultimately chose the correct sense and one stuck with the incorrect. One of the FE students who vacillated but chose correctly commented that the correct and incorrect senses were "not similar," but could both do in the provided context (!). Another FE who chose the incorrect sense, noted that one of the illustrative examples for the correct sense of *sharp* had "outfit" in it, similar to "suit" from the provided context. Despite this similarity, after much deliberation, this student still moved back to a more common but incorrect sense. For *mean*, there was the same type of behavior from two FEs: They went back and forth between the correct meaning and an incorrect one before ultimately choosing the correct, less frequent sense.

What we see from the FEs is, first, a great expenditure of time. They took a long time to give up on their preconceived notions of the meanings of the target words and move over to accept senses that were infrequent and tied to their provided contexts. If they had not been participating in a study where they knew that they were supposed to choose just one sense from the dictionary, it is doubtful that these students would have persisted. Most likely, the difficulty of the task would have caused them to give up; this is what one FE student themselves told us in response to an earlier question of our study (see 3.3 above).

On *tax*, *mean*, and *sharp*, the FAs did better; for all three words seven of eight students chose the correct sense. For *tax*, the students made their choice quickly and most did not even consider a more common sense. For *sharp* and *mean*, there was vacillation between a more common sense and the correct infrequent sense that was similar in nature to the vacillation demonstrated by the FEs. For *sharp*, four FA students considered a more common sense and three of these ultimately chose the correct sense. Here, the difference in performance as compared with the FEs relates to the amount of time that this deliberation process went on; it appeared to us that the FAs were quicker in moving to the sense that became their final answer. That being said, the FAs

took longer to arrive at their answers for *sharp* than they usually took with the other entries. For *mean*, the same state of affairs held: Similar to the FEs, two FAs considered an incorrect sense but quickly moved over and decided upon the correct sense. Overall, for all three of these words, the FAs performed better than did the FEs.

3.5.2 Problems with parts of speech

We have seen above that overall, the philologist-FAs did better in look-up and that the FEs had more problems navigating meaning during the look-up process, which resulted in less frequent selection of the correct sense. As has been noted, we attribute this mainly to sensitivity (or lack thereof) to the new or unknown (to the participants) aspects of meaning displayed in the contexts. In addition, other factors not directly associated with word meaning also affected the performance of both FEs and FAs. As compared with the FAs, the FEs had many more problems recognizing the part of speech of the words in their contexts and in the entries in the BD (see Farina et al. 2019).

Nevertheless, the philologist-FAs, while having vastly fewer difficulties in identifying part of speech, were not spared this problem entirely. The single word that gave the FAs the most difficulty was the noun *plug*; only five of the eight FA participants identified the correct sense. However, two of the three FA students who got this item wrong did not have a semantic problem but a problem recognizing that *plug* from their context was a noun. These two chose a verb sense over the correct noun sense. The verb sense that they chose was semantically correct, linked to the correct noun sense for *plug*. The FEs did better on *plug* than the FAs, with all but one of them getting it correct. Two FEs were distracted by a verb sense; however, one realized the mistake and moved over to the correct noun sense. This good performance by the FEs on the noun *plug* does not mitigate the serious problem they had throughout the study in identifying part of speech. On the other hand, part of speech mostly was not a problem for the linguistically more savvy FAs (see Vrbinc et al. in press).

Another, more minor problem than part of speech for the FEs was a problem in all three verb entries with the canonical form. In general, if an FE saw, for example, the form *fixing* in the provided context, they wanted to see that same form in the dictionary and sometimes encountered difficulties if they did not. The more linguistically aware FAs had no problems in adapting to the different canonical forms that they met in the verb contexts.

3.6 Comparing evaluation of dictionary components

In this section, we discuss the diverse ways in which the FAs and FEs speak about the components of the dictionary, and how both cohorts evaluate the usefulness of various parts of the dictionary in helping them to understand meaning.

3.6.1 Evaluating definition

A strong majority of the participants in both cohorts considered that the definition is the most useful element of the dictionary entry. There was close agreement among the FEs and the FAs on the general usefulness of the definition over the other parts of the dictionary entry. The FAs very often pinpointed which specific words or parts within the definitions were the most useful; the FEs sometimes identified specific parts but less often. There are two entries for which the two cohorts had differing views on the definition, but their comments do not allow for solid conclusions as to why they held the views that they did. For the noun *fix*, the FAs unanimously found the definition to be the most useful part, whereas only four of the nine FEs shared that view. Conversely, on the noun *ticket*, all nine FEs considered the definition to be the most useful, but only two of eight FAs did.

There were two entries where both FEs and FAs did *not* find the definition to be the most useful. For the adjective *mean*, only three people in both cohorts liked the definition. Elsewhere it has been proposed that perhaps the complex nature of the target sense of *mean* (Farina et al. 2019; Vrbinc et al. in press) drove the dislike of the definition. The adjective *rich* also had an unpopular definition:

... used to say that a person's comment or criticism is surprising or amusing because the same comment or criticism could be made about that person.

Only three FEs found it to be the most useful part of the entry and no FAs did: Three FAs found it to be too long and one said that it was not helpful; one FE said that it could be shortened. We speculate that perhaps it might not be the actual length of the definition that was problematic for these users, but rather its pragmatic format. Or, it could simply be that both the FEs and FAs liked the examples for *rich* more and found them to be the most useful part of the entry, edging out the definition. (Note that *rich* was the only one of the nine target words for which a dictionary example matched the context.)

3.6.2 Evaluating examples

Members of both cohorts stated much less frequently that they liked the examples best over the definition. Variants of the statement, "If the definition is good, you don't need examples" were repeated often by the FAs. The philologist-FAs commented on the adequacy of the definition when discussing the examples. On the other hand, the FEs rarely made evaluative comments about the definition quality, but would say, for example, that they only read the examples if they "need" them. In only one instance did an FE say that since the definition was clear, examples were not needed. The FA belief that a good definition does not necessarily need examples was borne out by FA behaviors: We observed that the FAs did not dwell on the examples when they considered that they had already identified the correct meaning based on the definition.

The sole exception to the preference of the two cohorts for definition over examples was the entry for *rich*.

Above (section 3.3) we discussed how the FA students often called for more dictionary examples, longer examples, etc. when they were talking about their overall level of satisfaction with dictionaries. At that early point during the interviews (before the look-up process began), the FEs said almost nothing about examples. However, when the FEs were actually looking at specific examples, they were more forthcoming. They began to make statements similar to the FAs, concerning the need for more examples, longer examples, etc. It is interesting that until they looked at concrete examples, the FEs clearly had not thought much or at all about this dictionary component, in contradistinction to the philologist FAs, whose greater dictionary experience had allowed them to form opinions about examples prior to participating in this study.

It is interesting that it was rare for any FE to say that a combination of different dictionary elements was useful for a given context. Sometimes FAs said that a combination of definition *and* examples was useful, but this happened less frequently than their identifying either definition alone or examples alone as most useful.

3.6.3 Evaluating square-bracketed information in illustrative examples

As one might expect, there was a great diversity of opinion about the information provided in the square brackets of some (but not all) examples. There appeared to be a difference in kind between the comments of the FEs and the FAs. The FAs gave specific, detailed, and diverse evaluations of individual pieces of square-bracketed information within the examples. At times, different FAs liked or disliked the same piece of information, but what is telling is that they had specific judgments about how the information fit within the broader scheme of examples and entries, how the information contributed (or not) to the betterment of an example or dictionary entry. We labeled these types of comments as "stylistic assessments": The FAs were deciding how the square-bracketed information fit into the whole of a dictionary entry or at least into the ensemble of a sense's examples.

