

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

Lexikos 33

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

Lexikos 33

Redakteur

Editor

André H. du Plessis

Resensieredakteur

Review Editor

Dané Claassen



African Association for Lexicography

AFRILEX-REEKS 33:2023

AFRILEX SERIES 33:2023



BURO VAN DIE WAT

STELLENBOSCH

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2023 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2023

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

ISBN 978-1-990998-51-5
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	x
Foreword	xii
<i>André H. du Plessis</i>	
'n Woord van AFRILEX	xiv
A Few Words from AFRILEX	xvi
<i>Langa Khumalo</i>	
Redaksionele doelstellings	xviii
Editorial Objectives	xvix

Artikels / Articles

Verifying the Terminology of Arabic Rhetorical Tradition: The Case of <i>'Ilṭifāt</i>	448
<i>Emad Abdul-Latif</i>	
A Further Look into the Use of a Dictionary APP in EFL Writing: A Replication Study	324
<i>Yuzhen Chen and Suping Liu</i>	
The Future of Metalexigraphy: Reaching for the Mesosphere	136
<i>Gilles-Maurice de Schryver</i>	
Frequency or Keyness	184
<i>Zorica Đurović</i>	
Lexical Analysis of Nautical and Marine Engineering Corpora: Similar or Different Lexicographic Results	255
<i>Zorica Đurović and Milena Dževerdanović-Pejović</i>	
The Treatment of Argument Structure Constructions in Online English Learners' Dictionaries: The Case of <i>V N ADJ</i> ASCs	207
<i>Yongfang Feng and Hai Xu</i>	

The Concept of Lexicographic Condensation: A Review of and Perspectives on Digital Lexicography <i>M.^a Teresa Fuentes Morán</i>	404
Nuwe raamstruktuurtypes in aanlyn woordeboeke <i>Rufus H. Gouws</i>	382
Synonymy from a Prototype Theory Perspective and its Symbiosis with Polysemy: Towards a New Dictionary of Synonyms <i>Ana Halas Popović</i>	1
The Operative Function in Spanish Lexicography Exemplified through Sport Dictionaries and Other Reference Works <i>Ángel Huete-García</i>	277
Mit Wortschatz und lexikografischen Ressourcen handeln: kritische Überlegungen zur Anwendung lexikalischer, lexikografischer und digitaler Kompetenzen im virtuellen Raum beim akademischen Schreiben im DaF-Bereich <i>Meike Meliss</i>	227
<i>A New English–Serbian Dictionary of Sports Terms in the Light of Contemporary Challenges</i> <i>Mira Milić, Olga Panić Kavgić and Aleksandra Kardoš Stojanović</i>	47
Phonostilistische und regionale Variation der deutschen Aussprache in den orthoepischen Nachschlagewerken der dritten Generation <i>Krzysztof Nycz und Zygmunt Tęcza</i>	166
The Language of Ethnic Conflict in English Online Lexicography: Ethnophaulisms in "powered by Oxford" Lexico.com <i>Silvia Pettini</i>	299
Developing Dictionary Skills through Monolingual and Bilingual English Dictionaries at Tertiary-level Education in Hungary <i>Katalin P. Márkus and Ida Dringó-Horváth</i>	350
The Names of the Balkan Peoples and the Names of the Inhabitants of Balkan Countries in Lexicography (on the Example of the <i>Dictionary of Montenegrin National and Literary Language</i>) <i>Sanja Šubarić and Jovana Đurčević</i>	68

- A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and Implications for the Future 425
Sven Tarp and Rufus H. Gouws

Projekte / Projects

- KIU (Kiswahili–Italiano–UniOr): The UniOr online Dictionary for Italian L1 Swahili Learners 90
Flavia Aiello and Rosanna Tramutoli
- African Englishes in the *Oxford English Dictionary* 24
Danica Salazar

Leksikonotas / Lexiconotes

- Phonological Inclusion, and Exclusion, Regarding South African English in the Online OED 318
Alex Baratta

Leksiko-opname / Lexicosurvey

- Evaluating the Usefulness of the Learning Tools in Monolingual Online Dictionaries for Learners of English: Gauging the Preferences of Polish Students of English 109
Bartosz Ptasznik

Resensies / Reviews

- Bonny Sands en Kerry Jones (Hoofreds.). *N|uuki Namagowab Afrikaans English: ꞤXoakiꞤxanisi / Mîdi di ꞤKhanis / Woordeboek / Dictionary* 474
A.E. Cloete
- Heming Yong and Jing Peng. *A Sociolinguistic History of British English Lexicography* 41
Yanxiu Li
- John Considine. *Sixteenth-Century English Dictionaries* 125
Ai Zhong
- Annette Klosa-Kückelhaus and Ilan Kernerman (Eds.). *Lexicography of Coronavirus-related Neologisms* 131
Jian Zhou and Shuliang Dong

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

Publikasieaankondigings / Publication Announcements	479
Voorskrifte aan Skrywers	481
Instructions to Authors	482

Voorwoord

Ek is al sedert 2020 deel van die *Lexikos*-redaksie, maar vanjaar het ek die groot voorreg om vir die eerste keer in die hoedanigheid van hoofredakteur daarby betrokke te wees. Hoewel die (hoof)redakteur se taak uitdagend en tydrowend is, is dit 'n ongelooflike leerskool. Ek het nooit gedink dat ek in my relatief kort leksikografiese loopbaan in hierdie wonderlike posisie sou wees nie. Ek het opregte waardering vir die vertroue wat die AFRILEX-raad en my mederedakteurs in my geplaas het om vanjaar en volgende jaar aan die stuur van hierdie vername internasionale vaktydskrif te staan. Ek wil graag my kollegas op die AFRILEX-raad en *Lexikos*-redaksie bedank vir hulle onwrikbare en volgehoue ondersteuning. 'n Spesiale woord van dank aan proff. Elsabé Taljard en Dion Nkomo vir die uitstekende raad en bystand wat hulle gebied het, asook vir die tyd wat hulle bestee het en werk wat hulle gedoen het met van vanjaar se bydraes. Baie dankie ook aan dr. Steve Ndinga-Koumba-Binza vir sy onverdroete ywer met die Franse bydraes ten spyte van sy moeilike omstandighede vanjaar. Ek wil ook graag vir me. Tanja Hartevelt en me. Hermien van der Westhuizen van die WAT hartlik bedank vir hulle toewyding aan die stiptelike en professionele set en afronding van vanjaar se uitgawe, asook vir hulle bekwaamheid met die algehele produksie van die tydskrif. *Lexikos* is uiteraard 'n spanpoging en dit is werklik 'n plesier om deel van so 'n bedrewe span te wees.

Lexikos is nie 'n volledige produk sonder sy resensies nie. Na baie jare se getroue diens tree Tanja in 2023 uit die pos van resensieredakteur. Haar kundige en noukeurige hantering van resensies word opreg waardeur en *Lexikos* is innig dankbaar vir die stil, dog betroubare wyse waarop sy die resensies behartig het. In haar plek het die redaksie en AFRILEX-raad eenstemmig vir dr. Dané Claassen aangewys. Dané is 'n jong en ywerige leksikograaf wat tans by die Woordeboek van die Afrikaanse Taal werkzaam is, en sal nuwe energie aan die resensieredaksie verleen. Baie welkom, Dané.

Vanjaar se bydraes illustreer weereens die diversiteit van ons vakgebied, nie net ten opsigte van die onderwerpe wat bespreek is nie, maar ook ten opsigte van die tale waarin en waaroor daar gepubliseer is. Dit is merkwaardig om te sien hoe die leksikografie vanuit verskillende perspektiewe bestudeer word. Die veelsydigheid van die meta- sowel as praktiese leksikografie word ook hierdeur weerspieël. Sommige artikels werp lig op die uitdagings van terminologie-ontwikkeling en -ontginning, woordeboekdidaktiek, woordeboek- en metaleksikografiese navorsing, en die noodsaaklike aanpassing van die metaleksikografie. Ander wys weer hoe leksikografiese en linguistiese teorie en die leksikografiese praktyk sinvol bymekaarkom. Die alomteenwoordigheid van gevorderde tegnologieë en die impak daarvan op ons vakgebied is ook, soos verwag, 'n belangrike onderwerp in verskeie bydraes. Hierdie uitgawe bevat, buiten die

plaaslike bydraes, insiggewende bydraes van, onder andere, China, Katar, Pole, Montenegro, Italië en Engeland. Dit is wel van uiterste belang dat *Lexikos* ook die mondstuk van Afrika-leksikografie bly. Daarom rig ek, soos my voorgangers, 'n uitnodiging aan alle Afrika-leksikograwe om meer gereeld hul bydraes aan ons te stuur om sodoende te verseker dat plaaslike leksikografiese uitmuntendheid die internasionale blootstelling ontvang wat dit verdien.

2023 het ongelukkig ook hartseer nuus gebring. Ons het met groot leedwese verneem van die heengaan van Thierry Declerck wat tydens die eLex-kongres in Brno skielik oorlede is. Ons dink veral aan elkeen van ons kollegas en vriende wat nou saam met hom gewerk het. Hoewel hy nie 'n aktiewe bydraer in *Lexikos* was nie, word sy navorsing en kundigheid, veral oor rekenaarlinguistiek en taaltegnologie, deur verskeie *Lexikos*-outeurs en AFRILEX-lede waardeer.

Geen geakkrediteerde vaktydskrif kan sonder die insae en kundigheid van portuurbeoordelaars bestaan nie. Die redaksie is elke keurder dank verskuldig vir die soms ondankbare taak van portuurbeoordeling. Ons ervare keurders verseker dat die hoë wetenskaplike standaard van hierdie tydskrif gehandhaaf word, en dat die kwaliteit van die bydraes ook telkens verhoog is. Hoewel dit nie gebruiklik is om keurders by die naam te noem nie, wil ek tog vir proff. Pedro Fuertes-Olivera, Herman Beyer, Rufus Gouws, drs. Sascha Wolfer en Hanelle Fourie-Blair uitsonder vir hulle buitengewone toewyding aan die keuring van en hulp met verskeie artikels. Die *Lexikos*-redaksie is innig dankbaar vir die onskatbare bydraes van al ons kundige portuurbeoordelaars.

Laastens, wat sal ons tydskrif sonder sy outeurs wees! Baie dankie aan die outeurs wat *Lexikos* as platform kies om hulle waardevolle kennis, kundigheid en navorsing te deel. Dit maak enige redakteur se taak makliker as die outeurs met 'n positiewe en geduldige gees op keurders en redakteurs se terugvoer reageer. Daarvoor, en vir die outeurs se goeie samewerking, is ek opreg dankbaar.

Die werk wat vorige redakteurs gedoen het, maak dat *Lexikos* steeds stewig gevestig is as een van die voorkeur- internasionale vaktydskrifte vir die publikasie van artikels oor alle leksikografiese sake. Mag hierdie uitgawe, soos voriges, die leksikografie ter wille wees en 'n beduidende rol speel in die internasionale leksikografiese diskoers.

André H. du Plessis
Redakteur

Foreword

I have been part of the *Lexikos* editorial team since 2020, but this year I have the great privilege of taking up the role of editor(-in-chief). Although the editor's role is a challenging and laborious one, it has provided me with an unbelievable learning experience. I would never have thought of being in this wonderful position in my relatively short lexicographic career. My most sincere appreciation goes out to the AFRILEX Board and my co-editors for entrusting me with this and next year's issue of this prominent international journal. I would like to thank my colleagues on the AFRILEX Board and the *Lexikos* editorial team for their unwavering and continual support. A special word of thanks to Profs. Elsabé Taljard and Dion Nkomo for their excellent mentorship and support, as well as for the time and effort they have put in with some of the submissions. I also cannot thank Dr Steve Ndinga-Koumba-Binza enough for his untiring efforts with the French submissions, despite his difficult circumstances this year. I would furthermore like to give my heartfelt thanks to Ms Tanja Hartevelde and Ms Hermien van der Westhuizen for their devotion in promptly and professionally typesetting and finalising this year's issue, as well as for their proficiency with the overall production of the journal. *Lexikos* is ultimately a team effort, and it is a pleasure to be part of such an adept one.

Lexikos is not a complete product without its reviews sections. In 2023, after many years of loyal service, Tanja steps down from her role as reviews editor. Her always expert and precise work on the reviews is greatly appreciated. *Lexikos* is thankful for the untroubled, yet reliable way she managed the reviews. In her place the editors and AFRILEX Board unanimously appointed Dr Dané Claassen. Dané is a young and diligent lexicographer and researcher currently employed at the Woordeboek van die Afrikaanse Taal, who will bring a new energy as reviews editor. A warm welcome to Dané.

This year's contributions once again typify the diversity of our field; not only in terms of the topics discussed, but also in terms of the languages published in and about. It is remarkable to see how lexicography is approached from different perspectives. This also displays the versatility of the metalexicography and lexicography in practice. Some articles shed light on the trials and tribulations of the development and extraction of terminology, dictionary didactics, dictionary or metalexicographic research, and the necessary adaptation of the metalexicography. Other articles show how lexicographic and linguistic theory meet lexicographic practice in meaningful ways. The ever-presence of advanced technologies is, as expected, also an important theme in different contributions. Aside from the local contributions, this issue includes thought-provoking contributions from, among others, China, Qatar, Poland, Montenegro, Italy and England. It is imperative however that *Lexikos* remains the mouth-

piece of African lexicography. Therefore, like my predecessors, I would like to extend an invitation to all lexicographers from Africa to send us contributions on a regular basis to ensure that local lexicographic excellence gets the international exposure it deserves.

Unfortunately, 2023 also brought with it some sad news. We learnt with heavy hearts about the sudden passing of Thierry Declerck during the eLex Conference in Brno. Our thoughts go out to all our colleagues and friends who worked closely with him. While he was not an active contributor to *Lexikos*, his research and expertise, especially on computational linguistics and language technology, are appreciated by many *Lexikos* authors and AFRILEX members.

No accredited journal can exist without the insights and expertise of its peer reviewers. The editors are deeply indebted to each adjudicator for the sometimes thankless task of peer reviewing. Our experienced adjudicators guarantee that the high scientific standard of this journal is upheld, and that the quality of the contributions are improved time and time again. Though it is not usual to mention reviewers by name, I would like to single out Profs. Pedro Fuertes-Olivera, Herman Beyer, Rufus Gouws, Drs Sascha Wolfer and Hanelle Fourie-Blair for their extraordinary dedication to reviewing and assisting with multiple articles. The *Lexikos* editors are very thankful for the priceless contributions from all our expert peer reviewers.

Lastly, what would our journal be without its authors! A big thank-you to the authors who chose *Lexikos* as a platform to share their valuable insights, expertise and research. The task of any editor becomes far easier when one works with authors who engage with reviewers and the editorial team in such a positive and patient way. I am truly thankful for the collaborative spirit of our authors.

The work done by previous editors has ensured that *Lexikos* is still firmly established as one of the international journals of choice for publications of articles on all matters lexicographical. As with prior issues, may this issue benefit lexicography and play a key role in the international lexicographic discourse.

André H. du Plessis
Editor

'n Woord van AFRILEX

Die leksikografiese verhouding tussen *Lexikos* in die AFRILEX-reeks en die African Association for Lexicography (AFRILEX) leef steeds voort aangesien albei op hul unieke wyses toegewyd is aan die teorie, praktyk, beginsels, en gebruik van woordeboeke in Afrika en globaal. Die onlangse besluit van die AFRILEX-raad om voort te gaan om hierdie verhouding deur 'n aanpasbare volhoubaarheidsmodel in stand te hou, is dus strategiese versierendheid.

By hul 26ste jaarlikse internasionale konferensie verlede jaar het AFRILEX met 'n spesiale uitgawe van *Lexikos* 32(2), 'n huldigingsbundel ter ere van prof. Danie Prinsloo, 'n gepaste eerbetoon gebring aan een van die doyens van die teoretiese leksikografie. Die Buro van die WAT het dit opgevolg met nog 'n spesiale uitgawe, *Lexikos* 32(3), as huldeblyk aan dr. Willem F. Botha. Hierdie huldigingsbundel is 'n gepaste erkenning van dr. Botha se uitsonderlike bydrae as Hoofredakteur en Besturende Direkteur van die *Woordeboek van die Afrikaanse Taal* (WAT). Beide prof. Prinsloo en dr. Botha het die leksikografiese gemeenskap in Suid-Afrika en globaal met soveel toewyding en groot onderskeiding gedien.