The FEs, after their attention had been drawn to it, certainly would comment on the square-bracketed information. They would state that such information was "useful" or "helpful." At times they stated that without it, they would (or would not) understand an example. One FE said that this was the first time they had ever noticed brackets in a dictionary, another said that the equal sign [=] sometimes present within brackets was confusing, yet another characterized such elements as square brackets in a dictionary as "noise." The FE comments about square-bracketed information reflect that, as they approached the dictionary of the study, they had no schema in mind for dictionaries; they did not know what to expect. So, each encounter with a square bracket took place as if no square bracket — or no dictionary — had ever been seen before.

The FAs looked at the dictionary entry as a whole consisting of various familiar components. When they evaluated the square-bracketed information within the examples, certainly they thought about how such information contributed to a specific example, but they were also assessing the functionality of the bracketed information within the sense, within the full entry, and within the dictionary as a whole.

4. Looking up *sharp*: Conclusions

These two in-person qualitative studies, where the users were directly observed and asked what they thought as they used an online dictionary, allow us to obtain information that would be impossible to obtain otherwise. At looking up *sharp* and at looking up the other target senses of these two studies, certainly the 3rd-year undergraduate English majors from the Faculty of Arts (FA) at the University of Ljubljana excelled. Their 3rd-year undergraduate School of Economics and Business (FE) peers performed well but their approach to the look-up tasks was radically different and consequently their end results were not the same.

While almost any lexicographer would predict that English philology students would have more "dictionary-awareness" and perform better than general educated dictionary users, it is important to understand exactly how and why the former group performed better. It is important as a factor that, we hope, could influence the construction of future online dictionaries.

First, coming to the study with a schema of what a dictionary is, what parts it consists of and how to make one's way through it, was essential to the FAs' consistent success during the look-up processes of their study. The FAs knew the terrain very well and they had traveled it frequently. On the other hand, as the study involving them shows, the FEs were using their GPS and often had to "recalculate." For the FEs, their sincere efforts at navigating the dictionary were often unsatisfying and frustrating; parts of the dictionary to them were just "noise." The lesson here is that the general dictionary user is always using a GPS and the lexicographer should never construct any lexicographic product as if the user knows the terrain.

Taking *sharp* as an example, we proposed (Farina et al. 2019) that the positioning of the correct information (sense #12 of 13 senses) was the main obstacle to the FE users' efforts to find the correct information:

12

informal: stylish or fashionable

- He's a *sharp* dresser.
- a *sharp* outfit
- You're looking very *sharp* today.

If this proposal is correct, it means that the FE users had to travel farther down the entry than expected and many were not persistent enough to do so. In

addition, the third sense, "noticeable," close to the beginning of the entry, deceived some FEs and they did not continue down the list of senses once they had landed on sense #3. Other available online learner's dictionaries also present this sense far down in the entry; the *Oxford Advanced Learner's Dictionary* has this as sense #12 out of 15; *Longman Dictionary of Contemporary English* has it as #12 of 16, and *Macmillan English Dictionary* as #8 of 11. Irrespective of the quality of information to be found, many dictionary users simply will not travel this far. Nevertheless, the content of the sense remains very important, because if a user does manage to get all the way to this sense, it is necessary for the information there to be useful. The *Oxford* sense has the following:

12. [usually before noun] (of clothes or the way somebody dresses) fashionable and new
- *The consultants were a group of men in sharp suits.*
 - *Todd is a sharp dresser.*

We consider that this would have been more helpful for our FEs, because the examples are full sentences; many FEs (as well as some FAs) commented that they do not like short dictionary examples (Farina et al. 2019; Vrbinc et al. in press). The longer *Oxford* examples do not take up much more screen space.

Another factor that might make access to the *sharp* information easier is that *outfit*, which appeared in the BD example is a hypernym whereas *Oxford* used *suit*, a hyponym. Generally, the FAs liked the example with *outfit* and appeared familiar with the word; on the other hand, it is not clear to us that most/all FEs were fully familiar with the hypernym.

Apart from the issues of word choice within illustrative examples and navigation of long entries, dictionaries are not personal to the FEs. Compared to the FAs, the FEs do not have a preferred dictionary, dictionaries are nameless, and a dictionary is not a "thing." This matters immensely; the dictionary has already been integrated into the life experience of the FAs and this helps drive what happens when they consult one online. If there is any hope at all that lexicographic tools will become integrated into the lives of well-educated people like the FEs, then this should change.

What does this mean for the lexicographer and for future lexicographic work? It means that the lexicographer should be extremely concerned about the success of users who are like the FEs. The lexicographer cannot make *any* assumptions about the familiarity of the educated general user with the lexicographic medium. Online lexicographic media of today are still too deeply rooted in the print dictionaries of the past. They are still more oriented toward persons with experience like the FAs, who can meet practically any linguistic or lexicographic challenge thrown at them. The target users (certainly not the FAs) of any modern online dictionary do not know the terrain, do not have any interest in learning it, and will not learn it. They must be provided with an online format that requires as few navigation skills as possible, an "uber-GPS"; any small flaw in the presentation of the information will without doubt cause

some users to go astray. And if they are led astray, not only will they receive minimal-to-no benefit from the dictionary at hand, but they will hesitate before using any online dictionary again.

Acknowledgements

The authors acknowledge the project, *Dictionary User Groups: What They Can Teach Lexicographers* (grant number: BI-US/17-18-033), which was financially supported by the Slovenian Research Agency. They also acknowledge the approval (2 March 2018) of the New Jersey City University (NJCU) Institutional Review Board for the Protection of Human Participants in Research, and the approval (5 March 2018) of the Ethics Committee of the Faculty of Arts, University of Ljubljana. The authors thank Peter Sokolowski, Editor at Large of Merriam-Webster, who is always willing to answer our questions and for the present endeavor provided information regarding *The Britannica Dictionary*. We are appreciative of the meticulous efforts of the two anonymous reviewers, who helped us to improve this article. Finally, the authors are forever grateful to the undergraduate study participants from the English Department in the Faculty of Arts and from the School of Economics and Business at the University of Ljubljana.

Endnotes

1. This abbreviation, derived from the previous name of the School of Economics and Business (i.e., the Faculty of Economics), is being retained here for the convenience of readers of Farina et al. (2019), which describes these students and uses the same "FE" abbreviation.
2. This dictionary is identical to what it was under the previous title. This information was confirmed by Peter Sokolowski, Editor at Large of Merriam-Webster. Persons accessing the former website of this dictionary (<http://learnersdictionary.com/>) are redirected to the new website, <https://www.britannica.com/dictionary>.
3. Please refer to Farina et al. (2019) for a lengthier discussion of previous studies.
4. For more information about the educational development of both groups' language proficiency, consult Farina et al. (2019) and Vrbinc et al. (in press).
5. The full explanation of the investigation of sense frequency in preparation for the two studies can be found in Farina et al. (2019).