Die 27ste jaarlikse internasionale AFRILEX-konferensie is in vennootskap met die Departement Suid-Afrikaanse Gebaretaal en Dowestudies aan die Universiteit van die Vrystaat (UV), in Bloemfontein, Suid-Afrika, georganiseer, en het saam met die 7de Internasionale Simposium oor Plekname (ISPN) plaasgevind. Daar was baie innemende referate, insluitende die stimulerende voorleggings deur die twee hoofsprekers, prof. Myriam Vermeerbergen van die Katholieke Universiteit Leuven, in Antwerpen, België en prof. Thapelo Otlogetswe van die Universiteit van Botswana, in Gabarone, Botswana. Dit is opmerklik dat sommige van hierdie interessante referate omskep is in wetenskaplike artikels wat in hierdie *Lexikos*-nommer gepubliseer is. Die *Lexikos*-joernaal ontvang steeds 'n groot aantal bydraes van internasionale outeurs, terwyl die getal plaaslike outeurs kommerwekkend afgeneem het, soos duidelik blyk uit hierdie nommer. Die *Lexikos*-redakteurspan het nog 'n uitgawe van *Lexikos* suksesvol voltooi om sodoende hierdie belangrike kanaal van akademiese diens in stand te hou. Namens die Raad, wil ek veral die redakteur, mnr. André H. du Plessis, bedank vir sy konsensieuse en onberispelike stiptheid en sorgvuldigheid met die huidige nommer, wat kenmerkend van die joernaal geword het. Mnr. Du Plessis is kundig bygestaan deur die vorige en huidige resensieredakteurs, me. Tanja Harteveld en dr. Dané Claassen, sowel as me. Hermien van der Westhuizen met tegniese ondersteuning.

Ek wil die belangrike rol wat die Buro van die WAT oor die jare as uitgewer van *Lexikos* vertolk het, herhaal. Die Buro van die WAT bly steeds 'n sleutel- en strategiese vennoot van AFRILEX. Namens die AFRILEX-raad, en inderdaad ook die algemene lede van AFRILEX, bedank ek die redaksiespan, die Buro van die

WAT, en die bydraende outeurs van harte vir nommer 33 van *Lexikos*. Dit is nog 'n belangrike bydrae tot die wetenskap van die leksikografie.

Langa Khumalo
President: AFRILEX

A Few Words from AFRILEX

The lexicographic liaison between *Lexikos* in the AFRILEX series and the African Association for Lexicography (AFRILEX) continues to exist as both are dedicated in their unique ways to the theory, practice, principles, and usage of dictionaries in Africa and globally. The recent decision by the AFRILEX Board to continue to preserve this liaison through an agile sustainability model is thus a strategic foresight.

At its 26th annual international conference last year, AFRILEX bestowed a befitting honour to one of the doyens of theoretical lexicography through a special issue of *Lexikos* 32(2), a festschrift in honour of Prof. Danie Prinsloo. The Bureau of the WAT followed this up with another special issue, *Lexikos* 32(3) as a tribute to Dr Willem F. Botha. This festschrift is a fitting recognition of Dr Botha's exceptional contribution as the Editor-in-Chief and Managing Director of the *Woordeboek van die Afrikaanse Taal* (WAT). Both Prof. Prinsloo and Dr Botha served the lexicographic community in South Africa and globally with so much dedication and great distinction.

The 27th AFRILEX annual international conference was organized in partnership with the Department of South African Sign Language and Deaf Studies at the University of the Free State (UFS), in Bloemfontein, South Africa, and was co-located with the 7th International Symposium on Place Names (ISPN). It had very engaging papers, including the stimulating presentations by the two keynote speakers Prof. Myriam Vermeerbergen from the Katholieke Universiteit (KU) Leuven, in Antwerp, Belgium and Prof. Thapelo Otlogetswe from the University of Botswana, in Gaborone, Botswana. It is notable that some of these engaging presentations were converted into scientific articles that are published in this *Lexikos* volume. The *Lexikos* journal continues to attract a large number of contributions from the international authors, while worryingly the number of local authors is dwindling, as is shown in this volume. The *Lexikos* editorial team has successfully concluded another issue of the *Lexikos*, thereby maintaining this important channel of academic service. On behalf of the Board, I want to especially thank the editor, Mr André H. du Plessis, for his conscientious and impeccable stringency and care with the current volume, which has become a hallmark of the journal. Mr Du Plessis was ably assisted by the review editors, Ms Tanja Harteveld and Dr Dané Claassen, as well as Ms Hermien van der Westhuizen, with technical support.

I want to reiterate 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 33 of *Lexikos*. It is another important contribution to the scholarship of lexicography.

Langa Khumalo
President: AFRILEX

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

- (1) om 'n kommunikasiekanaal vir die nasionale en internasionale leksikografiese gesprek te skep, en in die besonder die leksikografie in Afrika met sy ryk taleverskeidenheid te dien;
- (2) om die gesprek tussen leksikograwe onderling en tussen leksikograwe en taalkundiges te stimuleer;
- (3) om kontak met plaaslike en buitelandse leksikografiese projekte te bewerkstellig en te bevorder;
- (4) om die 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 vakspesialiste van hoë akademiese aansien voor te lê vir anonieme keuring.

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

Synonymy from a Prototype Theory Perspective and its Symbiosis with Polysemy: Towards a New Dictionary of Synonyms

Ana Halas Popović, *Department of English Studies,
Faculty of Philosophy, University of Novi Sad, Novi Sad,
Serbia (ana.halas@ff.uns.ac.rs)*

Abstract: This paper presents an attempt at providing a model for comprehensive, precise and systematic presentation of a word's synonymy within a dictionary entry, especially in the case of highly polysemous words. It is founded on the complementarity of polysemy and synonymy and a prototype-based view of the latter lexical relation. The proposed model is presented through four steps of tailoring an entry in a new dictionary of synonyms: the analysis of a word's polysemous structure, the compilation of a word's synonym sets viewed as prototype-based categories, the definition of synonyms within a set and the structure and organisation of an individual synonym set and the whole entry. Comprehensiveness of synonymy presentation according to the model is reflected in the fact that an entry lists synonym sets corresponding to different senses of the given headword as well as in the tendency for each set to offer a wide range of synonyms. Systematicity and precision in synonymy presentation are achieved primarily due to the application of prototype-based principles to the structuring and organizing of synonym sets and the defining of synonyms. The model also opens up the possibility of a transparent diagram-based visual representation of a word's synonymy in an electronic dictionary offering a more convenient visual organisation of large amounts of information.

Keywords: POLYSEMY, SYNONYMY, SYMBIOSIS, COMPLEMENTARITY, PROTOTYPE THEORY, MODEL, DICTIONARY ENTRY, SYNONYM SET, SENSE DEFINING, ENTRY STRUCTURE

Opsomming: Sinonimie vanuit 'n prototipeteorie-perspektief en die simbiose daarvan met polisemie: Op pad na 'n nuwe sinoniemwoordeboek. In hierdie artikel word 'n poging wat aangewend is om 'n model vir omvattende, eksakte en sistematiese voorstelling van 'n woord se sinonimie binne 'n woordeboekinskrywing daar te stel, veral by hoogs polisemiese woorde, bekend gestel. Dit is gebaseer op die komplementariteit van polisemie en sinonimie en 'n prototipegebaseerde siening van die laasgenoemde leksikale verwantskap. Die voorgestelde model word uiteengesit in vier stappe waarin 'n inskrywing in 'n nuwe sinoniemwoordeboek aangepas word: die analise van 'n woord se polisemiese struktuur, die samestelling van 'n woord se sinoniemstelle gesien as prototipegebaseerde kategorieë, die definiëring van

sinonieme binne 'n stel en die strukturing en organisering van 'n individuele sinoniemstel en die volledige inskrywing. Volgens die model word die omvattendheid van sinoniemaanbieding daarin gereflekteer dat 'n inskrywing sinoniemstelle lys wat ooreenstem met verskillende betekenisse van die gegewe lemma sowel as in die tendens van elke stel om 'n wye reeks sinonieme te vertoon. Konsekwenheid en akkuraatheid in sinoniemvoorstelling word hoofsaaklik verkry deur die toepassing van prototipegebaseerde beginsels op die strukturing en organisering van sinoniemstelle en die definiering van sinonieme. Die model ontsluit ook die moontlikheid van 'n deursigtige diagramgebaseerde visuele voorstelling van 'n woord se sinonimie in 'n elektroniese woordeboek wat 'n geriefliker visuele samestelling van groot hoeveelhede inligting aanbied.

Sleutelwoorde: POLISEMIE, SINONIMIE, SIMBIOSE, KOMPLEMENTARITEIT, PROTOTIPETEORIE, MODEL, WOORDEBOEKINSKRYWING, SINONIEMSTEL, BETEKENISDEFINIËRING, INSKRYWINGSTRUKTUUR

1. Introduction

This paper deals with a major lexicographic challenge: how to provide dictionary users with a comprehensive and, simultaneously, precise and systematic presentation of a word's synonymy within an entry, especially in the case of highly polysemous words. In regard to this, the aim of the paper is to propose a model for tailoring an entry in a dictionary of synonyms or thesaurus, as a possible solution to the aforementioned problem of an efficient synonymy presentation. The model, intended for both an electronic and a printed dictionary with polysemous headwords having more than one synonym in, at least, two of their senses, is illustrated in the paper through its application to English. Such a dictionary can serve as a vital resource to learners of a particular language, linguists and all other users interested in linguistic research of lexical synonymy and its functioning in the given language. The model is based on the complementarity of polysemy and synonymy and a prototype-based view of the latter lexical relation. This theoretical foundation should ensure greater efficiency in providing an account of a word's synonymy as a system of mutually related synonym sets clearly displaying their members' similarities and differences.

A word has different synonyms depending on the sense in which it is used. This clearly suggests that a lexicographer intending to provide a complete picture of a word's synonymy cannot disregard the sense-specific nature of this lexical phenomenon. Hence, the proposed model promotes the symbiosis of polysemy and synonymy and its contribution to a comprehensive lexicographic account of the latter lexical relation. Moreover, various studies have pointed out that prototype theory, as a cognitive-linguistic approach, is highly beneficial for lexicographic practice (Atkins and Rundell 2008, Geeraerts 2001, Halas 2013, 2016a, 2016b, Jehle 2004, Jiang and Chen 2017, McKeown 1991, Molina 2008, Ostermann 2015, Rundell 2012, Van der Meer 2000). It has been used in both semasiological and onomasiological studies of the lexicon. In this

paper, this theory is significant as the one that provides a deep insight into the internal organisation of a synonym set.

1.1 Existing lexicographic models of synonymy treatment

This subsection briefly sets forth existing models of lexicographic treatment of synonymy, which take into account the sense-specific nature of this lexical phenomenon. The overview is based on models identified in English lexicographic practice, whose products dominate the world-wide dictionary market and, more importantly, which is highly productive, diverse and innovative in terms of dictionaries of synonyms and thesauri.

There are two lexicographic models pointing out the sense-specific nature of synonymy. According to the first one, an entry contains several synonym sets, each related to a different sense of the given headword. However, such entries provide only bare lists of synonyms, lacking definitions or any explanation of their similarities or differences. Within each set, a meaning of the headword is illustrated with example sentences but it is not defined. An example of this model can be found in *Collins Thesaurus: The Ultimate Wordfinder* (second edition, 2002) (Figure 1) and *Oxford Thesaurus of English* (third edition, 2009) (Figure 2). As explained in the front matter of *Oxford Thesaurus of English*, the first synonym in a set is bolded and, thus, marked as the core one, which can be interpreted as a recognition of prototypicality, the notion that is significant for the model proposed in this paper, which will be elaborated on in this paper.

cut verb **1** = slit, saw, score, nick, slice, slash, pierce, hack, penetrate, notch ... *Thieves cut a hole in the fence ... You can hear the saw as it cuts through the bone ...*
2 = chop, split, divide, slice, segment, dissect, cleave, part ... *Cut the tomatoes into small pieces ...*
3 = carve, slice ... *Mr Long was cutting himself a piece of cake ...*
4 = sever, cut in two, sunder ... *I cut the rope with scissors ...*
5 = shape, carve, engrave, chisel, form, score, fashion, chip, sculpture, whittle, sculpt, inscribe, hew ... *Geometric motifs are cut into the stone walls ...*
6 = slash, nick, wound, lance, gash, lacerate, incise ... *I cut myself shaving ...*

Figure 1: The adaptation of a section of the entry for the verb *cut* in *Collins Thesaurus: The Ultimate Wordfinder*

cut ► **verb** 1 *the knife slipped and cut his finger* **GASH**, slash, lacerate, slit, pierce, penetrate, wound, injure; scratch, graze, nick, snick, notch, incise, score; lance.
— COMBINING FORM: -tomy.
2 *cut the red pepper into small pieces* **CHOP**, cut up, slice, dice, cube, mince; carve; divide; *N. Amer.* hash.
3 *they cut the rope before he choked | he has cut his ties with the church* **SEVER**, cleave, cut in two; *poetic/literary* rend; *archaic* sunder; *rare* dissever.
4 *she's had her hair cut | cut back the new growth to about half its length* **TRIM**, snip, clip, crop, bob, barber, shear, shave; pare; prune, pollard, poll, lop, dock; mow.
5 *I went out into the garden to cut some flowers* **PICK**, pluck, gather; harvest, reap; *poetic/literary* garner, cull.
6 *she gazed at the lettering cut into the stonework* **CARVE**, engrave, incise, etch, score; chisel, whittle.

Figure 2: The adaptation of a section of the entry for the verb *cut* in *Oxford Thesaurus of English*

The same model is followed in WordNet, an electronic lexical semantic database originating in the 1980s (Miller et al. 2008), which most significantly demonstrates the cooperation of polysemy and synonymy, hence it receives special attention here. Its main feature is the organisation of lexical information according to word meanings instead of word forms (Miller et al. 2008: 329). Thus, it is widely considered as closer to an online thesaurus than a dictionary. Its organisation is based on a lexical matrix in which there is mapping between a word form and a word meaning, including instances in which a word form has several different meanings (the case of polysemy) or ones in which a meaning can be realized through several different forms (the case of synonymy). In such a mapping, polysemy and synonymy are complementary. WordNet assumes that its users are already familiar with a particular concept (meaning) and so only identifies this concept, rather than defining it. This is why meanings are represented by lists of word forms that can express them, allowing users to distinguish between different meanings. In this database, different senses of a word are represented through sets of synonyms. Hence, WordNet can be regarded in essence as a representation of polysemy which relies on synonymy. In addition to synonym sets, which are structured taxonomically so that semantic relations among them are shown, the identification of a particular meaning is also enabled by short glosses and illustrative examples. However, no meaning has a fully-developed dictionary definition. The aforementioned features of WordNet are illustrated with Figure 3:

- S: (v) **cut** (separate with or as if with an instrument) "*Cut the rope*"
- S: (v) reduce, cut down, cut back, trim, trim down, trim back, **cut**, bring down (cut down on; make a reduction in) "*reduce your daily fat intake*"; "*The employer wants to cut back health benefits*"
- S: (v) swerve, sheer, curve, trend, veer, slue, slew, **cut** (turn sharply; change direction abruptly) "*The car cut to the left at the intersection*"; "*The motorbike veered to the right*"

Figure 3: The adaptation of a section of the entry for the verb *cut* in WordNet

The second model also recognizes the sense specific nature of synonymy offering several synonym sets within an entry which correspond to different senses of a polysemous headword. Still, an entry formed according to this model provides considerably more information on synonyms comprising, primarily, their definitions accompanied by example sentences while some other pieces of information, such as additional notes on a difference between synonyms quite close in meaning, etc. can also be included. A typical example of this model is *Oxford Learner's Thesaurus* (first published in 2008), as shown in Figure 4:

cut *verb*

- 1 cut taxes
- 2 cut the bread
- 3 have your hair cut
- 4 cut your finger

1 See also the entries for ABOLISH, REDUCE and SAVE 2

cut • slash • cut sth back • cut sth down • scale sth back • rationalize • downsize • scale sth down

These words all mean to reduce the amount or size of sth, especially of an amount of money or a business.

PATTERNS AND COLLOCATIONS

- ▶ to cut sth/ cut sth back/ cut sth down/ downsize sth/ scale sth down **from** \$50 000 **to** \$40 000
- ▶ to cut sth/ cut sth back/ cut sth down/ scale sth down **by** \$5 000/ 30%
- ▶ to cut back/ cut down **on** sth
- ▶ to cut/ slash/ cut back on **jobs**
- ▶ to cut/ slash/ downsize **the workforce**
- ▶ to cut/ slash/ rationalize **the cost** of sth
- ▶ to cut sth/ slash sth/ cut sth back/ cut sth down/ scale sth down **drastically**
- ▶ to cut sth/ cut sth back/ cut sth down **considerably**

cut [T] to reduce sth, especially an amount of money that is demanded, spent, earned, etc. or the size of a business:

The president has promised to cut taxes significantly. ◇ *Could you cut your essay from 5 000 to 3 000 words?* See also **cut** → REDUCTION noun

slash [T, often passive] (rather informal, journalism) to reduce sth by a large amount: *The workforce has been slashed by half.* ◇ *A slump in the retail trade has forced the company to slash prices.*

[...]

2 cut • chop • slice • carve • dice

These words all mean to make smaller pieces of sth by using sth sharp such as a knife.

PATTERNS AND COLLOCATIONS

- ▶ to cut/ chop/ slice/ carve sth **into** sth
- ▶ to cut/ chop/ slice sth **off** sth
- ▶ to cut/ slice sth **in half/ two**
- ▶ to cut/ chop/ slice sth **up**
- ▶ to cut/ chop/ slice/ carve/ dice sth **meat**
- ▶ to cut/ slice **bread/ cake**
- ▶ to chop/ slice an **onion**
- ▶ to cut/ chop/ slice/ dice sth **finely**
- ▶ to cut/ slice sth **thinly**

cut [T] to remove sth or a part of sth, or divide sth into two or more pieces with a knife, etc.; to make or form sth by removing material with a knife, etc.: *He cut four slices from the loaf.* ◇ *He cut the loaf into thick slices.* ◇ *Shall I cut you a piece of cake?* ◇ *Don't cut the string; untie the knot.* ◇ *The climbers cut steps in the ice.*

chop (-pp-) [T] to make smaller pieces of sth using sth sharp such as a knife: *He was chopping logs for firewood.* ◇ *Roughly chop the herbs.*

slice [T, I] to cut sth into slices; to cut sth easily with or as if with a sharp blade: *Slice the cucumber thinly.* ◇ *a sliced loaf* ◇ *a loaf of sliced bread* ◇ *He accidentally sliced through his finger.* See also **slice** → PIECE

carve [T, I] to cut a large piece of cooked meat into smaller pieces for eating: *She taught me how to carve a leg of lamb.* ◇ *Lunch is ready. Who's going to carve?*

dice [T] to cut meat, vegetables, etc. into small square pieces: *diced carrots/lamb*

3 cut • trim • shave • mow • lop • shear • snip • crop • clip

These words all mean to make sth shorter or neater by removing part of it with a sharp tool.

[...]

Figure 4: The adaptation of a section of the entry for the verb *cut* in *Oxford Learner's Thesaurus*

However, it should be noted that this thesaurus takes only certain possible senses of a polysemous word into consideration. In addition to this, it applies greater selectivity in comparison with the aforementioned thesauri in determining synonyms for inclusion in a set to meet the needs of its target group (Lea 2008: 546).

In summary, both models recognize the sense-specific nature of synonymy. However, the first one typically provides a longer list of synonyms within a set but scarce information on them (only example sentences as indicators of meaning and register/regional labels), while the second one applies greater selectivity in compiling lists of synonyms, it also provides more detailed information on synonymy within entries. It should be noted, though, that these are print dictionaries where one of the main reasons for the constraints in their models is that of space, which makes it really difficult to provide an exhaustive list of synonyms and clearly and neatly present complete information on each synonym. This constraint is, in some cases, addressed by online thesauri which are linked to dictionaries, which makes it possible to provide a list of synonyms and link them to the relevant senses of their dictionary entries (offering definitions, grammatical information, example sentences, etc.). Still, the question that remains is whether it is possible to merge the two models within a single dictionary, i.e. a dictionary of synonyms, and provide users with an inventory of a word's synonyms which is exhaustive in terms of the number of synonyms included as well as the information on each of them presented in a systematic, neat and clear way.

2. Synonymy from a prototype theory perspective

A general definition of synonymy frequently referenced in linguistic literature is that formulated by Cruse (2002: 486), in which it is understood as the 'similarity or identity of meaning between senses associated with two (or more) different lexical forms.' Šarić (2011: 305) points out that a cognitive linguistic theoretical framework quite suitably answers various questions related to synonymy. A number of existing cognitive linguistic interpretations of synonymy also rely on the principles of prototype theory.

The key term in a prototype-based examination of synonymy is a synonym set (e.g. *big*, *huge*, *enormous*), dominated by a neutral member or a semantic dominant (Dragičević 2007: 258), the prototype, more general than the other members of the given set. It functions as the basis for the other members' sense definitions (e.g. *huge* as 'extremely big') acting as the semantic base of the given set, representing the central semantic component common to all its members. For example, in the synonym set *cut*, *chop*, *slice*, *carve*, *dice*, the prototype is the verb *cut* since its meaning is the most neutral and general ('to divide something into two or more pieces using a sharp tool'). It represents the semantic base for the other members that express more specific meanings. These distinct meanings are, in the given example, determined by a specific referent or a specific

manner of cutting (e.g. *carve* 'divide a large piece of cooked meat into smaller pieces', *slice* 'divide something into slices', etc.).

When applied to a study of synonymy, the main postulates of prototype theory, summarized by Geeraerts (1989) as the four prototypicality effects, provide deeper insight into the internal organisation of a synonym set. The first effect is related to degrees of typicality. The position of synonyms in a set depends on their similarity to the prototype. The distance between a particular synonym and the prototype is inversely proportional to their similarity. Hence, the cline from centre to periphery is the central organisational principle of a synonym set.

The second prototypicality effect refers to the flexible boundaries of prototypical categories. Certain peripheral members might be so loosely related to the prototype that they are treated as boundary cases that can belong to another set as well. For example, in the synonym set *cut*, *reduce*, *decrease*, *slash*, *lessen*, *diminish*, *lower*, *retrench*, the last member has a rather specific meaning ('(of a business, government, etc.) make costs smaller') in comparison with the prototype ('reduce the amount/quantity of something'). As the periphery of the set, it represents a boundary case that can be justifiably treated as a member of another set: *retrench*, *economize*, *cut back*, *budget*, *save*. Such examples indicate that synonym sets are prototypical categories whose edges are blurred rather than rigid or clear-cut. This also implies that synonym sets can be extended by new members at any point.

Following Geeraert's third prototypicality effect, each member of the synonym set needs to possess only some of the prototypical features. No single set of features is necessarily shared by all the synonyms in a set, as evident in this one: *cut*, *clip*, *trim*, *snip*, *shear*, *crop*, *barber*, *bob*, *pare*, *prune*, *pollard*, *mow*. The prototypical features of this set are contained in the definition of its dominant member *cut* 'make something shorter with a sharp tool.' In the definition of the peripheral member *mow* ('make grass shorter using a machine'), the target is precisely specified ('grass'), while 'with a sharp tool' replaced by 'using a machine'. *Mow* shares some but not all the features with the prototype. As a prototypical category, a synonym set acts as a cluster of such partially related descriptions.

The fourth effect entails that synonym sets exhibit family resemblance. Organised around the same prototype, synonyms in a set are mutually related, sharing some sense components and overlapping to a certain extent.

2.1 The complementarity of polysemy and synonymy

Cruse's definition of synonymy claims that it is a relation between senses of different words and not words themselves. From this perspective, an in-depth study of synonymy needs to rely on polysemy, defined by Evans (2005: 33) as a phenomenon in which 'a single linguistic form' has various but mutually related senses. A more precise and thorough account of synonymy between two lexemes requires an analysis of their polysemous structures and the identification of

their synonymous senses. Rasulić (2016: 129) observes that synonymy between two lexemes is primarily based on the sameness of their primary (prototypical) senses, while it is additionally reinforced by the sameness of their secondary senses. For instance, the verbs *cool* and *chill* are synonymous in their primary sense 'to make somebody or something become colder' but not in some secondary senses, such as *chill* 'to frighten somebody' or *cool* 'become calmer.' In contrast, the adjectives *close* and *near* are synonymous in their primary sense 'a short distance or time away' (e.g. *Her house is close to/near the hospital.*) and in their secondary sense 'near or close in family relationship' (e.g. *Even their close/near relatives didn't know about the engagement.*).

This clear synergy between polysemy and synonymy implies that a thorough and detailed examination of a single word's synonymy fundamentally requires an analysis of its polysemous structure, followed by a search for synonyms corresponding to different senses of the given word. Dirven and Verspoor (2004: 44) support this idea, claiming that a systematic combination of the semasiological and onomasiological approach contributes to 'a fully integrated conception of lexicology.'

3. The proposed model for tailoring an entry in a dictionary of synonyms

This paper presents a model as a possible solution to the aforementioned issue of comprehensive and, simultaneously, systematic and clear presentation of a polysemous word's synonymy. For the purpose of the illustration of its functioning, the model is, in this paper, applied to the case of the English verb *cut*.

The model starts from two fundamental theoretical principles:

- (1) the symbiosis of polysemy and synonymy,
- (2) a prototype-based view of synonymy.

The essential idea of the proposed model is that a lexicographer must first analyse the word's polysemous structure and compile an exhaustive inventory of its senses. This inventory will then serve as the structural and organisational basis for an entry, representing a system of synonym sets related to different senses of the given word. An important issue in the analysis of a word's semantic structure is sense differentiation, which might refer to either the splitting of identified senses into a larger number of more specific senses or the lumping of identified senses into a smaller number of more general senses. Which sense differentiation strategy is applied is of crucial significance for the creation of synonym sets since it directly influences their scope. Namely, depending on the applied sense differentiation strategy (lumping or splitting), a particular use of a word can be more or less specifically demarcated. The more specifically a particular sense is demarcated, the narrower the range of its synonyms is. For instance, one of the uses of the verb *cut* is 'reduce the amount or quantity of something,' which can be treated as a superordinate sense that

subsumes two more specific senses; 'reduce the amount of money, especially costs, prices, etc.' and 'reduce the number of people in an organisation.' In each of these three senses, the verb *cut* has a different synonym set: 'reduce the amount or quantity of something:' *cut, reduce, decrease, slash, lessen, diminish, lower*; 'reduce the amount of money, especially costs, prices, etc.:' *cut, retrench, mark down, discount*; 'reduce the number of people in an organisation:' *cut, slim down, downsize*. If the strategy of lumping is used, the two more specific senses will not be so finely differentiated but instead will be subsumed under the more general, superordinate sense. This will clearly affect the range of synonyms for this use of the verb *cut* since the grouping of these senses leads to the merging of three potentially narrower synonym sets into a single, wider one: *cut, reduce, decrease, slash, lessen, diminish, lower, retrench, mark down, discount, slim down, downsize*. The proposed model is based on the assumption that a dictionary of synonyms should provide more complex synonym sets addressing and clearly outlining the similarities and, more importantly, the differences between synonyms within an extensive range. Hence, the practice promoted in this paper employs the strategy of lumping.

The aforementioned analysis is a necessary prerequisite for the second step in the given process — compiling synonym sets for the given word.¹ The formation of a synonym set starts with the establishment of its prototype, relative to which all its other members are identified and organised. Each prototype holds the most neutral and general meaning in each set, central and common to all other set members. These members are selected as words that express the same or a more specific meaning in comparison with the prototype. They generally represent a specialisation of the basic, general, prototypical meaning, as in the following example of the set corresponding to one of the senses of the headword *cut*: *cut, slit, slash, lacerate, gash, nick, notch*. However, it might happen that the given headword is not the prototype in each of its synonym sets. For example, *cut* is not the prototype in the synonym set corresponding to its sense 'hurt someone's feelings'. The prototypical member carrying the meaning central to all other members is *hurt*, while *cut* is one of the peripheral members (being more restricted in its use since its typical subjects are words or actions, not people and it is typically used with an adverbial modifier, such as 'deeply'). Still, the given set is listed within the entry for *cut*, but without any information on the synonyms (e.g. definitions, illustrative examples, etc.). Instead, cross referencing is applied, as can be seen in the illustration of the model in Section 4. Namely, the sense is cross-referenced, linking the primary entry to the entry for *hurt* and the given synonym set to which *cut* belongs and where the full information on all members of this set is provided. Hence, within an entry, complete information on synonyms is provided only in the case of sets whose central member is the headword itself. Furthermore, the overlapping of synonym sets across different entries is avoided through such a practice since polysemous headwords appear in only one completely evolved synonym set in each of their senses. It should also be noted that certain entries

may contain only cross-references if their headword is not the prototype in any of its synonym sets. For example, the verb *slash* has two senses. The first, 'violently make a long opening or wound in something with a sharp tool,' places it as a member of the synonym set formed for sense 1 in the entry for *cut*, as shown in the illustration of the model in Section 4. The second, 'make something significantly smaller in quantity, amount, size, etc.' makes it a member of the synonym set corresponding to sense 3.a. in the entry for *cut* as well (see Section 4). The entry for *slash* would, thus, contain two cross references, as can be seen in the following illustration:

slash *verb* /BrE slæʃ, AmE slæʃ/ slashes, slashing, slashed, slashed

See the entry for **cut** *verb*

1. MAKE AN OPENING OR A WOUND

cut, slit, slash, lacerate, gash, nick, notch

See the entry for **cut** *verb*

3.a. REDUCE THE AMOUNT OF SOMETHING

cut, reduce, decrease, slash, lessen, diminish, lower, retrench, mark down, discount, slim down, downsize

However, if a headword is a member in the set whose prototype does not have more than one synonym in, at least, two of its senses, this set cannot be included into the given dictionary since it does not adhere to the principle of the dictionary's compilation stated in the introduction to the paper.

The third step refers to the defining of synonyms included into an entry. The definition of the prototype serves as the basis for formulating the definitions of its synonyms. As Pisárčiková and Benko (1996: 691) underline, the explanation of the prototypical meaning is to be general enough to encompass the essential meaning of all synonyms in a set. Furthermore, a prototype definition used as a common pattern for defining all synonyms in a set acts as a formal indication of their relatedness to the prototype and, simultaneously, their mutual connection. For example, a sense of the verb *cut*, for which a rather complex synonym set can be compiled, is defined as 'make an opening or a wound in something with a sharp tool or object.' Acting as a prototypical definition, it serves as the base for all other synonym definitions in the set. Their specific sense components are tacked on to the prototypical definition, so their specific variations of its meaning are demonstrated. In this way, within a set, differences in meaning between any particular synonym and the prototype, as well as between synonyms themselves, is made apparent. As Adamska-Salaciak (2013: 330) emphasizes, when choosing between alternative synonyms, it is not only their similarities but also their differences that play an important role. Thus, in the definition of the synonym *slit*, 'an opening or a wound' is more precisely and specifically determined by adding the adjectives 'long and narrow' to the basic, prototype pattern. The final formulation of this *slit* is 'make a long and narrow opening or

wound in something with a sharp tool or object.' Similarly, in the case of the synonym *slash*, the way in which an opening or a wound in something is made is specified with the adverb 'violently,' while the opening or wound itself is described as 'long.' This yields the final formulation; 'violently make a long opening or wound in something with a sharp tool or object.' In the definition of the synonym *lacerate*, the added information refers to the fact that there are typically a number of openings or wounds as well as the object in which they are made ('make a number of openings or wounds in skin or flesh with a sharp tool or object').

Definitions can also include stylistic and pragmatic information on synonyms since the contrast among them is often reflected in their specific collocational range or associative meaning. This can include expressive features (e.g. approving, disapproving) or stylistic features related to a specific dialect (e.g. British, American, etc.), sociolect (e.g. standard, non-standard, slang, etc.), register (e.g. formal, informal, spoken, written) or field (e.g. geometry, law, computing, etc.). These pieces of information can be provided in the form of labels before the very text of definitions. Typical collocations of a particular synonym that distinguish it from other synonyms in the given set are highlighted in example sentences following the definition. In addition, typical referents used with a particular synonym are provided in brackets before its definition (e.g. *retrench* '(of a business, government, etc.) make costs smaller').

The fourth step deals with the structure and organisation of each individual synonym set as well as each entry as a whole. An entry represents a system of a headword's synonym sets. Since each synonym set in this system is related to a particular sense of the headword, these sets are arranged so that their mutual relations correspond to appropriate senses in the polysemous structure of the headword. In other words, synonym sets are mapped according to, and following, the organisation of their corresponding senses in a particular polysemous structure. Therefore, sets corresponding to closely related senses are grouped within the same numbered section. Every set within a section is positioned within its specific field. The first set in a section is the one corresponding to the most general sense in the given group. This set regarded as the superordinate one is followed by sets corresponding to more specific senses, probably derived from the most general one through metaphorical extension, specialization or any other semantic mechanism. These subsets are numbered according to the model 1.a, 1.b, 1.c, 2.a, 2.b, etc. and ordered according to the descending scale of their corresponding senses' frequency of use in the contemporary language.²

The internal structure and organisation at the field level in an entry or an individual synonym set itself follow the principle 'from centre to periphery.' The prototype is followed by other members ordered according to the declining scale of their similarity to it. For example, in the synonym set of the headword *cut* corresponding to the sense 'reduce the amount or quantity of something,' members positioned immediately after the prototype are *reduce* and *decrease*,

since they show relatively close similarity to it. These two synonyms are followed by *slash*, whose meaning is slightly more specific in comparison with the prototype, as the extent to which something is reduced is implied ('significantly'). *Slim down* and *downsize* are the most peripheral members of the set on account of the precise specification of the referent ('companies, organisations') and their respective, specific components of manner ('by reducing the number of jobs') and aim ('in order to reduce costs'). If two or more synonyms are equally similar to the prototype (for example, *reduce* and *decrease* in the given set), their placement in the set is also determined according to the declining scale of their frequency of use in the contemporary language. The gradient relatedness of synonyms to the prototype is also visually presented in the sample entry in Section 4 by the indentation of their definitions; the greater the indentation the weaker the synonym's similarity to the prototype. If two synonyms are equally similar to the prototype, their indentation is the same. Synonyms practically identical to the prototype are not indented.