References

A. Dictionaries

- The Britannica Dictionary*: <https://www.britannica.com/dictionary>. Merriam-Webster Learner's Dictionary. Accessed on 3 July 2022. (BD)
- Longman Dictionary of Contemporary English*: <https://www.ldoceonline.com/>. Accessed on 16 October 2022.

Macmillan English Dictionary: <https://www.macmillandictionary.com/>. Accessed on 16 October 2022.
Oxford Advanced Learner's Dictionary: <https://www.oxfordlearnersdictionaries.com/>. Accessed on 16 October 2022.

B. Other literature

Assigning CEFR Ratings to ACTFL Assessments: https://www.actfl.org/sites/default/files/reports/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf. Accessed on 5 July 2022.

Béjoint, H. 2010. *The Lexicography of English: From Origins to Present*. Oxford: Oxford University Press.

Caulfield, J. 2022. *How to Do Thematic Analysis: A Step-by-Step Guide & Examples*. <https://www.scribbr.com/methodology/thematic-analysis/>. Accessed on 11 July 2022.

De Schryver, G.-M. and D.J. Prinsloo. 2011. Do Dictionaries Define on the Level of their Target Users? A Case Study for Three Dutch Dictionaries. *International Journal of Lexicography* 24(1): 5-28. doi:10.1093/ijl/ecq045.

Farina, D.M.T.Cr., M. Vrbinc and A. Vrbinc. 2019. Problems in Online Dictionary Use for Advanced Slovenian Learners of English. *International Journal of Lexicography* 32(4): 458-479. doi.org/10.1093/ijl/ecz017.

Finegan, E. 2011. *Language: Its Structure and Use*. Sixth edition. Boston, MA: Wadsworth, Cengage Learning.

Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7-11 August 2012, Oslo*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Gillham, B. 2000. *The Research Interview*. London: Continuum.

Gouws, R.H. 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177. doi.org/10.5788/24-1-1256.

Hannabuss, S. 1996. Research Interviews. *New Library World* 97(5): 22-30. doi.org/10.1108/03074809610122881.

Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. Tübingen: Max Niemeyer.

Hatherall, G. 1984. Studying Dictionary Use: Some Findings and Proposals. Hartmann, R.R.K. (Ed.). 1984: 183-189.

Herbst, T. 1996. On the Way to the Perfect Learners' Dictionary: A First Comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography* 9(4): 321-357. doi.org/10.1093/ijl/9.4.321.

Lew, R. 2011. Space Restrictions in Paper and Electronic Dictionaries and their Implications for the Design of Production Dictionaries. Bański, P. and B. Wójtowicz (Eds.). 2011. *Issues in Modern Lexicography*. München: Lincom Europa. Accessed on 14 July 2022. https://repozytorium.amu.edu.pl/bitstream/10593/799/1/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf.

Lew, R. and G.-M. de Schryver. 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359. doi:10.1093/ijl/ecu011.

Lorentzen, H. and L. Theilgaard. 2012. Online Dictionaries — How Do Users Find Them and What Do They Do Once They Have? Fjeld, R.V. and J.M. Torjusen (Eds.). 2012: 654-660.

Maguire, M. and B. Delahunt. 2017. Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars. *AISHE-J* 9(3): 1-14.

- Merriam, S.B.** 2009. *Qualitative Research: A Guide to Design and Implementation*. (Revised and expanded from *Qualitative Research and Case Study Applications in Education*). San Francisco, CA: Jossey-Bass.
- Mitchell, E.** 1983. *Search-do Reading: Difficulties in Using a Dictionary*. Formative Assessment of Reading, Working Paper 21. Aberdeen: Aberdeen College of Education.
- Newton, N.** 2010. The Use of Semi-Structured Interviews in Qualitative Research: Strengths and Weaknesses. *Exploring Qualitative Methods* 1(1): 1-11.
- Qu, S.Q. and J. Dumay.** 2011. The Qualitative Research Interview. *Qualitative Research in Accounting & Management* 8(3): 238-264. doi.org/10.1108/11766091111162070.
- Rubin, H.J. and I.S. Rubin.** 1995. *Qualitative Interviewing: The Art of Hearing Data*. Second edition. London: Sage Publications.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322. doi.org/10.5788/25-1-1301.
- Ryan, Gery W. and H.R. Bernard.** 2003. Techniques to Identify Themes. *Field Methods* 15(1): 85-109. doi:10.1177/1525822X02239569.
- Tono, Y.** 1984. *On the Dictionary User's Reference Skills*. Unpublished B.Ed. dissertation. Tokyo: Tokyo Gakugei University.
- Učni načrt Angleščina.* 2008. http://www.mss.gov.si/fileadmin/mss.gov.si/pageuploads/podrocje/ss/programi/2008/Gimnazije/UN_ANGLESCINA_gimn.pdf. Accessed on 11 July 2022.
- Vrbinc, M., A. Vrbinc and D.M.T.Cr. Farina.** In press. The Lexicographer's Dream Audience: Dictionary Use among English Majors at a Slovenian University. *International Journal of Lexicography*.
- Wingate, U.** 2002. *The Effectiveness of Different Learner Dictionaries*. Tübingen: Max Niemeyer.

Appendix A: List of Contexts and their References

Context 1

This small desert nation of six million opened its doors to the newcomers but was quickly overwhelmed as they gobbled up jobs, **taxed** scarce water resources and forced schools into double shifts. About two-thirds of the refugees are squatting in Jordanian cities and villages, but the pathos and problems are most profound in Zaatari [...].

Context 2

ROVANIEMI, Finland — A man arrived at the police station here in 2011 with an unusual tip. He told the police that a Singaporean man was **fixing** matches with the local professional soccer team. The police were incredulous.

Context 3

Back when Patrick had a job at an auto-parts store and as a banquet server, his morning routine involved driving to Lawrence before work and **scoring** his daily fix.

Then he would shoot up with heroin or fentanyl at the wheel of the car while driving back to New Hampshire.

Context 4

With no discussion of a business model and only vague statements that offer no numbers, investors will be unlikely to take a **pitch** like this seriously. You need to put numbers to the idea and make the business case. Moreover, if you are going to ask for money, investors need to know how much, what it will be used for and what kind of return will be generated.

Context 5

Microsoft says that the wireless sharing is a new way to discover music. But you can't shake the feeling that it's all just a big **plug** for Microsoft's music store. If it's truly about the joy of music discovery, why doesn't Microsoft let you buy your discoveries from any of the PlaysForSure stores?

Context 6

The state's Republican chairman, Edward F. Cox, offered a respectful, if cautious statement about Mr. Kolb's candidacy. 'We're glad he has formally announced his intentions,' Mr. Cox said, before saying he was excited by the possibility of other candidates, too. [...]

He added that the party would be 'working collaboratively with our county chairs over the coming weeks and months to put together the statewide **ticket**.'

Context 7

Her rock 'n' roll friends might have expected a hip 'n' cool outfit for her English country wedding. But it was her husband, Jamie Hince, the guitarist from The Kills, in his sharp blue Yves Saint Laurent suit, who brought a touch of music-world fantasy.

Context 8

My biggest help in mothering with MS? I give myself permission to not be perfect. I let myself adjust things according to what kind of day I'm having. When I can't go for walks with Jerry, I lie on the bed and watch him play Nintendo or read with him or play a mean game of Checkers [...]. I am still involved very much in my kids' lives.

Context 9

'Jesus,' Arden says, stacking his hands on the top of his head. 'I can't win.'