4. The illustration of the proposed model

The application of the proposed model representing a system of synonym sets is illustrated by the entry for the verb *cut*:³

cut *verb* /BrE kʌt, AmE kʌt/ cuts, cutting, cut, cut

1. MAKE AN OPENING OR A WOUND

cut, slit, slash, lacerate, gash, nick, notch

cut make an opening or a wound in something with a sharp tool or object:

[cut something + adjective] *She tripped over a stone and cut her head open.*

slit /BrE slit, AmE slit/ make a long and narrow opening or wound in something with a sharp tool or object: [slit something + adjective] *He slit open the envelope and took out the money.*

slash /BrE slæʃ, AmE slæʃ/ violently make a long opening or wound in something with a sharp tool or object: *My neighbour was so angry that he slashed the tires on my car.*

lacerate /BrE 'læsəreɪt, AmE 'læsəreɪt/ FORMAL make a number of openings or wounds in skin or flesh with a sharp tool or object: *I have seen that her leg was badly lacerated.*

gash /BrE gæʃ, AmE gæʃ/ make a long and deep opening or wound in something, especially human skin, with a sharp tool or object: *He gashed his hand with a broken bottle.*

nick /BrE nɪk, AmE nɪk/ make a small wound in something with a sharp tool or object: *I nicked myself while shaving so I'll just put a bit of toilet paper on my face.*

notch /BrE nɒtʃ, AmE nɑːtʃ/ make a small V-shaped opening in an edge or a surface with a sharp tool or object: *He marked the place by notching one of the planks.*

1.a. HURT SOMEONE'S FEELINGS

hurt, wound, pain, cut, grieve, distress, upset, sting

See the entry for **hurt** *verb*

2. DIVIDE SOMETHING INTO PIECES

cut, slice, dice, cube, chop, carve

cut divide something into pieces with a sharp tool: [[cut something in/into something](#)] *She cut the cake into twelve pieces.*

slice /BrE slaɪs, AmE slaɪs/ divide something into thin pieces with a sharp tool: *Slice the tomatoes and put them on the pizza.*

dice /BrE daɪs, AmE daɪs/ divide something, especially food, into small cubes with a sharp tool: *Dice the ham and then mix it with the vegetables to finish the salad.*

cube /BrE kjuːb, AmE kjuːb/ divide food into small cubes with a sharp tool: *The bread should be cubed and dried in the oven.*

chop /BrE tʃɒp, AmE tʃɑːp/ divide something into pieces with repeated blows of a sharp tool, especially an axe or a knife: [[chop something into something](#)] *Please chop the carrots into equal pieces.*

carve /BrE kɑːv, AmE kɑːrv/ divide a larger piece of cooked meat into smaller pieces with a sharp tool for eating: *On this special day, my father always ceremoniously carves the turkey.*

2.a. DIVIDE SOMETHING INTO TWO PIECES

cut, cleave, sever

cut divide something into two pieces with a sharp object: [[cut something in/into something](#)] *You should cut the potatoes in half. After the accident, the car was cut in two by the train.*

cleave /BrE kliːv, AmE kliːv/ OLD-FASHIONED/ LITERARY divide something into two pieces with a sharp and heavy object: *He cleaved the log in half with an axe to light the fire.*

sever /BrE 'sevə(r), AmE 'sevər/ divide something into two pieces with a sharp object, especially suddenly and forcibly: *The bullet went through the window and severed the phone cord.*

2.b. DIVIDE A LINE IN TWO

cross, intersect, cut, bisect

See the entry for **cross** *verb*

3. MAKE SOMETHING SHORTER

cut, clip, trim, snip, shear, crop, barber, bob, pare, prune, pollard, mow

cut make something (grass, hair, etc.) shorter with a sharp tool: [**cut something+adjective**] *For this role, I had to cut my hair short. Your task is to cut the grass/lawn/hedge every week.*

clip /BrE klɪp, AmE klɪp/ make something shorter with scissors or shears: *He was clipping the hedge all morning.*

trim /BrE trɪm, AmE trɪm/ make something shorter and neater with a sharp tool: *If you want to be a good gardener, you have to trim hedges regularly.*

snip /BrE snɪp, AmE snɪp/ make something shorter with scissors using short quick strokes: [**snip at/through something**] *She grabbed a pair of scissors and started snipping at loose threads.*

shear /BrE ʃɪə(r), AmE ʃɪr/ FORMAL make someone's hair shorter: *The soldiers' hair was shorn before the ceremony.*

crop /BrE krɒp, AmE krɑ:p/ make someone's hair very short: [**crop something+adjective**] *She wanted a change so she cropped her long brown hair short.*

barber /BrE 'bɑ:bə(r), AmE 'bɑ:rbər/ make a man's hair shorter: *He wanted to neatly barber his hair before the party.*

bob /BrE bɒb, AmE bɑ:b/ make someone's hair shorter so that it is of the same length all the way around and sits above the shoulders: *She bobbed her hair so that everyone could see her new earrings.*

pare /BrE peə(r), AmE per/ BRITISH make something, especially one's nails, shorter so that it becomes smooth and neat: *When your nails are soft, it is easy to pare them.*

prune /BrE pru:n, AmE pru:n/ make a tree, shrub or bush shorter by removing its dead branches or stems with a sharp tool: *This month is the perfect time to prune roses.*

pollard /BrE 'pɒləd, 'pɒlə:d, AmE 'pɑ:lərd/ SPECIALIST make a tree shorter by removing branches at its top with a sharp tool: *Although the trees were pollarded, this avenue is still beautiful.*

mow /BrE məʊ, AmE moʊ/ make grass shorter using a machine: *He regularly mows the lawn/grass in front of his house.*

3.a. REDUCE THE AMOUNT OF SOMETHING

cut, reduce, decrease, slash, lessen, diminish, lower, retrench, mark down, discount, slim down, downsize

cut make something smaller in quantity, amount, etc.: *We can't buy the house if they don't cut the cost. The citizens asked from the local authority to cut spending/*

prices/taxes. [cut something by] *The company cut our salaries by 20%.* [cut something from something to something] *Is it possible for you to cut the article from 9000 to 7000 words?*

reduce /BrE rɪ'djuːs, AmE rɪ'duːs/ make something smaller in quantity, amount, size, etc.: [reduce something from something to something] *The company had to reduce the number of employees from 100 to 60.* [reduce something by something] *We managed to reduce our costs by 5%.*

decrease /BrE dɪ'kriːs, AmE dɪ'kriːs/ FORMAL make something smaller in quantity, amount, size, etc.: *I should decrease my sugar intake.*

slash /BrE slæʃ, AmE slæʃ/ INFORMAL make something significantly smaller in quantity, amount, size, etc.: *The company slashed the costs/prices by 50%.*

lessen /BrE 'lesn, AmE 'lesn/ make something smaller, weaker, etc.: *Such a tactic will help to lessen the influence of reforms.*

diminish /BrE dɪ'mɪnɪʃ, AmE dɪ'mɪnɪʃ/ make something smaller, weaker, etc.: *The captain's sickness will diminish our team's chances.*

lower /BrE 'ləʊə(r), AmE 'loʊər/ make something smaller in degree, value, quality, etc.: *The temperature should be gradually lowered.*

retrench /BrE rɪ'trentʃ, AmE rɪ'trentʃ/ FORMAL (of a business, government, etc.) make costs smaller: *The company will have to retrench.*

mark down /BrE mɑːk daʊn, AmE mɑːrk daʊn/ make a price of something smaller: *During the weekend, we will mark down all goods by 50%.*

discount /BrE dɪs'kaʊnt, AmE dɪs'kaʊnt, 'diskaʊnt/ make a price of something smaller: *This supermarket is discounting the price of sweets.*

slim down /BrE slɪm daʊn, AmE slɪm daʊn/ make a company or an organisation smaller by reducing the number of jobs: *The reforms require slimming down the workforce.*

downsize /BrE 'daʊnsaɪz, AmE 'daʊnsaɪz/ BUSINESS make the number of employees smaller in order to reduce costs: *At the meeting, they decided to downsize the cabinet.*

3.b. SHORTEN A TEXT, FILM, PERFORMANCE

cut, abridge, condense

cut make a text, film, performance shorter by removing some of its parts: *All the scenes that were not believable enough were cut.*

abridge /BrE ə'brɪdʒ, AmE ə'brɪdʒ/ make a text, film, performance shorter by removing some of its parts: *The first version of his speech was abridged.*

condense /BrE kən'dens, AmE kən'dens/ make a text shorter and concise: [condense something into something] *I think that the whole chapter can be condensed into a one-page text.*

4. MAKE SOMETHING WITH A SHARP TOOL

cut, carve, score, engrave, incise, etch, chisel, whittle

cut make something by using a sharp tool to remove material: [cut something in something] *They had to cut a hole in the floor to escape.*

carve /BrE kɑ:v, AmE kɑ:rv/ make something by using a sharp tool to remove material from wood or stone: [carve something from/out of something] *I carved the whole figure from a block of jade.* [carve something into/in something] *She carved a piece of stone into the shape of a heart.*

score /BrE skɔ:(r), AmE skɔ:r/ make a mark on a surface by using a sharp tool to remove material: *In order to recognise it later, score the plywood with a knife.*

engrave /BrE ɪn'ɡreɪv, AmE ɪn'ɡrev/ make a text or designs on the surface of a hard object by using a sharp tool to remove material (wood, stone, metal, etc.): [engrave something on something] *Her name was engraved on the ring.* [engrave something with something] *The plate was engraved with their symbol.*

incise /BrE ɪn'saɪz, AmE ɪn'saɪz/ FORMAL make a text or designs on the surface of a hard object by using a sharp tool to remove material (wood, stone, metal, etc.): [incise something in/on/onto something] *Human figures were incised in the stone.*

etch /BrE etʃ, AmE etʃ/ make a text or picture in a piece of glass, metal, etc. by using a sharp tool to cut lines: [etch something in/into/on something] *His initials were etched into the metal plate.* [etch something with something] *The glass is etched with her name.*

chisel /BrE 'tʃɪzl, AmE 'tʃɪzl/ make a shape in wood or stone by using a specific tool (a chisel) to remove material: [chisel something+ adverb/preposition] *She chiselled my name into the marble.*

whittle /BrE 'wɪtl, AmE 'wɪtl/ make a shape of a piece of wood, etc. by using a sharp tool to remove small pieces from it: [whittle something from something] *He whittled a female figure from a piece of wood.* [whittle something into something] *He whittled a piece of wood into a female figure.*

4.a. MAKE A SOUND RECORDING

record, cut, tape-record

See the entry for **record** *verb*

4.1 The proposed model in an electronic environment

Given the visual and processing efficiency of electronic dictionaries, there is an obvious need to examine the application of the proposed model in an electronic environment.

The main advantage offered by this format is its enabling of a more convenient visual organisation of large amounts of information. This is particularly

evident in such an entry as the one presented in the previous section of the paper. This entry can be displayed in the same visual layout in a print dictionary and an electronic dictionary. The only exception is cross referencing to another entry in the case of synonym sets in which the headword is not the prototype (e.g. See the entry for **hurt** *verb*). This note is omitted in an electronic dictionary as unnecessary since, even in the aforementioned case, the information on all synonyms can be accessed from the given headword's entry through hyperlinking, the additional benefit possessed by this format. Hyperlinking enables a visual representation that, in its basic form, offers a smaller amount of initial information, yet provides simple access to more detailed information. This initial presentation can be limited to only sense definitions of the given headword and synonym sets corresponding to them yet each synonym in these sets can be hyperlinked to more complete information. This can include its definition, example sentences, pronunciation, grammatical patterns in which it appears, and more. In this way, the basic appearance of an entry is not burdened with a large amount of detailed information, enabling a simpler and more efficient search experience. Hyperlinking is especially convenient for cross-referencing, a practice frequently used by the proposed model and already discussed in Section 4 of the paper.

Moreover, as the proposed model is based on the symbiosis of polysemy and synonymy according to the principles of prototype theory, an electronic environment allows for the visual representation of an entry in the form of a diagram, such as the following one:

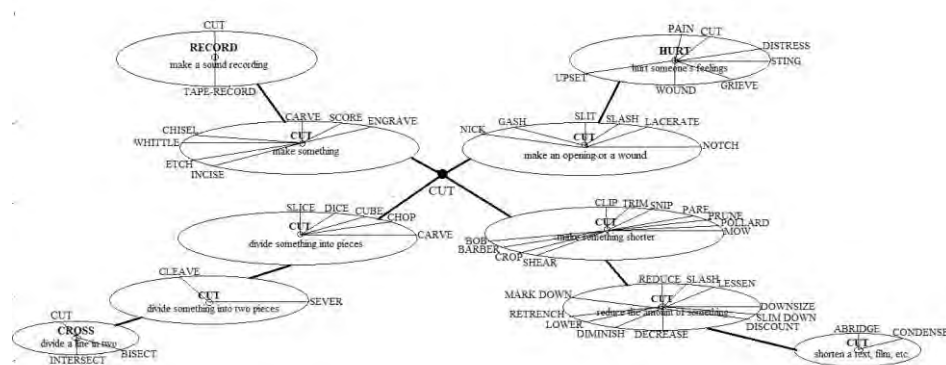


Figure 5: A diagram-based entry for *cut* in an electronic dictionary of synonyms

The diagram has already been recognised in lexicographic practice as suitable for presenting semantic structures of words and lexical relations, as in *The Thinkmap Visual Thesaurus* or *Visuwords*, an online graphical dictionary. The diagram above is based on the prototype theory view of a polysemous struc-

ture as a hierarchically organised set of derivationally related senses gathered around their prototype⁴. This representation of a word's synonymy includes only a part of its corresponding polysemous structure; only those senses for which synonym sets can be formed. Each of the ellipses in the diagram for the verb *cut* establishes both one of its meanings and the corresponding synonym set, whose internal organisation is prototype-based. In each case, the prototype is in the centre of the area enclosed by an ellipse, while its synonyms are represented by points on the given closed curve. The position of a synonym depends on its similarity with the prototype — the shorter the distance between a point on an ellipse and its centre (shown by the line connecting them), the greater the similarity between the given synonym and the prototype, and vice versa. Synonyms that are equally similar to their prototype are symmetrically positioned on a curve in relation to the centre (e.g. *cut* and *tape-record* in the ellipse presenting the sense 'make a sound recording'). In an electronic dictionary, each synonym on an ellipse would be hyperlinked to the complete information on it, including all elements contained in the sample entry shown in Section 4. All pieces of information contained in the textual entry can thereby also be provided through a diagram. Such a diagram-based entry might offer its users a clearer and more succinct representation of a word's integrated synonymy. Furthermore, this visual format enables a quick and easy search for the desired word, including all the necessary information on semantic contrasts among synonyms.

5. Conclusion

Compared to existing ones, this model functions as an attempt to unify such synonym lists as found in *Collins Thesaurus: The Ultimate Wordfinder* or the *Oxford Thesaurus of English* with information on such synonyms as offered in the *Oxford Learner's Thesaurus*. Providing users with as much relevant information as possible about a word's synonyms within one entry is the driving concept behind this model. In fulfilling this goal, the model cannot follow the selective approach employed by the *Oxford Learner's Thesaurus*, but rather embraces one of inclusiveness.

The proposed model achieves comprehensiveness due to its relying on synonymy's dependence on polysemy, as one of its fundamental principles. It provides synonym sets related to different senses of the given word. The model is inclined towards inclusiveness within an individual synonym set as well, which can be, thus, complex including a wide range of synonyms.

Such an approach to a word's synonymy requires a long and complex dictionary entry providing a vast amount of information, which, however, needs to be clearly presented to dictionary users. The principle of complementarity between polysemy and synonymy also contributes to systematicity in the organization of synonym sets within an entry, which follows the organization of appropriate senses in the given polysemous structure. Hence, there are sec-

tions including sets corresponding to closely related senses. This model demonstrates that a word, beyond its polysemous structure, can acquire an organised structure encompassing all its synonyms. Such organization of synonym sets accompanied by indicators of corresponding word meanings is expected to enable an easier and quicker search for the needed set and, simultaneously, individual synonym.

Further systematicity and consistency in synonymy presentation is achieved by the application of the prototype-based view of synonymy. In each synonym set, all its members are identified and organised relative to the prototype according to a declining scale of their similarity to it, which can also be visually represented in an entry. The organisation of synonyms in a set obviously follows a clearly established system. In addition to this, there is a set pattern in defining synonyms according to which the definition of the prototype serves as the base for definitions of all other synonyms in the set. The applied prototype-based strategies can help users easily understand similarities and differences in meaning between any particular synonym and the prototype, as well as between synonyms themselves. Due to efficient demonstrating of nuances in meaning, it is believed that users can quickly identify synonyms that most accurately and precisely express the meanings sought. The application of prototype-based principles also affords a transparent diagram-based visual representation of a word's synonymy in an electronic dictionary offering a more convenient visual organisation of large amounts of information.

The cooperation of polysemy and synonymy, as a principle, is common to the proposed model and WordNet. However, in this model, based on a lexicographic treatment of synonymy that relies on polysemy, the cooperation of these two relations is viewed from a perspective opposite to that adopted by WordNet, in which polysemy depends on synonymy. It might be assumed that the lexicographic outcome would be the same regardless of which of the two perspectives is employed. Yet this is not true, despite their similarities. While in both cases the outcome is an entry containing a number of synonym sets corresponding to different senses of a polysemous word, a more detailed examination shows that synonym sets in the two models have different functions. As illustrated in Figure 3, synsets in WordNet serve to help a user to identify a particular lexical meaning. The selection of their members and their internal organisation are not meant to provide a comprehensive list of various ways of expressing a particular meaning, or to show mutual semantic distinctions among synonyms they comprise. This indicates that such sets do not fulfil the recognised primary functions of a dictionary of synonyms. In contrast, in the model proposed in this paper, synonym sets are designed precisely to serve these functions. Therefore, a synonym set corresponding to a particular meaning in WordNet will almost certainly be measurably different from one related to the same meaning in the proposed model. For example, the synset formed in WordNet for one of the senses of the verb *cut*, 'reduce the amount of something,' contains the following synonyms: *reduce, cut down, cut back, trim, trim*

down, trim back, cut, bring down. The members of the synonym set formed for the same meaning according to the model proposed in this paper are the following: *cut, reduce, decrease, slash, lessen, diminish, lower, retrench, mark down, discount, slim down, downsize*. Clearly, the representation of polysemy relying on synonymy and the representation of synonymy relying on polysemy are two different models of cooperation between the two lexical phenomena whose differing intrinsic purposes result in markedly different lexicographic outcomes.

Endnotes

1. In this paper, only the lexical segment of the lexicon is taken into account in the formation of synonym sets. Thus, these sets include only words.
2. For the purpose of illustrating the proposed model in Section 4 of the paper, the information on the frequency of particular senses of the verb *cut* has been obtained from English general-purpose and learner's dictionaries listed in the References.
3. For the purpose of forming a sample entry as an illustration of the proposed model, English thesauri listed in the References have been used as sources of possible synonyms of the verb *cut*.
4. More on polysemy from a prototype theory perspective including the presentation of a polysemous structure according to the radial set model in Halas (2016a).

References

A. Dictionaries

- Cambridge Advanced Learner's Dictionary*. Accessed on 1 November 2019. <https://dictionary.cambridge.org/>.
- Collins English Dictionary*. Accessed on 1 November 2019. <https://www.collinsdictionary.com/dictionary/english>.
- Lea, D. (Ed.). 2008. *Oxford Learner's Thesaurus. A Dictionary of Synonyms*. Oxford: Oxford University Press.
- Longman Dictionary of Contemporary English*. Accessed on 1 November 2019. <https://www.ldoceonline.com/>.
- Mackie, W. (Ed.). 2002. *Collins Thesaurus: The Ultimate Wordfinder*. Second edition. Glasgow: HarperCollins.
- Macmillan English Dictionary for Advanced Learners*. Accessed on 1 November 2019. <http://www.macmillandictionary.com/>.
- Merriam-Webster's Collegiate Dictionary*. Accessed on 1 November 2019. <http://www.merriam-webster.com/>.
- Oxford Advanced Learner's Dictionary*. Accessed on 1 November 2019. <http://www.oxfordlearnersdictionaries.com/>.
- Oxford Dictionary of English*. Accessed on 1 November 2019. <https://www.lexico.com>.
- The Visual Thesaurus*. Thinkmap, Inc. Accessed on 1 February 2021. <https://www.thinkmap.com/>.
- Visuwords*. Accessed on 1 February 2021. <https://visuwords.com/>.

Waite, M. (Ed.). 2009. *Oxford Thesaurus of English*. Third edition. Oxford: Oxford University Press.
WordNet. Accessed on 1 February 2020. <http://wordnetweb.princeton.edu/perl/webwn>.

B. Other literature

- Adamska-Salaciak, A. 2013. Equivalence, Synonymy, and Sameness of Meaning in a Bilingual Dictionary. *International Journal of Lexicography* 26(3): 329-345.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Cruse, D.A. 2002. Paradigmatic Relation of Inclusion and Identity III: Synonymy. Cruse, D.A., F. Hundsnurscher, M. Job and P.R. Lutzer (Eds.). 2002. *Lexikologie / Lexicology: Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen / An International Handbook on the Nature and Structure of Words and Vocabularies*. Vol. 1: 485-497. Berlin: De Gruyter.
- Dirven, R. and M. Verspoor. 2004. *Cognitive Exploration of Language and Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Dragičević, R. 2007. *Leksikologija srpskog jezika*. Beograd: Zavod za udžbenike.
- Evans, V. 2005. The Meaning of Time: Polysemy, the Lexicon and Conceptual Structure. *Journal of Linguistics* 41(1): 33-75.
- Geeraerts, D. 1989. Prospects and Problems of Prototype Theory. *Linguistics* 27(4): 587-612.
- Geeraerts, D. 2001. The Definitional Practice of Dictionaries and the Cognitive Semantic Conception of Polysemy. *Lexicographica* 17: 6-21.
- Halas, A. 2013. The Representation of a Multidimensional Semantic Structure in English Monolingual Learner's Dictionaries. Gudurić, S. and M. Stefanović (Eds.). 2013. *Languages and Cultures in Time and Space II*: 269-280. Novi Sad: Faculty of Philosophy, University of Novi Sad.
- Halas, A. 2016a. The Application of the Prototype Theory in Lexicographic Practice: A Proposal of a Model for Lexicographic Treatment of Polysemy. *Lexikos* 26: 124-144.
- Halas, A. 2016b. Primena teorije prototipa u predstavljanju polisemije. Ristić, S., I. Lazić Konjik and N. Ivanović (Eds.). 2016. *Leksikologija i leksikografija u svetlu savremenih pristupa*. Zbornik naučnih radova: 455-469. Beograd: Institut za srpski jezik SANU.
- Jehle, G. 2004. Prototype Semantics and Learners' Dictionaries of English. Herbst, T., G. Lorenz, B. Mittmann and M. Schnell (Eds.). 2004. *Lexikografie, ihre Basis- und Nachbarwissenschaften*: 53-69. Tübingen: De Gruyter.
- Jiang, G. and Q. Chen. 2017. A Micro Exploration into Learner's Dictionaries: A Prototype Theoretical Perspective. *International Journal of Lexicography* 30(1): 108-139.
- Lea, D. 2008. Making a Thesaurus for Learners of English. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, Spain, July 15-19, 2008*: 543-550. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- McKeown, M.G. 1991. Learning Word Meanings from Definitions: Problems and Potential. Schwanenflugel, P.J. (Ed.). 1991. *The Psychology of Word Meanings*: 137-156. Hillsdale, NJ: Lawrence Erlbaum.
- Miller, G.A., R. Beckwith, C. Fellbaum, D. Gross and K.J. Miller. 2008. Introduction to Wordnet: An On-line Lexical Database. Fontenelle, T. (Ed.). 2008. *Practical Lexicography: A Reader*: 327-334. Oxford: Oxford University Press.

- Molina, C.** 2008. Historical Dictionary Definitions Revisited from a Prototype Theoretical Standpoint. *Annual Review of Cognitive Linguistics* 6: 1-22.
- Ostermann, C.** 2015. *Cognitive Lexicography*. Berlin/Boston: De Gruyter.
- Pisárčiková, M. and V. Benko.** 1996. Slovak Synonym Dictionary. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C. Rödger Papmehl (Eds.). 1996. *Papers Submitted to the Seventh EURALEX International Congress on Lexicography, EURALEX '96, Göteborg, Sweden, August 13–18, 1996, Vol. 2*: 689-696. Göteborg: Göteborg University, Department of Swedish.
- Rasulić, K.** 2016. *Exploring Semantic Relations*. Belgrade: Faculty of Philology.
- Rundell, M.** 2012. 'It Works in Practice but Will it Work in Theory?' The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the Fifteenth EURALEX International Congress, EURALEX 2012, Oslo, Netherlands, August 7–11, 2012, Vol. 1*: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Šarić, Lj.** 2011. Kognitivna lingvistika i sinonimija: teorija i leksikografska praksa. *Croatica et Slavica Iadertina* 7(2): 305-325.
- Van der Meer, G.** 2000. Core, Subsense and the *New Oxford Dictionary of English* (NODE). On How Meanings Hang Together, and Not Separately. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8–12, 2000, Vol. 1*: 419-431. Stuttgart: Institut für Maschinelle Verarbeitung, University of Stuttgart.

African Englishes in the *Oxford English Dictionary*

Danica Salazar, *Oxford Languages, Oxford University Press,
Oxford, United Kingdom (danica.salazar@oup.com)*

Abstract: In some of its recent quarterly updates, the *Oxford English Dictionary* has published particularly large batches of new and revised entries from South African English, Nigerian English, and East African English. The present article is a detailed discussion of the editorial work behind these updates for African varieties of English, whose distinctive vocabularies the OED is currently taking steps to cover more widely. The paper explains how words are selected for inclusion, how new entries are researched and written, how new sources of textual evidence such as social media give OED editors greater insight into African Englishes, and how the dictionary's African consultants and users contribute to ensuring the accuracy and authenticity of the OED's definitions, etymologies, pronunciations, and quotation evidence. The article also highlights a selection of African words and phrases that have recently been added to or revised in the OED, all of which show the many different forms of lexical innovation that shape the distinctive vocabularies of African Englishes.

Keywords: OXFORD ENGLISH DICTIONARY, HISTORICAL LEXICOGRAPHY, AFRICAN ENGLISHES, EAST AFRICAN ENGLISH, SOUTH AFRICAN ENGLISH, WEST AFRICAN ENGLISH, NIGERIAN ENGLISH

Opsomming: Afrika-Engels in die *Oxford English Dictionary*. In sommige van die onlangse kwartaallikse bywerkings van die *Oxford English Dictionary* is besonder baie nuwe en hersiene inskrywings uit Suid-Afrikaanse Engels, Nigeriese Engels en Oos-Afrikaanse Engels gepubliseer. Hierdie artikel is 'n gedetailleerde bespreking van die redaksionele werk agter hierdie bywerkings vir Afrika-variëteite van Engels waarvoor die OED tans poog om wyer dekking aan elkeen se kenmerkende woordeskatte te verleen. Hier word uiteengesit hoe woorde vir insluiting geselekteer word, hoe nuwe inskrywings nagevors en gedefinieer word, hoe nuwe bronne van tekstuele bewys soos die sosiale media die OED-redakteurs groter insig in Afrika-Engels gee, en hoe die woordeboek se Afrika-adviseurs en -gebruikers daartoe bydra om die akkuraatheid en geloofwaardigheid van die OED se definisies, etimologieë, uitspraak aanduidings en aanhalingsbewysmateriaal te verseker. Die artikel plaas ook klem op 'n groep Afrika-woorde en -frases wat onlangs toegevoeg is tot of hersien is in die OED, waarvan almal die verskillende vorme van leksikale vernuwing toon wat die kenmerkende woordeskatte van Afrika-Engels vorm.

Sleutelwoorde: OXFORD ENGLISH DICTIONARY, HISTORIESE LEKSIKOGRAFIE, AFRIKA-ENGELS, OOS-AFRIKAANSE ENGELS, SUID-AFRIKAANSE ENGELS, WES-AFRIKAANSE ENGELS, NIGERIESE ENGELS

1. Introduction

The English language has a long and complex history in Africa. It first reached the shores of the continent's West Coast, where trade contact was established with British merchants as early as in the 15th century. English gained a firmer foothold in Africa with the arrival of British settlers in the early 19th century, with these settlers and their descendants forming a significant population of native English-speaking Africans (Van Rooy 2020). Finally, the colonization of large swathes of African territory by the British beginning in the late 19th century led to further entrenchment of English as an African language, to such a degree that in the mid-20th century, when British colonies in Africa became independent nations, these new states chose to continue using English, with many of them even giving the language official status (Schmied 1991).

Today, English is seen in Africa as a prestige language, one that is associated with educational and professional success and upward social mobility. It also serves a very useful role as a neutral lingua franca for the highly multilingual, multi-ethnic, and multicultural communities that form part of modern African countries. However, the violent, repressive, and exploitative imperial history behind the spread of English in Africa is far from forgotten, and the lasting hegemony of this colonial language is often decried as a threat to the autonomous development of local languages and cultures. Its presence in Africa resulted in the creation of an English-speaking elite and further deepened inequalities between socioeconomic classes on the continent.

Despite its contradictory roles and the conflicting attitudes towards English in Africa, what is clear is that with their continued use of English throughout the centuries, Africans have been able to adapt the language to suit their communicative needs, thereby giving rise to indigenized varieties of the language with their own unique phonological, morphological, lexical, and pragmatic features, now known collectively as African Englishes. It is the documentation of these distinctive features of African varieties of English in a historical dictionary that will be the focus of the present article.

2. Documenting African Englishes in the *Oxford English Dictionary*

2.1 World Englishes in the OED

The *Oxford English Dictionary* (OED) is widely regarded as the undisputed authority on the history of the English language. It is an historical dictionary of unparalleled size and scope, showing not only the current meanings of hundreds of thousands of words, but also tracing their chronological evolution through millions of quotations taken from written examples of authentic language use.

The OED is presently undergoing its first thoroughgoing revision and update since it was first published in full in 1928. This endeavour, the biggest

humanities project in the world, employs over 70 editors based mostly in Oxford and New York. The dictionary's third and latest edition is accessible through *OED Online* (oed.com), where updates are published quarterly. This makes the OED a work in progress, a hybrid text which combines unrevised entries from previous editions with fully revised and newly added entries. One of the key components of the dictionary's revision project is improving and widening its coverage of varieties of English used outside of the United Kingdom and the United States. This is in response to the global spread of English and the emergence of new varieties of the language — collectively termed World Englishes — that are developing their own standards of grammar, pronunciation, and more importantly for a historical dictionary, lexis.

In recent years, the OED has undertaken targeted projects to broaden its coverage of several World Englishes, publishing particularly large batches of new and revised entries for varieties spoken in the Philippines, Malaysia, Singapore, Hong Kong, India, Canada, Bermuda, the Caribbean, and Ireland. The dictionary's latest quarterly updates have also featured the results of ongoing editorial work on African Englishes. The OED's December 2018 update included 26 new entries from South African English (Salazar 2018), while its January 2020 update contained 29 new entries from Nigerian English (Salazar 2020). More recently, the June 2022 update saw the publication of close to 200 new and revised entries for words from East African Englishes, chiefly covering words from Kenya, Tanzania, and Uganda (Salazar 2022).

2.2 Selecting words for inclusion and editing entries

Lexicographers working on the *OED* today have several resources at their disposal to track the emergence of new words and senses from different parts of the English-speaking world. The principal research tool for this lexical monitoring is Oxford Languages'¹ monitor corpus of English (henceforth the Oxford Monitor Corpus), which currently contains over 14 billion words of web-based news content from 2017 to the present day, and is updated each month. This corpus also consists of subsections with texts from different countries, which make it possible to carry out variety-specific searches and comparisons between different varieties of English. As of this writing, the Oxford Monitor Corpus contains around 284 million words of South African English, 158 million words of Nigerian English, 133 million words of Ghanaian English, 129 million words of Kenyan English, and 16 million words of Ugandan English.

Another important resource for lexical monitoring is the OED's Reading Programme database, which currently comprises over 1.3 million separate quotations drawn from a huge variety of writing. The Reading Programme has been in existence since 1857, when the British Philological Society began to recruit volunteer readers to collect quotations for a planned historical dictionary that would later become the OED. The database started as a collection of quotations

written on slips of paper filed alphabetically, but has now evolved into a digitally stored compendium that is fully searchable. An increasing number of publications written in or about World Englishes is now being included in the Reading Programme, so that a greater quantity of quotations from a wider geographical range of texts can be added to the database.

In addition to their in-house resources, the OED's lexicographers also make use of external evidence sources to monitor lexical developments in different varieties of English. They regularly consult text databases, newspapers, journals, and books from across the globe, as well as a number of regional dictionaries and grammars. They are also aided by contributions from members of the public, and specialist advice from an international network of consultants. Various forms of social media have also given lexicographers a view into current, informal, idiosyncratic uses of words from many different places, and even allow them to reach out to speakers of World Englishes to ask them about the words that characterize their vocabulary.

Once a list of candidates for inclusion has been established, OED editors begin to carefully research both print sources and electronic databases to make sure that there are several independent examples of each word being used, for a reasonable amount of time and reasonable frequency in the types of text in which it can be expected to be found. There is no exact timespan and frequency threshold, as this may vary depending on each word. Some words are relatively young, but they were quickly added to the OED because of the huge social impact they had in such a short space of time; other words are not overwhelmingly frequent, but are included because they are of specific cultural, historical, or linguistic significance (see Diamond 2015).

The OED's editors consider thousands of word suggestions every year, reviewing each and every one. Words that have not yet accumulated enough evidence for permanent record in the OED remain on the watch list for continued monitoring, while suggestions for words with sufficiently sustained and widespread use are assigned to an editor.

Editors begin the process of drafting a dictionary entry by reviewing the information gathered so far for their assigned word, before embarking on their own research to trace the word's development. This research might lead them to search newspaper archives, online forums, academic studies, magazines, law tracts, recipe books, or social media for dated or dateable evidence of the word. If a key example is available in a library or archive that they cannot access digitally, they can enlist the help of the OED's network of researchers, who are based at institutions around the world, to track down this example.