Oh, now that's rich. 'You can't win? *You?* Arden Moss? You've already won, idiot. You have everything you've ever wanted in life, all handed to you on a silver platter.'

References for Contexts

- [1] **Rudoren, Jodi.** 2013. A Lost Generation: Young Syrian Refugees Struggle to Survive. *The New York Times*, 8 May 2013. Accessed on: 15 January 2018. <https://www.nytimes.com/2013/05/09/world/middleeast/syrian-refugees-in-jordan-struggle-to-survive.html>.
- [2] **Hill, Declan.** 2014. Inside the Fixing: How a Gang Battered Soccer's Frail Integrity. *The New York Times*, 1 June 2014. Accessed on: 23 January 2018. <https://www.nytimes.com/2014/06/02/sports/soccer/match-fixing-gang-strikes-at-soccers-fragile-legitimacy.html>.
- [3] **Seelye, Katharine Q.** 2018. 1 Son, 4 Overdoses, 6 Hours: Drug Deaths Draw the Most Notice, but More Addicted People Live Than Die. For Them and their Families, Life Can Be a Relentless Cycle of Worry, Hope and Chaos. *The New York Times*, 21 January 2018. Accessed on: 12 February 2018. <https://www.nytimes.com/2018/01/21/us/opioid-addiction-treatment-families.html>.
- [4] **Roth, Carol.** 2012. A Pitch to Collect Charity from Casino Gamblers. *The New York Times*, 24 April 2012. Accessed on: 15 January 2018. <https://boss.blogs.nytimes.com/2012/04/24/a-pitch-to-collect-charity-from-casino-gamblers/>.
- [5] **Pogue, David.** 2006. Trying Out the Zune: iPod It's Not. *The New York Times*, 9 November 2006. Accessed on: 15 February 2018. <https://www.nytimes.com/2006/11/09/technology/09pogue.html>.
- [6] **Mckinley, Jesse.** 2017. Republican Assemblyman Enters Race to Challenge Governor Cuomo. *The New York Times*, 12 December 2017. Accessed on: 24 January 2018. <https://www.nytimes.com/2017/12/12/nyregion/brian-kolb-cuomo-governor-new-york.html>.

- [7] **Menkes, Suzy.** 2011. 3 Weddings and a Dior Carnival. *The New York Times*, 4 July 2011. Accessed on: 3 February 2018. <https://www.nytimes.com/2011/07/05/fashion/first-dior-couture-show-since-john-gallianos-departure-offers-confetti-but-no-finesse.html?mtrref=www.google.com&gwh=451F4172A6491A2CF6C456ADB08888CF&gwt=pay>.
- [8] **Nichols, Judith Lynn.** 2000. *Living Beyond Multiple Sclerosis: A Women's Guide*: 134. Nashville, Tennessee: Hunter House. Accessed on: 23 February 2018. <https://goo.gl/pXwYya>.
- [9] **Banks, Anna.** 2015. *Joyride*: 46. New York: Feiwel & Friends. Accessed on: 14 January 2018. <https://goo.gl/DoxqmB>.

Clifford Simango. *Concise Chindau Dictionary*. 2019, xii + 263 pp. ISBN: 978-1-77929-682-5. Harare: Dream Discovery. Price US\$9.

As a language that was first described and written by Portuguese and English explorers, traders, and missionaries, ChiNdau possesses the sort of intermittent lexicographic tradition seen in several other indigenous African languages. A vocabulary wordlist in Wilhelm Bleek's (1856) *The Languages of Mozambique* that records some Ndau words and their glosses under the Sofala column set the stage for lexicographic development in Ndau. Whereas Bleek's effort may sometimes fail to qualify as a 'dictionary' in several contested senses of the word, succeeding bilingual paper dictionaries such as Dysart and Orner's (1915) *ChiNdau–English English–ChiNdau Vocabulary with Grammatical Notes* and Manuel dos Anjos Martins's (2013) *Elementos Da Língua Ndau: Gramática, Literatura Oral e Dicionário* provided a springboard for a meaningful development of Ndau lexicography in Zimbabwe and Mozambique. The dictionary under review, Clifford Simango's (2019) *Concise ChiNdau Dictionary* (hereafter CCD), is then a fourth lexicographic endeavour to be published in Ndau.

It is thus not only a welcome addition to the growing Ndau corpus but equally a strategic intervention "to document ChiNdau words" as it "is suffering from language degradation", as Simango (2019, iii-iv) says in his introduction. Because "ChiNdau-speaking students are taught Shona at school" notwithstanding its constitutionally acquired official language status in Zimbabwe, the CCD was excellently designed to be a significant turning-point in the lexicographic history of Ndau. The CCD is the first Ndau dictionary written for use in schools in view of conscious language development events transpiring at both macro and micro language planning levels in Zimbabwe. At a macro level, Ndau was constitutionally recognised as one of the 16 official languages that must be legally supported to function in all high status domains. Thus, with government authorisation and sufficient activism from mother-tongue speakers, it has somewhat penetrated domains such as broadcast media on radio and television, parliament and education. At a micro level, higher and tertiary education institutions such as Midlands State University, Great Zimbabwe University, Masvingo Teachers' College and Marymount Teachers' College have, to varying degrees of success, already introduced and implemented Ndau training personnel programmes in their institutions. In view of such a promising background for Ndau's development in Zimbabwe, a Professor of Biomedical Pathology at the University of Zimbabwe, Clifford Simango (now deceased), diverged from his own area of speciality to, despite his limited lexicographic skills, knowledge and experience, compile after a century the second bilingual ChiNdau–English and English–ChiNdau dictionary.

The CCD is a medium-sized, general-purpose, bilingual dictionary with 263 pages of main text and 13 pages of front matter. In line with normal dictionary practice, the compiler acknowledges, in one front matter text, seven

lexicographic and other works that were "used extensively in compiling the dictionary". The listed works are *Nduyo Dzokudira* (1907), *ChiNdau–English English–ChiNdau Vocabulary with Grammatical Notes* (1915) and the *Holy Bible* in Nda (1957), *Standard Shona Dictionary*, *Duramazwi Guru ReChiShona*, *A Practical Ndebele Dictionary* and *Scholar's Zulu Dictionary*. It is, of course, unfortunate that Simango did not refer to Manuel dos Anjos Martins's (2013) *Dicionario* section in his *Elementos Da Língua Nda: Gramática, Literatura Oral e Dicionário*. It is apparent, to the extent that dictionary data is distilled from previous works in Nda and other languages, that CCD is not a corpus-based dictionary but a lexicographic effort that was largely influenced by its compiler's intuition as a mother-tongue speaker. It also depended on his own personal judgment and preference in selecting data sources, an approach that, though somewhat understandable given the lack of a Nda corpus, is still regrettable because the front matter texts do very little to describe and explain the principles followed in compiling the dictionary. For instance, some reviewers and ordinary dictionary users might require further explanations on the lexicographic benefit/s of 'extensively consulting' Shona, Ndebele and Zulu dictionaries for a medium-sized bilingual Nda dictionary such as the CCD. Again, in the absence of a befitting justification, the value of centuries-old evangelical literature such as *Nduyo Dzokudira* (1907) and dictionaries such as *Chindau–English English–ChiNdau Vocabulary with Grammatical Notes* might be critiqued against the needs of present-day school users of Nda lexicographic products. To what extent is the CCD responsive to the needs and expectations of its target users?