For historical research on African English words, online book, journal, and newspaper depositories such as Google Books, Hathi Trust, Internet Archive, Proquest, JSTOR, and Gale are especially useful for finding antedatings — the earliest evidence of a word being used in English. Works of British traders, explorers, and colonizers in Africa, as well as documents relating to British colonial administration, are particularly rich sources of these antedatings. For

example, *First Footsteps in East Africa*, British explorer Richard Francis Burton's record of his first expedition to Somalia, published in 1856, is the source of the OED's antedating for the word *shuka*, a noun referring to a long piece of fabric usually worn as a loincloth or used as a bedsheet: 'He had ... a Shukkah or half Tobe for his daughter' (Burton 1856: 92).

An invaluable source of historical data for South African English is the *Dictionary of South African English* (DSAE), first published in print by Oxford University Press in 1996 and now maintained as an online dictionary. The DSAE follows the same historical principles as the OED, and now consists of over 4,000 uniquely South African words across three centuries. OED editors consult the DSAE for further information on South African vocabulary and usage, as it contains material that is more detailed as regards to this particular variety of English than that published in the OED.

For more contemporary quotations, OED lexicographers working on African English entries usually consult the newspaper database Nexis, which includes the latest issues of a number of newspapers published across Africa. As previously mentioned, social media sites, particularly Twitter, also provide plenty of evidence for current African English vocabulary, especially for slang and colloquial usages. The tweet shown in Figure 1, posted by a Tanzanian on Twitter in 2020, is the last quotation for the OED entry for *chips(i) mayai*, the name of a thick omelette having fried potatoes (chips) mixed in with the eggs during cooking, often eaten as a street food in Kenya and Tanzania.

Figure 1: Tweet cited in the *OED* as an example for the Tanzanian noun *chips(i) mayai*



Although the *OED* now has access to a more geographically diverse range of research resources, it remains true that the amount of lexicographical evidence available for region-specific vocabulary is still very small compared to that for general English. This is also due to the fact that publications in postcolonial nations continue to be edited following British or American standards, so that

lexical innovations in speech may not always make it into published writing (Salazar 2014). For this reason, OED editors are increasingly turning to less mediated forms of writing, such as social media posts and song lyrics, in order to find quotations that more closely approximate the way that speakers of World English talk, and therefore serve as more authentic illustrations of how World English words are used in real-life contexts (Salazar 2021).

Yet another important data source for contemporary evidence of African English usage is the Corpus of Global Web-based English (GloWbE), which consists of about 1.9 billion words of text in 1.8 million web pages from 340,000 websites in 20 different Anglophone countries, including South Africa (45 million words), Nigeria (43 million words), Ghana (39 million words), Kenya (41 million words), and Tanzania (35 million words). Just like the Oxford Monitor Corpus, GloWbE allows for many types of searches and comparisons between different varieties.

Once an editor has collected enough evidence to form a detailed picture of the word, they begin to build their dictionary entry, adding each element, beginning with the headword and including its pronunciation, variant forms, etymology, definition, example quotations, and any other senses or associated phrases it may have. For new senses of existing words, these are included in their chronological position in the entry, with the definition and example quotations.

In choosing the illustrative quotations to include in a World English entry, it is now of utmost importance for an OED editor to ensure that these quotations are taken from works written by speakers of the variety in question, and published in the word's place of origin. Consider this entry for *buka*, a Nigerian English word for a roadside restaurant or food stall (Figure 2).

Figure 2: OED entry for the Nigerian noun *buka*

The screenshot shows the OED entry for 'buka, n.'. The entry includes the following information:

- Pronunciation:** Brit. ˈbuːkɑː; U.S. ˈbuːkɑː; West African English ˈbuka
- Inflections:** Plural *bukas*, unchanged.
- Forms:** 1900s–*buka*, 1900s–*bukka*.
- Origin:** Of multiple origins. Partly a borrowing from Yoruba. Partly a borrowing from Hausa. **Etymons:** Yoruba *bùkà*, Hausa *bùkkà*.
- Etymology:** < (i) Yoruba *bùkà* hut, market stall, and its etymon (ii) Hausa *bùkkà* grass shed, hut.
- Nigerian English.**
- Definition:** A roadside restaurant or street stall with a seating area, selling cooked food at low prices. Cf. *BUKATERIA* *fl.*, *MAMA PUT fl.* Frequently as a modifier, as in *buka food*.
- Quotations:** 1972 K. OMOTOSO *Combat* 23 When they had finished reading the paper, they went down to a *buka* and had a big breakfast of eko and akara. 1991 B. OKRI *Panished Road* (1992) n. vii.146 But the only thing he did...was go into a *bukka*, put away a great bowl of pounded yam. 2009 @bellanaija 16 Feb. in *twitter.com* (accessed 16 July 2019) Buka food was tastier than home made...at least for me. 2017 @Ezendinwanyi 15 Oct. in *twitter.com* (accessed 6 Feb. 2019) I really can't wait to take Muiyiwa to Lagos and show him my former life. Take him to the joints and *bukas*.

On the right side of the entry, there is a note: 'This is a new entry (OED Third Edition, December 2019; most recently modified version published online December 2021).' Below this note are links for 'Entry history' and 'Entry profile'. At the bottom right, there is a section 'In this entry:' with the text 'buka food'.

It can be seen that all the quotations for this entry come from Nigerian sources, from its earliest attested use in 1972 in *The Combat*, a historical novel written by Nigerian author Kole Omotoso, to a 1991 quotation taken from Nigerian writer Ben Okri's Booker Prize-winning novel *The Famished Road*, to tweets posted by Nigerians.

In their definition, an editor can also add geographical labels, as well as provide more discursive information to accompany these labels. Note how the entry for *buka* includes a *Nigerian English* label. The definition also comments on the frequent use of *buka* as a modifier, and cross-refers to *bukateria* and *mama put*, synonymous words for which the OED also has entries.

After the editor finishes their draft, the entry is passed on to several specialist teams at the OED, such as the etymologists, who check the accuracy of and add more detailed information on the origin of the word, and the bibliographers, who review the quotations to ensure that sources are cited correctly.

Table 1: List of varieties of English for which the *OED* has a pronunciation model (as of November 2022)

Australian English	Indian English	Singapore and Malaysian English
Bermudian English	Irish English	South African English
Canadian English	Manx English	Welsh English
Caribbean English	New Zealand English	West African English
East African English	Philippine English	
Hong Kong English	Scottish English	

Another specialist group in the OED is the Pronunciations team. The dictionary's Pronunciation editors have formulated specific pronunciation models for several World English varieties (see Table 1 for a complete list), which can all be consulted on the *OED* website. Each pronunciation model has a dedicated page with a discussion of the rationale for the model, along with major reference sources and potential discrepancies and compromises. These models were developed using pronunciation data drawn from varietal dictionaries, knowledgeable native speakers and expert consultants, and online evidence of natural language use. The OED's East African pronunciation model (Sangster and Moreland 2022) is based primarily on the work of Josef Schmied, who was also the dictionary's phonetic consultant; its pronunciation model for South African English largely drew from the work of Sean Bowerman and the contrasts found in the DSAE (Sangster and Moreland 2016); and its West African English pronunciation model is chiefly informed by the work of Ulrike Gut, who similarly acted as the OED's phonetic consultant (Sangster and Moreland 2020).

These pronunciation models form the basis of variety-specific transcriptions in the International Phonetic Alphabet (IPA) for World English entries, accompanied by recorded audio pronunciations. See, for instance, how in the OED's entry for *buka* (Figure 3), there are the IPA transcriptions and audio pro-

nunciations in British and American English that are standard for all non-obsolete OED entries, but alongside these is the IPA transcription in West African English, and a blue button that, when clicked, will play a recording of the word being pronounced in a West African accent, as determined by the OED's pronunciation model for West African English.

Figure 3: IPA transcriptions and audio pronunciations for the Nigerian noun *buka* in British English, American English, and West African English



The audio pronunciations are not synthesized, but are recordings made by actor-phoneticians recruited for each variety, who came to Oxford University Press' recording facilities in Oxford to read each transcription aloud. The OED currently features 1,016 South African English transcriptions, for which there are 1,010 sound files; 299 West African English transcriptions, for which there are 256 sound files; and 140 East African transcriptions and sound files. The dictionary's review of World English pronunciations is a work in progress, and pronunciation models for more varieties, along with transcriptions and audio recordings for corresponding entries, are continuously added to the dictionary by its Pronunciations team.

Once the dictionary entry has been signed off by each specialist team, it is given a final review by the dictionary's Chief and Deputy Chief Editors, before it finally takes its place in the OED.

The publication of the OED's World English updates is accompanied by detailed release notes on the OED blog. The varieties of English covered by the dictionary are also featured in the dictionary's World English Hub, a section of its publicly accessible pages which serves as a central repository for the content and resources related to World Englishes on the OED site². The OED also promotes further engagement with its World English content by regularly hosting webinars³ on various topics relating to variation in English, as well as bigger events such as the Oxford World English Symposium, held online in April 2022⁴.

3. Elements of the African English lexicon in the OED

This section will highlight a selection of African words and phrases that have recently been added to or revised in the OED, all of which show the patterns of lexical innovation in African Englishes.

3.1 Borrowings

A significant proportion of the OED's new and revised entries for African varieties have been borrowed into English from some of the most widely spoken languages on the continent. For South African English, Afrikaans is a particularly important source for such loanwords, lending two of the oldest words in the dictionary's December 2018 update. *Deurmekaar*, first attested in 1871, is an adjective applied to something that is confused, muddled, or mixed up. The adverb *voetstoots* was first used in English in 1883 as a legal term describing the buying or selling of items in their existing condition, but nearly a hundred years later, it also began to be used more generally to describe actions done unconditionally, without reservation or qualification.

Later borrowings from Afrikaans were first seen in English in the first half of the 20th century. They include *eina* (first attested 1913), an interjection expressing sharp pain or distress, and *dwaal* (1957), a noun referring to a dreamy, dazed, or absent-minded state, frequently used in the phrase *in a dwaal*.

Other words in the OED's recent update for the English of South Africa have their roots in two of the country's other official languages — Xhosa and Zulu. The oldest of these loanwords date to the late 19th century: *amakhosi* (1857), a collective term of Xhosa and Zulu origin for tribal leaders or chiefs in traditional Nguni societies, and *ubuntu* (1860), a word signifying the fundamental values of humanity or of Africanness, also borrowed partly from Xhosa and partly from Zulu. *Ingcibi*, first used in English in 1937, is a Xhosa word for a person who performs circumcisions on young men as part of a traditional rite of passage, while the more contemporary borrowing *Mzansi*, dating from 1999, is the Xhosa name for South Africa, and for South Africans as a people.

Two years later, the OED's new additions for Nigerian English included such loanwords as the aforementioned *buka*, which was borrowed from the Yoruba word *búkà*, meaning a hut or market stall, and was itself borrowed from the Hausa word *búkkàa*, signifying a grass shed or hut. Two other Yoruba borrowings in the Nigerian update are the noun *danfo* (first attested 1973), which refers to a yellow minibus that carries passengers for a fare as part of an informal transport system in Lagos, and *tokunbo* (1990), an adjective denoting an imported second-hand product, especially a car. The latter comes from the Yoruba word *tòkunbò*, which literally translates into English as 'from overseas', and is formed by the words *ti* 'from, belonging to', *òkun* 'ocean, sea', and *bò* 'to return'.

Nigerian Pidgin is another fount of new words for Nigerian English. *Sef*, first evidenced in Ben Okri's novel *Flowers and Shadows*, published in 1980, is an adverb borrowed from Pidgin, which itself could have been an adverbial use of either the English adjective *safe* or the pronoun *self*. It is an emphatic marker added to the end of statements or rhetorical questions, often to express irritation or impatience, as in this quotation from Nigerian author Chimamanda Ngozi Adichie's 2013 novel *Americanah*: 'He could have given you reduced rent

in one of his properties, even a free flat *sef*' (Adichie 2013: 396).

Also coming from pidgin contexts is the verb *chop*, which is a common colloquial word in Ghana and Nigeria meaning 'to eat'. However, beginning in the 1970s, *chop* also developed the sense of acquiring money quickly and easily, and often dishonestly. The negative sense of misappropriating, extorting, or embezzling funds is also in the earlier reduplicative noun *chop-chop* (first attested 1966), which refers to bribery and corruption in public life. This likening of stealing money to actually devouring it is also reflected in the even earlier synonymous phrase *to eat money* (1960), as in the following quotation from the 22 August 2016 issue of Nigeria's *News Chronicle*: 'Our roads were not done. By the end of this year, you will know who ate the money of these roads'.

More recently, the OED's update for East African Englishes published in its June 2022 update is dominated by loanwords from Swahili, the region's main lingua franca. This long list of borrowings includes the oldest of the new entries in the batch, *jembe*, referring to a hoe-shaped hand tool used for digging, which is first attested in an article by Richard Francis Burton published in the *Journal of the Royal Geographical Society* in 1860. Over a hundred years later, renowned Kenyan writer and academic Ngũgĩ wa Thiong'o used the same word in his historical novel *A Grain of Wheat*, first published in 1967.

One of the newest words in the East African update is also a Swahili loan: *sambaza*, a verb originally used to mean 'to send mobile phone credit to someone', but is now used more generally to mean 'to share or send something'. Dating back to 2007, the English word comes from the Swahili word *-sambaza* meaning 'to spread, disperse, scatter', and now also 'to transfer mobile phone credit'. The transmission into English of this usage with reference to mobile phone credit may have been reinforced by the use of the Swahili word in the name for a credit sharing system in Kenya, introduced in 2005.

Other borrowings in this batch include Swahili forms of address such as *mwalimu* 'teacher' (first attested 1884), as well as *Bwana* (1860) and its abbreviation, *Bw* (1973), a title of courtesy or respect prefixed to the surname or first name of a man. There are also expressions and discourse markers of Swahili origin such as *asante sana* (1911) 'thank you', *pole sana* (1966) 'sorry', and *ati* (2010) 'as someone said; reportedly, allegedly'.

In addition to words used throughout the region, the OED's East African update also features words unique to the varieties of English spoken in Kenya, Tanzania, and Uganda. The lexicon of Kenyan English is represented by borrowings from a few of its many languages: for example, *kiondo* (1902) from Kikuyu and *Isukuti* (1972) from Luhya. A *kiondo* is a handwoven bag made from cord or string, now usually of sisal, with long handles or straps that can be slung over the shoulder, typical of the traditional handicraft of the Kikuyu and Kamba peoples of Kenya. An *isukuti* is a wooden drum, traditionally made from a hollowed log, which is usually hung over the shoulder and played by striking with the fingers and palms. *Isukuti* is also the name of a rhythmic, energetic traditional celebratory dance accompanied by drumming and sing-

ing, performed typically at festivals and weddings by the Luhya peoples of Western Kenya, such as the Isukha and Idakho.

Also included in this update are names of Kenyan dishes such as *githeri* (1973), a traditional central Kenyan dish consisting of boiled maize and legumes, typically beans; and *irio* (1931), a dish consisting of mashed potatoes or sweet potatoes with maize, peas, and leafy green vegetables such as spinach, typically eaten as an accompaniment to other dishes. Also newly added to the *OED* are names of traditional Kenyan home-brewed alcoholic drinks: *muratina* (1904), made from the fermented fruit of the sausage tree, also known locally as *muratina*; *busaa* (1967), made from fermented millet, maize, or sorghum flour; and *changaa* (1975), made from fermented millet, maize, or sorghum grains, a liquor so strong its production and distribution were illegal in Kenya until 2010.

As for Tanzanian English, one of the most widely known words from this variety is *daladala*, the name of a van or minibus that carries passengers for a fare as part of a local informal transport system. Dating back to 1983, the English word comes from Swahili, with *daladala* being a reduplication of *dala* 'dollar', perhaps originally as a bus driver's call. *Dala* is also the nickname of the Tanzanian 5-shilling coin, which used to be the typical fare for *daladala* minibuses.

The vocabulary of Ugandan English draws primarily from Luganda, one of the country's major languages (Isingoma 2016). Examples of Lugandan borrowings in this batch are *kaveera* (1994) 'a plastic bag, plastic packaging'; *kwanjula* (1973) 'an engagement ceremony where the families of the bride and groom formally meet'; and *nkuba kyeyo* (1991) 'a Ugandan person working overseas, especially one doing a low-paid or unskilled job' — the Lugandan phrase literally means 'someone who sweeps'. *Katogo* (1940) is another loan word from Luganda — it is the name of a typical Ugandan breakfast dish consisting of *matoke* (banana or plantain) boiled in a pot with various other ingredients. The word later developed a figurative sense, as it began to be used to mean 'a mixture or fusion of disparate elements; a mess, a muddle'.

3.2 Other lexical innovations

The African English word stock is characterized not just by borrowings, but also by lexical innovations based on English elements, several of which have now made their way into the *OED*. They include words formed through suffixation, such as the East African word *unprocedural* (first attested 1929), meaning irregular or illegal; or through compounding, such as *bunny chow* (1972), the name of a popular South African takeaway dish consisting of a hollowed-out loaf of bread filled with curry; *barbing salon* (1979), the Nigerian expression for a barber's shop; and *deskmate* (1850), a compound East Africans use to refer to a person who sits next to another at school.