Indeed, notwithstanding the dictionary's limited sources of data, the CCD is a very useful lexicographic product that functions "as a reference book for teachers and students of Nda" not only in the education sector (primary, secondary and tertiary education) but also in the media, legislature and the society at large.

Firstly, pursuant to its interventionist goal of stopping the degradation of Nda, the dictionary has a combined total of approximately 7703 entries where around 4124 entries feature in the ChiNdau–English section whereas 3579 lemmas appear in the English–ChiNdau section. At a microstructural level, the dictionary subscribes to commonly accepted modern standards by consistently utilising different kinds of typographical conventions in each treatment of an ordinary headword. This helps in illustrating both the morphological and grammatical links between the headword (which is always in a larger bold lower case), tone (pronunciation), word type (part of speech), number (singularity and plurality), definition as well as illustrative example/s (consistently in italics and bold when in Nda but just italicised when in English), for example,

muisa, [LHH] n., pl. vais-, man. *Pane muisa uri kutengese urimbo.*

There is a man who is selling bird lime.

chiro, [H] n., pl. zvir-, fierce animal; thing.

The treatment of microstructural elements of noun entries in the CCD provides a solid foundation upon which orthographic and morphological properties of Ndau can be appreciated. For example, indicating plurality on the singular nouns above is instructive not only on the morphological structures pertaining to the **mu-/va-** or **chi-/zvi-** prefixal agreement but also on the orthographic conventions governing the language regarding spelling.

Secondly, the compiler commendably adheres to best lemmatisation practices by ensuring that all noun word lemmas i.e. noun headwords are entered with their prefixes. Following standard practice for Bantu languages, the CCD also utilises stem lemmatisation by stripping verbal prefixes from their stems which are then entered as headwords in a dictionary as seen in three examples below,

-beuka, [LLH] v.i., be open.

-beura, [H] v.t., open e.g. door; school. *Chikora chakabeurwa zona. The school was opened a day before yesterday.*

-beurira, [H] v.t., open for someone or something. *Endai mwobeurire mwombe kuchibaya. Go and open the cattle kraal.*

Research reveals that stem lemmatisation may, in appropriate environments, prove superior, more scientific and more economical than word lemmatisation (Van Wyk 1995). Verbs, however, typically have a productive series of inflectional suffixes which could needlessly multiply the number of entries. In general, Simango has taken an intermediate approach, giving the three derivatives of **-beu-** (open) above, but omitting others, such as **-beurika** (be openable), **-beurwa** (be opened). This is one aspect among many others that future compilers of Ndau dictionaries might need to decide and improve on to enhance dictionary quality.

Thirdly, in view of Ndau's marginalisation and "degradation" as well as the dictionary's target users in Zimbabwe, it can be argued that the dictionary's lemma stock offers a fair coverage of the language's lexicon. The dictionary describes diverse vocabulary items such as nouns, verbs, pronouns, adjectives, demonstratives, quantitatives and ideophones that broadly identify with the targeted users' needs. Further to this, the CCD commendably provides a full description of the language as it includes several loanwords from Zulu, Shona and English. But unlike preceding dictionaries such as Dysart and Orner's (1915) *ChiNdau-English and English-ChiNdau Vocabulary with Grammatical Notes* which marked Zulu loanwords by a superscript (z), the CCD does not indicate the origin of borrowings in Ndau, probably because they have already been naturalised as permanent features of the Ndau lexicon. The examples of Zulu headwords include 68 click words as exemplified below and several non-click lexemes such as **chibochwa** (slave), **chibhedhlera** (clinic) derived from **isibotshwa** and **isibhedhlela** respectively. Some English loanwords include but are not limited to **alithari** (altar) and **awa** (hour) derived from **altar** and **hour** respectively.

Without a corpus approach, it becomes difficult to know how to limit the increasing number of borrowings from English or any other language with which Ndaui is in contact. However, to comply with certain macrostructural requirements for larger projects, future Ndaui lexicographers might need to select lemmata based on principled criteria set to better respond to the different needs of target users. This suggests the imperativeness of conducting some prior dictionary research of some sort to improve the compilation of new and big size dictionaries in Ndaui or to enhance the revision of the CCD.

Fourthly, examples cited in this review show that the CCD follows proper defining principles where all Ndaui headwords are provided with English equivalents and vice versa. Because a dictionary explains the meaning of words while simultaneously demonstrating how they combine with each other to form phrases, sentences and clauses, the dictionary succeeds in giving simple definitions that provide 'all the answers to the questions that made a user consult the dictionary' (Gouws and Prinsloo 2005: 144). Similarly, the compiler uses examples that provide additional contexts to strengthen given definitions as seen in provided examples. Although Simango does not explain the principles guiding their use, it is apparent that appropriate examples can actually be considered natural extensions of the definitions as has been illustrated already. Taken together, the qualities of the CCD, for instance, good definitions, proper examples, legible fonts, layouts and spacing of entries, presentation simplicity, as well as the medium and easy to handle size contribute in qualifying it as a user-friendly dictionary.

Fifthly, the lexicographer commendably attempts to reform and standardise the Ndaui spelling system. To that extent, the CCD's guide to the dictionary does well to list the single consonants, digraphs and trigraphs that feature in the Ndaui alphabet. All single consonants are included except the lateral /l/ which unfortunately precludes the inclusion of possible Ndaui headwords such as **lamba** (barrenness), **-lemuka** (cleverness) and **-lunga** (niceness). However, the CCD makes up by incorporating lateral sounds which it represents by graphemes /hl; dhl; ndhl/ in these entries:

hlanga, [H] n., pl. mahl-, dry maize stalk. *Isai mahlanga muchibaya kuti mwombe dzirye*. Put maize stalk in the cattle kraal as fodder for cattle.

madhleyo, [LHH] n., pl. only, grazing area; pasture land. *Majaha vaende kooshe mwombe kumadhleyo*. The boys have gone to herd cattle at the grazing area.

khandhlera, [LHH] n., pl. makh-, candle. *Basai khandhlera mumphatso mukhanye*. Light the candle so that there is light in the house.

The labelling of **madhleyo** as plural only instead of singular only is one of the few mistakes observed in this generally well-edited dictionary. Future dictionary compilers might therefore need to pay particular attention to minor descriptive, grammatical and typographical errors which can negatively impact

the quality of dictionaries in Nda.

In addition, the CCD deals decisively with clicks and aspirated sounds in Nda. With regard to clicks which have been variously and often inconsistently represented by different symbols in previous publications, for instance, /x/ as in **nxwadh**i (letter), /c/ **ncwadh**i (letter), the CCD adopted the voiceless uvular plosive /q/ as the standard orthographic symbol to represent all clicks as shown in these examples.

-qonda, [H] v.i., be straight. *Pato rinori rakaqonda. This road is straight.*

-qhasa, [L] v.t., clarify; explain.

-gqoka, [L] v.t., wear; put on clothes; dress oneself. *Uya ugqoke ngubo yako tombi. Come and wear your dress little girl.*

ngqondo, [L] n., sing. only, mind, power of reasoning.

-nqaka, [H] v.t., catch anything thrown at you e.g. ball. *Nqake bhora rendiri kuposhere kweuri. Catch the ball which I am throwing at you.*

nqwadhi, [L] n., pl. idzi, letter.

The uvular /q/ can stand on its own in **-qonda**, be aspirated in **-qhasa** and be prenasalised in **nqaka** and **nqwadh**i. The dictionary's contribution towards the standardisation of Nda spelling is also witnessed through the aspiration of all voiceless stops /t, k, p/ as in **thika** (hyena); **k^hamba** (leopard) and **p^hinda** (repeat) which were also seriously misrepresented by other writers of Nda. Evidently, aspiration remains an orthographic challenge in Nda that is best exhibited by different spelling systems on social media and other platforms in Zimbabwe (Sithole 2020). When the potential that dictionaries contribute towards the standardisation of languages (Hadebe 2002) is considered, it is therefore hoped that some form of standardisation of Nda spelling conventions will be achieved as the CCD entrenches its position as a lexicographic reference tool in education and other high status domains in Zimbabwe.

However, it may also be observed that the compiler adopts some peculiar sound combinations that are often difficult to explain both at phonetic and orthographic levels. For example, he uses /nts, ntsw, thsw/ to represent lexemes such as **ntsanangudzo** (explanation), **ntsentswa** (wood shaving), and **mupfuthswa** (dried vegetables). While this could be a mere reflection of dialectal variation in Nda and perhaps an acknowledgement of the variable nature of orthographies the world over, it is instructive that the prenasalisation of the affricate /ts/ as in these given examples is extremely rare in Nda, Shona and related languages. This may therefore imply that the compiler, as a new lexicographer without relevant training not only in the art, craft and science of making dictionaries but also lacking a sound basis in phonetics and phonology, made an orthographic error. Such an error creates a mismatch between the Nda words which appear in the CCD and those that are spoken and used in Nda speaking communities' daily communicative situations. One last observation on spelling is

that the compiler misspells **sonto** (Sunday) which was directly taken from Dysart and Orner's *ChiNdau–English English–ChiNdau Vocabulary with Grammatical Notes* (1915). Whereas this is not a good lexicographic practice, it is noted that the entry **sonto** might either be changed to **sondo** or be replaced by **soto**. Based on the reviewer's intuition as a mother-tongue speaker of Ndaou too, both suggestions enjoy popular usage in the Ndaou speaking community.

Finally, in appreciation of Clifford Simango's excellent work in contributing to the development of Ndaou lexicography and because Ndaou now has two bilingual dictionaries, one with translations into English and one into Portuguese, the next step could be the building of a corpus to assist with the creation of a monolingual Ndaou dictionary and other specialised school dictionaries.

References

- Bleek, W.H.I.** 1856. *The Languages of Mozambique*. London: Harrison and Sons.
- Chimhundu, H. (Ed.)**. 2001. *Duramazwi Guru ReChiShona*. Harare: College Press.
- Dent, G.R and C.L.S. Nyembezi**. 1995. *Scholar's Zulu Dictionary*. Third edition. Pietermaritzburg: Shuter and Shooter.
- Dysart, J.P. and A.J. Orner**. 1915. *ChiNdau–English English–ChiNdau Vocabulary with Grammatical Notes*. Mt. Selinda: American Board Mission, Rhodesian Branch.
- Gouws, R.H. and D.J. Prinsloo**. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: AFRICAN SUN MeDIA.
- Hadebe, S.** 2002. *The Standardisation of the Ndebele Language through Dictionary-making*. Unpublished D.Phil. Dissertation. Harare and Oslo: University of Zimbabwe and University of Oslo.
- Hannan, M.** 1959. *Standard Shona Dictionary*. Harare: Literature Bureau.
- Martins, Manuel dos Anjos**. 2013. *Elementos Da Língua Ndaou: Gramática, Literatura Oral e Dicionário*.
- Pelling, J.N.** 1971. *A Practical Ndebele Dictionary*. Harare: Longman Zimbabwe.
- Sithole, E.** 2020. Challenges Encountered on the *Rekete ChiNdau — Leave A Legacy* Facebook Page. *Language Matters* 51(1): 3-21. <https://doi.org/10.1080/10228195.2020.1738533>.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.

Emmanuel Sithole
School of Languages and Literatures: African Language Studies
Rhodes University
Makhanda
and
Faculty of Humanities
Northwest University
Potchefstroom
South Africa (emmanesu@gmail.com)

Heming Yong. *English Lexicography from British Tradition to World Englishes*. 2022, viii + 194 pages. ISBN 978-1-4331-8696-7 (Hardback), ISBN 978-1-4331-8697-4 (eBook PDF), ISBN 978-1-4331-9258-6 (ePub). New York/Bern/Berlin/Brussels/Oxford/Vienna: Peter Lang. Price: US\$ 97.80 (Hardback), US\$ 94.95 (eBook).

Broadly speaking, English lexicography encompasses dictionaries of British English and other native varieties such as American English, Australian English, Canadian English, etc. Dating from the Anglo-Saxon period, it has a history of at least 400 years (Cowie 2009). Since the 1970s, many monographs have explored English lexicography from various perspectives: Some provide a comprehensive account (e.g. Zgusta 1971; Landau 1989; Svensén 1993; Atkins and Rundell 2008; Svensén 2009); some give a historical description (e.g. Collison 1982; Cowie 2002, 2009; Béjoint 2010; Miyoshi 2017; Yong and Peng 2021); and some expose the existing limitations of present-day English dictionaries and offer some solutions to these issues (e.g. Dixon 2018). There are also some volumes focusing on the history of world lexicography (e.g. Considine 2019). However, except for Martin (2019) on American English lexicography, little research has been conducted to investigate the development of English lexicography in other English-speaking countries.

English Lexicography from British Tradition to World Englishes, the volume in question, explores the evolution of English lexicography from British tradition to other major native varieties, viz. American English, Australian English, Canadian English, New Zealand English, and South African English. It reveals the historical relations and inheritance between British lexicographical tradition and the development of English lexicography in those varieties, giving an integrated, diachronic, and comprehensive account of how English lexicography in those varieties started and has gradually evolved into modern shapes.

With a total of 194 pages, this volume comprises seven chapters, plus a bibliography and an index. The first two chapters, which make up half of the whole volume, depict British English lexicographical tradition and American English lexicography respectively. The following four chapters illustrate English lexicography with regard to other major native varieties such as Australian English, Canadian English, New Zealand English, and South African English. The last chapter describes the prospects for the development of world English lexicography.