A few of the African English words in recent *OED* updates were created by shortening existing English words, like the South African greeting *howzit?* (1918)

and the East African verb *collabo* (2008), short for 'collaborate', used especially of musicians. Another example is the Nigerian English adjective *guber* (1989), which is short for 'gubernatorial' — so Nigerians, for instance, would call a person running for governor a 'guber candidate'. Another frequently used clipping with a longer history in English is *agric*. It was originally used in American English around 1812 as a graphic abbreviation for the adjective *agricultural*, but is now used chiefly in this sense in West Africa. In the early 1990s, *agric* began to be used in Nigeria to designate improved or genetically modified varieties of crops or breeds of livestock, especially a type of commercially reared chicken that is frequently contrasted with native chicken. Two decades later, Nigerian students also started to use the word as a noun meaning agricultural science as an academic subject or course.

A number of the African English entries that have recently been documented in the OED are for words that have developed meanings specific to the region. One notable example is *K-leg*, earliest seen in 1842 in British English, but now used mostly in Nigerian English. It is another term for the condition of knock knees, as well as a depreciative name for a person affected with this condition, whose inward-turning knees often resemble the shape of the letter K. It is of such widespread use in Nigeria that by the early 1980s, it had also acquired a figurative meaning — a *K-leg* can now also be any sort of problem, flaw, setback, or obstacle.

Also in Nigerian English, a *gist* (1990) is a rumour, and *to gist* (1992) is to gossip; and something described as *qualitative* (1976) is excellent or of high quality. In East African English, the noun *tarmac* (1982) is also used as a verb meaning 'to walk the streets looking for work; to job hunt'; a person who is *pressed* (1958) needs to go to the bathroom, while a *stage* (1965) is a bus stop or a taxi rank. In Kenyan English, a *biting* (1997) is a bite-sized piece of food, a small snack, appetizer, or canapé; while a *merry-go-round* (1989) is an informal cooperative savings scheme, typically run by and for women, in which each participant regularly contributes an amount, and the whole sum is distributed to the members in turn. *To shrub* is to pronounce or write words in another language in a manner that is influenced by one's mother tongue, and a *shrub* (2008) is a word pronounced or written in this manner. *To shrub* and *shrub* are colloquialisms chiefly used with reference to English or Swahili words pronounced in a manner characteristic of another Kenyan language. In Ugandan English, *to cowardize* (2003) is to act like a coward or to lose one's nerve, while *to extend* (2000) is to move from one's position so as to make room for someone else.

African Englishes also have their share of idiosyncratic phrases. In South Africa, a non-committal, resigned, or ironic 'whatever' is expressed as *ja well no fine*, pronounced quickly, almost as one word. In Uganda, *well done* (1971) is used as a friendly greeting or salutation, especially when encountering a person at work or in a state of activity; *you are lost!* (2013) is also used as a greeting, or in response to a greeting, in a manner similar to 'long time no see'.

3.3 Semantic fields

Schmied (1991: 82, 84) gives the natural and built environment, food, people, and clothing as semantic fields in which Africanisms often occur. This is reflected in the OED, as many of its newly recorded and revised African English entries belong to these domains. Table 2 shows some examples from the dictionary for each of these areas of meaning.

Table 2: Examples of African English entries in the *OED* belonging to commonly occurring semantic fields

Semantic field	Word	Date of first quotation	Variety	Definition
Food	<i>bunny chow</i> , n.	1972	South African	a South African dish consisting of a hollowed-out loaf of bread (or part of a loaf) filled with curry, typically sold as take-away food; a serving of this
	<i>chapo</i> , n.	1993	East African	a thin pancake of unleavened wholemeal bread cooked on a griddle
	<i>mandazi</i> , n.	1937	East African	a small cake consisting of sweetened dough fried in oil, usually triangular in shape and typically eaten as a snack or as an accompaniment to other dishes; (as a mass noun) these cakes collectively
	<i>nyama choma</i> , n.	1980	East African	roasted or grilled meat
	<i>sarmie</i> , n.	1970	South African	a sandwich
Clothing	<i>buibui</i> , n.	1929	East African	a traditional garment worn by Muslim women in East Africa, typically a long black gown with a black head covering that leaves only the eyes or face exposed

	<i>kanga</i> , n.	1895	East African	a type of cotton fabric printed with designs in bright colours, typically in squares or rectangles featuring a border on all four sides, and used especially for women's clothing; a piece of this fabric, often worn as a shawl or wrap
	<i>kanzu</i> , n.	1870	East African	a long, loose-fitting white tunic worn by men
Environment	<i>banda</i> , n.	1908	East African	a hut or shed with a thatched roof, used typically as a rest house or shelter for travellers
	<i>boma</i> , n.	1860	East African	a barrier formed from thorny branches or wooden stakes, used for defence against attacks by enemies or wild animals; a fence, palisade, or stockade
	<i>duka</i> , n.	1912	East African	a small neighbourhood store selling a variety of goods
	<i>spaza</i> , n.	1988	South African	a small, unregulated and unlicensed grocery shop in a township, usually run from a private house
	<i>tembe</i> , n.	1860	East African	a rectangular house with mud walls and a flat roof
People	<i>district surgeon</i> , n.	1829	South African	a doctor appointed by the government to fulfil specific functions in a particular district; (now) spec. a police surgeon or forensic medical examiner
	<i>ingcibi</i> , n.	1937	South African	among Xhosas: a man who performs circumcisions on young males as part of the traditional initiation ritual symbolizing passage into manhood

	<i>non-indigene</i> , adj.	1907	West African	belonging to an ethnic group considered not to be indigenous to a par- ticular area
--	-------------------------------	------	-----------------	--

Another semantic field that is of particular significance to Africa has to do with street food. In Nigerian English, there are at least three words indicating a roadside restaurant or street stall that sells local fare at low prices. One of them is the aforementioned *buka*, while another is *bukateria* (first attested 1980), which adds to *buka* the *-teria* ending from the word *cafeteria*. An even more creative synonym is *mama put*, from 1979, which comes from the way that customers usually order food in a *buka*: they say 'Mama, put ...' to the woman running the stall, and indicate the dish they want. The word later became a generic name for the female food vendors themselves — Nobel Prize-winning Nigerian playwright Wole Soyinka notably includes a Mama Put character in one of his works.

The informal transport systems that emerged in Africa's huge, densely populated cities such as Lagos and Dar es Salaam have also necessitated lexical invention. Apart from the previously mentioned Nigerian *danfo* and Tanzanian *daladala*, there is also the Nigerian *okada* (1993), the term for a motorcycle that passengers can use as a taxi service. The name is a reference to Okada Air, an airline that operated in Nigeria from 1983 to 1997, and its reputation as a fast yet potentially dangerous form of transport, just like the motorcycle taxi.

Tanzania's contemporary music scene is represented in the OED by *Bongo Flava* (2003) and *singeli* (2015). *Singeli* is a Tanzanian style of fast-paced electronic dance music, combining elements of hip-hop with influences from East African popular music such as *taarab* (1969), a form of music originating in Zanzibar. *Bongo Flava*, another style of music from Tanzania, fuses elements of American hip-hop with influences from reggae, R&B, Afrobeat, dancehall, and traditional East African forms of popular music, and features lyrics in Swahili or English. *Bongo* (1993) is a nickname for the city of Dar es Salaam — *bongo* being the Swahili word for 'brain' or 'intelligence', something one needs a lot of in order to thrive in the most populous city in Tanzania.

4. Conclusion

As can be concluded from the many lexical examples given throughout this article, African speakers of English have long adopted the language and adapted it as a means of expressing their own identity, culture, and experience. Yet despite this, their performance in English continues to be unfairly judged based on exonormative norms set by more prestigious varieties such as British and American English.

It is hoped that the OED's continuing efforts to document the lexicon of African Englishes can play a part in changing this deficit view of these varieties.

As explained earlier in this paper, African English entries in the OED undergo the same thorough and comprehensive research process that all entries in the OED go through before publication. By providing all of this high-quality historical dictionary data for African English words, the OED offers proof of the longevity and stability of these words, and by extension, the longevity and stability of African Englishes. This data can also be used by other researchers to undertake different kinds of historical and lexical investigations of these varieties.

In addition, recording African English words in a dictionary, especially an authoritative historical dictionary such as the *OED*, is a highly visible way of recognizing the valuable contributions that African Englishes have made to the development of the English lexicon. As Nigerian writer T.J. Benson commented to Reuters in January 2020 in reference to the OED's then newly published Nigerian English update (Ukomadu and Carsten 2020), 'I think this (recognition) is empowering for lots of us writers and for everyday people, because at the end of the day it ties back to identity and how we perceive ourselves and how we express ourselves'.

Endnotes

1. Oxford Languages is the department of Oxford University Press that is home to the *Oxford English Dictionary* as well as a wide range of dictionaries and lexical datasets for English and other languages.
2. The OED's World English Hub can be found at <https://public.oed.com/world-englishes/>. It has individual pages for East African English, Kenyan English, Nigerian English, South African English, Tanzanian English, and Ugandan English.
3. Information on the OED's upcoming webinars, as well as recordings of its past webinars can be found at <https://public.oed.com/webinars-and-events/>. A recent webinar relating to African Englishes is 'Mama put in the OED: World Englishes and the *Oxford English Dictionary*', featuring Kingsley Ugwuanyi, the OED's consultant for Nigerian English.
4. All recordings from the Oxford World English Symposium 2022 can be found at <https://public.oed.com/world-englishes/oed-symposium-2022/>. The Symposium features three presentations on African Englishes: 'World Englishes and context-based ELT in Nigeria' by Joy Onyemaechi of the University of Abuja, 'South African English and OUP Southern Africa' by Phillip Louw, formerly of OUP South Africa and now Editor-in-Chief and Executive Director of the *Woordeboek van die Afrikaanse Taal*, and 'Ugandan English' by Bebwa Isingoma of Gulu University.

References

Dictionaries and tools

Corpus of Global Web-Based English (GloWbE). <https://www.english-corpora.org/glowbe/>. Accessed on 19 Nov. 2022.

Dictionary of South African English (DSAE). Van Niekerk, T. (Ed.). <https://www.dsae.co.za/>. Accessed on 19 Nov. 2022.

Oxford English Dictionary (OED) (2000–). Third edition (in progress). Simpson, J.A. (2000–2013) and M. Proffitt (2013–). (Eds.). Oxford: Oxford University Press. <https://www.oed.com/>. Accessed on 19 Nov. 2022.

Other literature

Adichie, C.N. 2013. *Americanah*. London: 4th Estate.

Burton, R.F. 1856. *First Footsteps in East Africa*. London: Longman, Brown, Green and Longmans.

Diamond, G. 2015. Making Decisions about Inclusion and Exclusion. Durkin, P. (Ed.). 2015: 532-545.

Durkin, P. (Ed.). 2015. *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press.

Isingoma, B. 2016. Lexical Borrowings and Calques in Ugandan English. Meierkord, C. et al. (Eds.). 2016: 149-172.

Meierkord, C., B. Isingoma and S. Namyalo (Eds.). 2016. *Ugandan English: Its Sociolinguistics, Structure and Uses in a Globalising Post-Protectorate*. Amsterdam: John Benjamins.

Salazar, D. 2014. Towards Improved Coverage of Southeast Asian Englishes in the *Oxford English Dictionary*. *Lexicography* 1(1): 95-108.

Salazar, D. 2018. *South African Additions to the OED*. Oxford. <https://public.oed.com/blog/south-african-additions-oed/>. Accessed on 19 Nov. 2022.

Salazar, D. 2020. *Release Notes: Nigerian English*. Oxford. <https://public.oed.com/blog/nigerian-english-release-notes/>. Accessed on 19 Nov. 2022.

Salazar, D. 2021. Documenting World Englishes in the *Oxford English Dictionary*: Past Perspectives, Present Developments, and Future Directions. Onysko, A. (Ed.). 2021. *Research Developments in World Englishes*: 271-294. London: Bloomsbury.

Salazar, D. 2022. *Release Notes: East African English*. Oxford. <https://public.oed.com/blog/release-notes-east-african-english-june-2022/>. Accessed on 19 Nov. 2022.

Sangster, C. and M. Moreland. 2016. *Pronunciation Model: South African English*. Oxford. <https://public.oed.com/how-to-use-the-oed/key-to-pronunciation/pronunciations-for-world-englishes/pronunciation-model-south-african-english/>. Accessed on 19 Nov. 2022.

Sangster, C. and M. Moreland. 2020. *Release Notes: West African English Pronunciations*. Oxford. <https://public.oed.com/blog/release-notes-west-african-english-pronunciations/>. Accessed on 19 Nov. 2022.

Sangster, C. and M. Moreland. 2022. *Pronunciation Model: East African English*. Oxford. <https://public.oed.com/how-to-use-the-oed/key-to-pronunciation/pronunciations-for-world-englishes/pronunciation-model-east-african-english/>. Accessed on 19 Nov. 2022.

Schmied, J.J. 1991. *English in Africa: An Introduction*. New York: Longman.

Ukomadu, A. and P. Carsten. 2020. Nigeria and the *Oxford English Dictionary* 'rub minds'. *Reuters*, 28 January 2020. <https://news.trust.org/item/20200128092541-su613>. Accessed on 19 Nov. 2022.

Van Rooy, B. 2020. English in Africa. Schreier, D., M. Hundt and E.W. Schneider (Eds.). 2020. *The Cambridge Handbook of World Englishes*: 210-235. Cambridge: Cambridge University Press.

Heming Yong and Jing Peng. *A Sociolinguistic History of British English Lexicography*. 2022, xiii + 254 pages. ISBN 978-1-032-02467-7 (Hardback), ISBN 978-1-032-02468-4 (Paperback), ISBN 978-1-003-18347-1 (eBook). London/New York: Routledge. Price \$130.00 (Hardback), Price \$38.99 (Paperback), Price \$35.09 (eBook).

The most striking feature of this book lies not in its diachronic and synchronic enquiry into British English lexicography, but in its adoption of a sociolinguistic perspective which complements the sole dictionary-ontology paradigm in dictionary history research. This volume traces the history of British English lexicography — from its Latin origin before the 16th century to the end of the 20th century — against the backdrop of sociocultural dynamics such as foreign invasion and rule, English language evolution, development of modern linguistics, sociocultural progression, and scientific and technological advancements. It is composed of an introduction section and seven chapters. In the introduction, Yong and Peng first provide the reader with a concise yet informative synopsis of the four paradigms in the evolution of British English lexicography, i.e., prescriptivism, historicism, descriptivism and cognitivism. Then, several key research gaps in previous diachronic studies are identified. In order to address these gaps, they propose their methodology, approaches and historical segmentation of British English lexicography.

Chapter One deals with the embryonic stage of British English lexicography prior to the 16th century, laying a solid foundation of linguistic data for future English monolingual/bilingual dictionary-making. Before touching upon the Latin origin of English lexicography and its lexicographical culture, the authors describe how the English language with its roots in Anglo-Frisian dialects had evolved from Old English to Middle English and how sociocultural factors such as the Roman Invasion, the Anglo-Saxon Invasion, Christianity, the Scandinavian occupation and rule, the Norman Conquest, printing technology, and British overseas colonisation, contributed to it. The authors move on to explicate how English lexicographical culture originated from the annotations of hard-word Latin glossaries compiled at the end of the 7th century and the start of the 8th century by tracing back to Ancient Greek and Roman lexicographical traditions. Afterward, they confirm that English lexicography was born with the annotations collected from Latin glossaries by monks and school masters to meet religious demand. The glossaries and bilingual dictionaries with educational and scholarly values are believed to be the primary achievement in English dictionary-making at this stage. At the end of this chapter, they argue that the four classic English–Latin or Latin–English bilingual dictionaries in the 15th century had already indicated a tendency of English lexicography towards its localisation and nationalisation.

Chapter Two centres on the early development of English bilingual dictionary-making and how it started to deviate from the Latin traditions in the 16th century. The authors largely ascribe the preliminary transition of English

lexicography, from Latin traditions to its localisation and nationalisation, to the ever-growing national awareness and the higher status of English language in British society. The economic, sociopolitical, linguistic and cultural events that facilitate this shift are adequately illustrated, including export-oriented economy and overseas colonial expansion, the English Reformation and strengthened state sovereignty, early modern English and the elevated status of the English language, and humanism and Renaissance in Britain. In view of this, the 16th century witnessed a relatively independent development of English lexicography with some of its own innovations. Although Latin–English dictionaries still dominated English lexicography in the first half of the 16th century, dictionaries with English as lemmas gained great momentum in the latter part of the 16th century. The functions of English dictionaries had transformed from basically explaining hard words to providing multiple functions by incorporating diversified lexicographical information. The authors consider the English bilingual dictionaries of the 16th century as necessary preconditions for compiling English monolingual dictionaries in the 17th century.