Chapter One documents British lexicographical tradition and the development of world English lexicography. Adopting a linguistic approach, this chapter traces the evolutionary trajectory of British English lexicography from its archetype to the prescriptive, the historical, the descriptive, and finally to the cognitive paradigms. It begins with an investigation into the seeds of Latin lexicography in Old English glossaries such as *The Leiden Glossary*, and the genesis of English bilingual lexicography from Latin traditions, especially the making of Latin–English and English–Latin bilingual dictionaries that embrace

a Latin heritage. Next, it explores the prescriptive tradition of Latin grammar and the making of monolingual English dictionaries in Britain. The publication of Samuel Johnson's monumental *A Dictionary of the English Language* (1755) signifies the firm establishment of the prescriptive paradigm in lexicography. Then, it examines the European philological tradition in English lexicography and the making of historical dictionaries in Britain. This phase is marked by *The Oxford English Dictionary* (Murray et al. 1884–1928), which illustrates the consistent, comprehensive, systematic and scientific implementation of the historical principles in dictionary-making and signifies the firm establishment and adequate application of the historical paradigm to English dictionary-making (p. 18). Subsequently, it surveys the descriptive linguistic paradigm and the development of English descriptive dictionaries. Philip Babcock Gove's (1902–1972) *Webster's Third New International Dictionary of the English Language, Unabridged* (commonly known as *Webster's Third*, or W3, 1961), is regarded as 'the first dictionary that applies the descriptive principle in a systematic and down-to-earth manner' (p. 21). Finally, it introduces the cognitive linguistic paradigm and the making of active English dictionaries, viz. British learner's dictionaries, which have undergone three generations of evolution represented by the *New Method English Dictionary* (West and Endicott 1935), *A Grammar of English Words* (Palmer 1938) and *The Idiomatic and Syntactic English Dictionary (ALD1)* (Hornby et al. 1942) in the first generation, *ALD2* (Hornby et al. 1963), *ALD3* (Hornby et al. 1974) and *LDCE1* (Procter 1978) in the second generation, and the *OALD4* (Cowie 1989) and *LDCE2* (Summers 1987) as the foundation and *COBUILD1* (Sinclair et al. 1987) as its representative work in the third generation. This chapter concludes with an analysis of British lexicographical tradition and its impacts upon the development of world English lexicography.

Chapter Two examines American English lexicography and its inheritance and reformation of British lexicographical tradition. It first explores the socio-linguistic background for the emergence and development of American English lexicography, and traces the extension and reformation of British lexicographical tradition in the making of first American English dictionaries. Then, it elaborates on Noah Webster's dictionaries and their leading innovation upon British lexicographical traits, and the American "War of Dictionaries" and Joseph E. Worcester's lexicographical predicament. As a strong rival and equivalent to *OED*, Webster's *American Dictionary of the English Language (ADEL, 1828)* is the most representative and notable one of his dictionaries, and its publication opened up a brand-new era for the compilation of American English dictionaries. Recognized by the American public as the symbol of the independent national identity and the signifier of cultural and linguistic divergence of the New World from its source language and culture, *ADEL* not only reflects Webster's linguistic and political views but also redefines, to a certain extent, the American social, cultural, and historical significance (p. 60). Worcester's most significant, influential, and widely recognized work is *A Dictionary of the English Language* (1860), and it sets the precedent of both using pictorial illustra-

tions to reinforce definitions and systematically treating synonyms in American English dictionaries. This chapter also discusses the compiling philosophy in other major American dictionaries of the 19th century. In addition, it highlights the link between British lexicographical tradition and the development of different types of American dictionaries in the 20th century, including the inheritance of British lexicographical tradition in the prosperity of American general dictionaries, the continuation of British pedagogical traditions and the making of American learner's dictionaries, the tolerant attitude toward regional variants and the making of American pronouncing and dialect dictionaries, the tradition of British usage studies and the making of American usage and slang dictionaries, and the tradition of British semantic and etymological studies and the making of American thesauruses and etymological dictionaries. This chapter further surveys the making of American encyclopedic, electronic, and online dictionaries in the 20th Century, and other types of dictionaries and the dictionary of world Englishes.

Chapter Three investigates British lexicographical tradition and the development of Australian English lexicography. It explores the extension of British lexicographical tradition in the sprouting of Australian English lexicography, analyzes the diversity of the colonial sociocultural setting and the early development of Australian English lexicography, discusses the extension of British slang studies and the making of Australian dictionaries of slang and colloquialisms, examines the strengthened sociocultural consciousness and *The Macquarie Dictionary*, and surveys Australia's Oxford path of lexicographical development and the making of *The Australian National Dictionary*.

Chapter Four focuses on British and American lexicographical tradition and the making of Canadian English dictionaries. It introduces Canadian English and British lexicographical legacies in the making of dictionaries of Canadianisms, explores the British and American lexicographical heritage in the making of Canadian national dictionaries, and discusses the Oxford collaboration and the making of general dictionaries of Canadian English.

Chapter Five illustrates British lexicographical tradition in the shaping of New Zealand English lexicography. It surveys the pervasion of Māori into New Zealand English and the making of first English and Māori bilingual dictionaries, traces British tradition in nonstandard English studies and their extension into the making of nonstandard English dictionaries in New Zealand, explores the pivotal role of Māori in the development of New Zealand bilingual lexicography, and analyzes the newly emerging factors and Oxford lexicographical involvement in the prosperity of New Zealand lexicography.

Chapter Six examines the extending influence of British lexicographical tradition and the shaping of South African English lexicography. It begins with an explanation of the background for the sprouting of South African English lexicography and the establishment of the Dictionary Unit for South African English (DSAE). Then, it introduces South African English as a new and developing variety and the inception of South African monolingual lexicography. It

also discusses the extending influence of British lexicographical tradition and the making of South African national dictionaries, analyzes the English domination and the development of South African bilingual lexicography, and surveys Pharos dictionaries and other types of dictionary making.

The final chapter depicts the prospects for the development of world English lexicography with respect to dictionary form and function, information presentation and organization, incorporation and interaction in terms of potential areas of lexicographical research, dictionary users, integration of digital, technological and corpus technologies with lexicography and creation of new dictionaries, internationalization and regionalization of English dictionaries, and user segmentation and dictionary serialization.

The volume under inspection is distinguished from other works on similar topics in at least three aspects. Firstly, different from other existing studies on English lexicography which mainly focus on the lexicography of British English or American English, this volume may be the first one to provide a panorama of how lexicography of major native-speaker varieties of English (in addition to British English) has evolved from their British roots to current shapes and prosperity. Secondly, unlike other works which are confined to the dictionary-ontology paradigm, this volume explores the evolution of world English lexicography from both sociolinguistic and lexicographic perspectives. It puts English dictionary-making and research against a broader setting of socio-cultural observations in the world, and traces their evolutionary links with and inheritance from British (occasionally American) lexicographical tradition, their interrelation to the socio-cultural backgrounds, as well as their reformation and divergences through innovation and self-expansion. Thirdly, this volume reveals the lexicographical theories that guide dictionary-making of each English variety, and provides a comprehensive, coherent, and unified theoretical framework of English dictionary paradigms from its archetype to the prescriptive, the historical, the descriptive and finally the cognitive model.

Admittedly, in addition to the merits mentioned above, there is still some room for improvement in this volume. Firstly, this book features in a complete picture of the development of world English lexicography. However, the introduction to the whole history of the lexicography of some English varieties seems a little rough and general. Secondly, as a monograph on the history of English lexicography, this volume could have elaborated on the design features of major dictionaries in different English varieties. A comprehensive diachronic and synchronic investigation would reveal the distinctive characteristics of different lexicographical works. Finally, a glimpse of the major references indicates that there are some important omissions. Although Yong has quoted many dictionaries in the main body of the text, no dictionary is listed in the References. Moreover, there are a few minor mistakes and inconsistencies in the References. Specifically, as far as *The Oxford Guide to Practical Lexicography* is concerned, this volume mistakenly takes Rundell as the first author and Atkins as the second (p. 185). Furthermore, some authors are given the full name,

while some are listed in the format of family name plus the initial letter of the given name.

In conclusion, in spite of the minor criticisms raised above, this monograph deserves recognition as the first volume on a comprehensive history of world English lexicography from both sociolinguistic and lexicographic perspectives, and as a rich resource to English lexicographers and linguists. It presents readers with a clear, coherent, and comprehensive view of how lexicography of core English varieties has developed from its British origin, and how it interrelates to the socio-cultural backgrounds in the world.

Acknowledgments

This study was supported by a grant from the Philosophy and Social Science Foundation of Guangdong Province of China (Grant No. GD21CWY04).

References

A. Dictionaries

- Cowie, A.P.** 1989. *Oxford Advanced Learner's Dictionary of Current English (OALD4)*. Oxford: Oxford University Press.
- Delbridge, A.** 1981. *The Macquarie Dictionary*. McMahons Point: The Macquarie Library.
- Gove, P.B.** 1961. *Webster's Third New International Dictionary*. Springfield, MA: G. & C. Merriam.
- Hornby, A.S., A.P. Cowie and J. Windsor Lewis.** 1974. *Oxford Advanced Learner's Dictionary of Current English (ALD3)*. London: Oxford University Press.
- Hornby, A.S., E.V. Gatenby and H. Wakefield.** 1942. *Idiomatic and Syntactic English Dictionary (ALD1)*. Tokyo: Kaitakusha.
- Hornby, A.S., E.V. Gatenby and H. Wakefield.** 1963. *The Advanced Learner's Dictionary of Current English (ALD2)*. London: Oxford University Press.
- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton, T. & T. Longman et al.
- Murray, J.A.H., H. Bradley, W.A. Craigie and C.T. Onions.** 1884–1928. *The Oxford English Dictionary (OED)*. Oxford: Clarendon Press.
- Palmer, H.E.** 1938. *A Grammar of English Words*. London: Longmans, Green & Co.
- Procter, P.** 1978. *Longman Dictionary of Contemporary English (LDCE1)*. Harlow: Longman.
- Sinclair, J., P. Hanks, G. Fox, R. Moon and P. Stock.** 1987. *Collins Cobuild English Language Dictionary (COBUILD1)*. London/Glasgow: HarperCollins.
- Ramson, W.S.** 1988. *The Australian National Dictionary: A Dictionary of Australianisms on Historical Principles*. Melbourne: Oxford University Press.
- Summers, D.** 1987. *Longman Dictionary of Contemporary English (LDCE2)*. Harlow: Longman.
- Webster, N.** 1828. *An American Dictionary of the English Language (ADEL)*. New York: S. Converse.
- West, M.P. and J.G. Endicott.** 1935. *New Method English Dictionary*. London: Longmans, Green & Co.
- Worcester, J.E.** 1860. *A Dictionary of the English Language*. Boston: Hickling, Swan and Brewer.

B. Other literature

- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. New York/Oxford: Oxford University Press.
- Béjoint, H.** 2010. *The Lexicography of English: From Origins to Present*. Oxford: Oxford University Press.
- Collison, R.L.** 1982. *A History of Foreign-language Dictionaries*. London: André Deutsch.
- Considine, J. (Ed.).** 2019. *The Cambridge World History of Lexicography*. Cambridge: Cambridge University Press.
- Cowie, A.P.** 2002. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press.
- Cowie, A.P. (Ed.).** 2009. *The Oxford History of English Lexicography*. Oxford: Oxford University Press.
- Dixon, R.M.W.** 2018. *The Unmasking of English Dictionaries*. Cambridge: Cambridge University Press.
- Landau, S.I.** 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Miyoshi, K.** 2017. *The First Century of English Monolingual Lexicography*. Cambridge: Cambridge Scholars Publishing.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Yong, H. and J. Peng.** 2021. *A Sociolinguistic History of British English Lexicography*. London/New York: Routledge.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

Yongfang Feng
School of English Education
Guangdong University of Foreign Studies
Guangzhou
China
(fengyf@gdufs.edu.cn)

Publikasieaankondigings / Publication Announcements

Howard Jackson (Ed.). *The Bloomsbury Handbook of Lexicography*. Second edition. 2022, 473 pp. ISBN: 978-1-3501-8170-0 (Hardback), 978-1-3501-8171-7 (ePDF), 978-1-3501-8172-4 (eBook). London: Bloomsbury. Price £130.00 (Hardback), £117.00 (eBook/ePDF). (Review in this issue.)

Mariusz Piotr Kamiński. *Defining with Simple Vocabulary in English Dictionaries*. 2021, xvi + 326 pp. ISBN: 978-9-0272-0859-0 (Hardback), 978-9-0272-6000-0 (e-Book), ISSN 1388-8455. Terminology and Lexicography Research and Practice 22. Amsterdam/Philadelphia: John Benjamins. Price € 99.00 (Hardback and e-Book). (Review in this issue.)

Anton Prinsloo, Leon de Stadler en Amanda de Stadler. *Spreekwoorde en ander segswoyses*. Derde hersiene uitgawe. 2022, x + 597 pp. ISBN: 9781868902194 (sagteband). Kaapstad: Pharos Woordeboeke. Prys: R360.

Clifford Simango. *Concise Chindau Dictionary*. 2019, xii + 263 pp. ISBN: 978-1-77929-682-5. Harare: Dream Discovery. Price US\$9. (Review in this issue.)

Elsabé Taljard. *Lexikos 32(2) — Huldeblyk aan / Tribute to Danie Prinsloo*. 2022, xix + 243 pp. ISBN: 978-1-990998-49-2, ISSN: 2224-0039. AFRILEX-REEKS 32A:2022. Stellenbosch: Buro van die WAT.

Heming Yong. *English Lexicography from British Tradition to World Englishes*. 2022, viii + 194 pages. ISBN 978-1-4331-8696-7 (Hardback), ISBN 978-1-4331-8697-4 (eBook PDF), ISBN 978-1-4331-9258-6 (ePub). New York/Bern/Berlin/Brussels/Oxford/Vienna: Peter Lang. Price: US\$ 97.80 (Hardback), US\$ 94.95 (eBook). (Review in this issue.)

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://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 een 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 evalueerend 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, Leksikofokus, 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 STELLENBOSCH, Republiek van Suid-Afrika.

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat versoenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie 8 000 woorde oorskry nie.

Elke artikel moet voorsien wees van 'n opsomming van ongeveer 200 woorde en ongeveer 10 sleutelwoorde in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutelwoorde in Engels. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigeen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n vertaling van die oorspronklike titel bevat.

2. Grafika

Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting. Buiten in spesiale gevalle moet verwysings na *Lexikos*-artikels tot twee of drie per artikel beperk word. Uitsonderings moet met die redakteur van *Lexikos* uitgeklaar word. Dit word gedoen om die status van *Lexikos* in verskeie internasionale indekse te behou.

4. Aantekeninge/voetnote/eindnote

Aantekeninge moet deurlopend in die vorm van boskrifte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website: <http://lexikos.journals.ac.za/>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

- (1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected in an original, synoptic, interpretative, comparative or critically evaluative manner.
- (2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources. Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.
- (3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
- (4) **Projects:** Discussions of lexicographical projects.
- (5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
- (6) **Lexicovaria:** Any of a large variety of articles containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
- (7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index (ISI)*.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free, however, to use their

material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

6. Free offprints and copies

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. Self-references to *Lexikos* should be limited to two or three per article, except in exceptional circumstances. Exceptions should be cleared with the editor of *Lexikos*. This is done to preserve the status of *Lexikos* in various international indices.

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**.