Chapter Three depicts the rapid growth in English monolingual dictionaries, the fading of hard-word traditions as well as the innovations of bilingual dictionary-making in the 17th century. The sociopolitical background is introduced at length, including the establishment of the East India Company, settlements in North America, the English Civil Wars, the Glorious Revolution, the population boom, promotion of elementary and higher education, and the founding of the Royal Society. At this stage, English had gradually become an international language; the Renaissance was reaching its peak with prominent cultural, scientific and technological advancements. Meanwhile, as the national awareness of the English language was increasingly heightened, heated controversies and discussions arose over its standardisation. Subsequently, dictionary compilation assumed the role of ensuring linguistic norms and purity, which naturally gave impetus to the burgeoning of English monolingual dictionaries. *A Table Alphabeticall* was regarded as the first English monolingual dictionary in the history of English lexicography. It was a hard-word dictionary compiled by Robert Cawdrey and published in 1604. The hard-word traditions of English monolingual lexicography gradually faded away with the closing of monolingual dictionary-making in the 17th century. In contrast, English bilingual lexicography made a great leap towards innovative compiling techniques such as providing explanatory notes, bidirectional configuration, lists of homophones, and a variety of appendixes. The social and academic values of English dictionaries in the 17th century are also discussed in relation to dictionary type, design features and linguistic purism.

Chapter Four ushers English lexicography into the stage of a prescriptive paradigm with the termination of hard-word traditions in the 17th and 18th centuries. The influential sociocultural factors contributing to the shaping of the English language and English lexicography in the 18th century are described in detail, for instance, the Industrial Revolution, the world's first English daily,

the first modern English novel, the establishment of the United Kingdom of Great Britain, Protestantism, and the American War of Independence. In 1664, the Royal Society passed a resolution to set up the English Reformation Committee coping with English standardisation and codification; the nationally unified writing system had been adopted step by step propelled by language-inherent motivations and social impetuses. The authors maintain that printing technology, industrialisation, public literacy, the mass media and publishing industry played an irreplaceable role in promoting and establishing language norms. Driven by linguistic purism, normalisation and prescription, dictionary compilers had to take linguistic standardisation and codification as their top priority in response to the more diversified user demands on authoritative reference works. Thus, English dictionary-making began to shift from a compiler-centred hard-word tradition to a more market-oriented fashion. In this sense, Samuel Johnson's monumental *A Dictionary of the English Language* published in 1755 is a direct outcome and reflection of the pressing need for national language standards. More importantly, it is also a typical showcase of descriptivism established as a paradigm in the history of British English lexicography. Johnson's success was further analysed with a reference to his *Plan of a Dictionary of the English Language* (1747) and the preface of his dictionary. However, the defects and mistakes in his dictionary with respect to headword selection, etymological information, definition, pronunciation, citation, spelling etc. are also mentioned. Other significant English monolingual dictionaries adhering to prescriptivism in the 18th century are furthermore covered and evaluated. The authors close this chapter with the humanistic and academic values of English monolingual, bilingual and other types of dictionaries in the 18th century.

Chapter Five explores the part that European philological traditions played in the creation of a historical paradigm of English dictionary-making in the 19th century. Overall, profound social transformations, economic prosperity and the fast pace of scientific and technological innovations took place in Britain, for example, the First and Second Industrial Revolution, its monopoly status in international trade, the strengthened sociopolitical privilege of the bourgeois class, the founding of the Philological Society of London, and the publication of *The Origin of Species* (1859). In this period, the transformations in British ideology and the progress of science and technology led to a considerable expansion of the English vocabulary. However, no significant changes arose in the pronunciation, grammar, or structure of English. Along with the expansion of the British Empire, English had gradually spread to almost every corner of the globe. As a result, regional varieties of English gradually emerged in different parts of the world. In the meanwhile, English standardisation and the linguistic purity movement did not come to an end. The Philological Society of London was established in 1842 to "investigate and promote the study of and knowledge of the structure, the affinities, and the history of languages" (p. 124). In Yong and Peng's words, it "played a key role in formulating the compiling notions of OED (*Oxford English Dictionary*) and the sophisticated plan for making

its compilation and publication possible" (p. 135). The proposals for and publication of OED (originally titled NED, *A New English Dictionary*) proclaimed the birth and successful implementation of the historical paradigm in English dictionary-making. The initiatives and the practice of compiling a dictionary based on historical principles by the Philological Society were directly linked with the European philological traditions of a historical and comparative paradigm in the study of language origin and families, which broadened the theoretical visions and academic foundation for English lexicography. Although descriptive elements can be found in Johnson's plan for making his dictionary, his masterpiece is prescriptive by nature. Yong and Peng reckon that historical descriptivism lies at the core of the compiling principles of the OED, which indicates the transition of English lexicography from prescriptivism to descriptivism in the 19th century. The authors assert that the academic significance and sociocultural values of the OED have gone far beyond its values as a cultural product, but also "as an icon of scholarship [...] a signifier of British civilisation and a symbol of national identity" (p. 159). The traditional inheritance and theoretical innovations of other major British English dictionaries in the 19th century are also discussed.

Chapters Six and Seven come to the descriptive and cognitive stage of British English lexicography, with a summary of the accomplishments, developments and prospects in the 21st century. The 20th century saw two world wars, the declining of British power, the rise of the United States, English as an international lingua franca, radical sociopolitical transformations, a striking advancement of science and technology, etc. The breakthroughs in computers, information processing and internet technology made possible the application of databases and corpora for dictionary compilation. The dictionary is a carrier of language and culture, a mirror of human communication and language evolution, and a document of the trajectory of civilisation advancement (p. 161). The prosperity of the 20th century English lexicography drew immensely on the emerging new disciplines such as modern linguistics, cognitive science, information science and technologies of the computer and internet in terms of theoretical underpinning and methodological support. The incorporation of information technology into dictionary-making has led to a revolution in the ways of data collection, processing, retrieval, storage, transmission etc. According to Yong and Peng, the pioneering lexicographical works that employed corpora in dictionary-making were *Longman Dictionary of Contemporary English* (1978) and *Collins COBUILD Dictionary of the English Language* (1987). As the flourishing of new linguistic theories, there was a noticeable transition "from the diachronic and comparative description of language to its structural description and from structuralism [...] to cognitivism" (p. 168). It is modern linguistics such as semantics, pragmatics, stylistics, structural linguistics, descriptive linguistics and cognitive linguistics that made the 20th century English dictionaries unusual and characteristic from previous stages. This is especially true of English learner's dictionaries with regard to compiling techniques, defining modes,

structural organisation and other design features. The authors claim that early learner's dictionaries had an evident impact on headword selection, phonetic notation, definition and usage explanation in subsequent philological dictionaries; in the second half of the 20th century, some features such as usage notes and synonym discrimination in English philological dictionaries were modelled on the design of learner's dictionaries. The English learner's dictionaries had gone through three generations of evolution by 1989. The period from 1978 to and after 1995 was an era of the thriving development of the British learner's dictionary family, and "the Big Four" in 1995 was expanded to "the Big Five" with the publication of *Macmillan English Dictionary for Advanced Learners* in 2002. At this stage, the cognitive idea of user-friendly was taken as an essential principle in the designing and compiling of learner's dictionaries so as to earn a competitive edge in the world market. "The Big Five", together with their bilingualised versions, served as the essential reference toolkit for English learners and teachers worldwide. As for the development of British philological dictionaries in the 20th century, OED is again recognised as an unparalleled representative of this type. Other dictionary types such as electronic and online dictionaries, pronouncing dictionaries, dialect dictionaries, slang dictionaries, dictionaries of etymology and usage, thesauruses, encyclopaedic dictionaries, dictionaries of new words, and quotation dictionaries are also investigated.

To some extent, two key words in the title of this volume — "sociolinguistic" and "British" — can distinguish itself from previous works such as *The Oxford History of English Lexicography* (2009) and *The Lexicography of English: From Origins to Present* (2010). First of all, it narrows its focus of enquiry down to British English lexicography rather than the lexicography of world Englishes, though the authors relate it to European and world lexicographical culture when the need arises. Secondly, its greatest innovation lies in incorporating the sociolinguistic perspective into the traditional dictionary-ontology paradigm of diachronic English lexicography, in an attempt to truly represent the interactions between sociocultural dynamics and British English dictionary-making. Thirdly, a trinocular perspective, which integrates the dictionary, its compiler, and its user, is employed to examine the evolution of lexicographical works to overcome compiler-centric limitations. Fourthly, the separate individual case study of English dictionaries and its theorisation are reviewed and refined in the holistic evolution of British English lexicography under a more coherent and unified sociolinguistic framework. Fifthly, a five-stage segmentation of British English lexicography is presented, based on the "regularities and distinctive features" (p. 8) in the evolution of English dictionary-making and previous relevant findings.

Admittedly, aside from the merits mentioned above, the book still has some room to achieve perfection. Chapter Six seems to end somewhat abruptly without providing a due section of the significance and values of the 20th century British English dictionaries as previous chapters do. While contending that descriptivism and cognitivism as paradigms are prevailing in English dictionary-

making in the 20th century and beyond, the authors fail to elaborate them in length in Chapter Six. Albeit, the readers can find some relevant account of the two paradigms scattered in the concluding Chapter Seven.

All things considered, this is probably the first volume on the history of British English lexicography adopting both sociolinguistic and lexicographical perspectives. It portrays a vivid, holistic and dynamic picture of how the paradigms of British English lexicography have developed from its earliest roots to prescriptive, historical, descriptive and eventually cognitive approaches in relation to the sociocultural momentum in British society. This volume will serve as a valuable resource for students and academics of English lexicography and English linguistics.

Acknowledgments

This study was supported by a grant from the National Philosophy and Social Sciences Foundation of China (Grant No. 22BYY010).

References

- Béjoint, H. 2010. *The Lexicography of English: From Origins to Present*. Oxford: Oxford University Press.
Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography*. Oxford: Oxford University Press.

Yanxiu Li
School of Foreign Languages
Jiangxi University of Science and Technology
Ganzhou
China
(yanxiu.li@jxust.edu.cn)
and
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou
China
(yanxiu-li@qq.com)

A New English–Serbian Dictionary of Sports Terms in the Light of Contemporary Challenges

Mira Milić, *Faculty of Sport and Physical Education, University of Novi Sad, Serbia* (miramilicns@gmail.com)

Olga Panić Kavgić, *Faculty of Philosophy, University of Novi Sad, Serbia* (olgapk@ff.uns.ac.rs)

and

Aleksandra Kardoš Stojanović, *Faculty of Philosophy, University of Novi Sad, Serbia* (sandra.kardosh@gmail.com)

Abstract: This paper deals with theoretical and practical aspects of the recently published *Novi englesko–srpski rečnik sportskih termina* [Eng. *A New English–Serbian Dictionary of Sports Terms*] (Milić et al. 2021). The aim is to shed light on solutions that have given rise to innovations in Serbian lexicography. Following the theoretical underpinnings of the lexicographic model and its description, the discussion centers on resolving macrostructural and microstructural issues that could not be brought into accordance with the existing lexicographic and lexicological norms. The problematic issues dealt with include the morphosyntax of headwords, the adaptation of Japanese-based and gymnastics terms from English into Serbian, and the Dictionary's medium in the digital era. The current research builds on previous works from 2017 to 2021, the main result of which was the publication of the Dictionary.

Keywords: DICTIONARY, ENGLISH, SERBIAN, LEXICOGRAPHY, LEXICOLOGY, SPORT, STANDARDIZATION, TERM, TERMINOLOGY

Opsomming: A New English–Serbian Dictionary of Sports Terms in die lig van kontemporêre uitdagings. In hierdie artikel word aandag geskenk aan teoretiese en praktiese aspekte van die onlangs gepubliseerde *Novi englesko–srpski rečnik sportskih termina* [Afr. 'n Nuwe Engels–Serwiese Woordeboek van Sportterme] (Milić et al. 2021). Die doel daarmee is om lig te werp op oplossings wat gelei het tot vernuwings in die Serwiese leksikografie. Ná die teoretiese onderbou van die leksikografiese model en die beskrywing daarvan word daar in die bespreking op die oplos van die makro- en mikrostrukturele kwessies wat nie met die bestaande leksikografiese en leksikologiese norme in ooreenstemming gebring kon word nie, gefokus. Die problematiese kwessies wat hanteer word, sluit die morfosintaksis van trefwoorde, die aanpas van Japannees-gebaseerde en gimnastiekterme uit Engels in Serwies, en die medium van die Woordeboek in die digitale era in. Die huidige navorsing bou voort op voorafgaande werk vanaf 2017 tot 2021, waarvan die belangrikste resultaat die publikasie van die Woordeboek was.

Sleutelwoorde: WOORDEBOEK, ENGELS, SERWIES, LEKSIKOGRAFIE, LEKSIKOLOGIE, SPORT, STANDAARDISERING, TERM, TERMINOLOGIE

1. Introduction

Under the circumstances of the ever-increasing global importance of English, non-native speakers of English, be it students, researchers, or professionals getting in direct or indirect contact with a specific domain, face an increasing need for achieving a good command of English for specific purposes (ESP), especially the register used in a particular field. To achieve this goal, ESP learners use bilingual specialized dictionaries more extensively than other learning resources. In light of this, specialized dictionaries, in general, are a means to preserve the convention, i.e., the linguistic standard, owing to which they help in the preservation of a national language. Thus, they are not only reference sources for translators, but also teaching/learning resources.

Narrowing down the scope of research to bilingual terminological dictionaries, English–Serbian in particular, it can be said that they have recently gained in importance as pedagogical resources, provided that they are conceived in such a way as to supply multifaceted information (linguistic, contact-linguistic, and professional), which is needed for a controlled influx of borrowed lexical material from English into Serbian, knowledge transfer, and active and uninterrupted communication in a particular subject field. That subsumes a scientifically based and user-oriented lexicographic product amenable to further improvements in accordance with users' feedback. The scientific basis of an English–Serbian terminological dictionary implies the application of the latest knowledge in linguistics, English–Serbian contact linguistics, as well as the latest developments in lexicographic theory (cf. Milić 2015b). The user-oriented quality of such a dictionary includes a multitude of ever-changing user needs and preferences, which might be considered from three aspects: firstly, among the most recent ones, it is the inclusion of lexical headwords other than nouns, as well as those extending beyond the lexical level, because they pose communication challenges in a particular field; secondly, it is the impact of the digital revolution on users' expectations in lexicography that requires full attention to the digital environment, which entails innovations in terms of form and content of future dictionaries (Knežević et al. 2021); thirdly, it is the functional profile of the Dictionary which is no longer communicative only (productive and receptive), but also cognitive and increasingly pedagogical, especially in teaching academic ESP (cf. Milić et al. 2018). In a nutshell, it turns out that the most salient issue is to ensure a principle-governed and high-quality lexicographic work (Milić 2015b: 184), as well as to devote more attention to the development of dictionary culture (Nkomo 2015; Prčić 2018).

Motivated by such aims, the project of a new English–Serbian dictionary of sports terms has been implemented, under the title *Novi englesko–srpski rečnik sportskih termina*, henceforth *NESRST*. The exposition is organized as follows:

after the introduction, Section 2 provides a brief account of *NESRST* from theoretical and practical viewpoints; Section 3 deals with elicited lexicographic problems and their solutions; finally, Section 4 draws conclusions and highlights the imported innovations in Serbian lexicography.

2. Theoretical underpinnings of *NESRST*

Since the subject of this paper is a lexicographic resource whose entries are sports terms, it is deemed necessary to point out that the analysis is based on a linguistically based approach to terminology (cf. Cabré 1999; L'Homme 2005; Milić 2004). The fact that professional knowledge and linguistics account for an approximately equal share in terminology is an argument in favor of the conclusion that "terminological analysis must incorporate aspects of both knowledge-driven and lexicon-driven approaches to linguistic content" (L'Homme 2020: 27), i.e., that a dictionary is both a cognitive and linguistic tool (Łukasik 2018: 201). Shifting the focus from the term to the dictionary, it is necessary to point out that a dictionary is not only a reference book consisting of a systematically arranged list of lexical units (cf. Zgusta 1971: 197), but it is also an act of communication between the lexicographer and the user (Milić 2015b: 182), which means that the process of dictionary-making is expected to fulfill the needs and expectations of the user, as particularly emphasized by the communicative theory (Cabré 1999; Yong and Peng 2007) and the more recent theory of lexicographic functions (Bergenholtz 2012; Fuertes-Olivera and Tarp 2014).

Therefore, in terms of the underlying lexicographic theory, *NESRST* is based on the foundations of the communicative theory that also guided the theoretical and practical elaboration of the first edition of the dictionary (Milić 2006). Accordingly, the lexicographic model is only concerned with synchronic aspects of Serbian terms that have been standardized using the first author's model of standardization, theoretically elaborated in Milić (2004), and upgraded on several occasions (Milić 2015a; Milić and Kardoš 2018; 2019). The fact that Serbian terms are regularly created by adaptation of sports terms from English is an argument in favor of the conclusion that the angloglobalized world (cf. Furiassi et al. 2012) leaves its imprint on lexicography, since the limits of language planning are being increasingly extended from the intralingual level to the interlingual one (Bergenholtz and Gouws 2006). Thus, the English-based model of standardization of sports terms in Serbian subsumes that, out of the existing ones, a lexicographer should impose the best fitting term, borrowed or translated from English, or prescribe a new one in accordance with the linguistic standard of Serbian. In other words, *NESRST* is descriptive with elements of prescriptivism. Concerning the attitude towards two opposite approaches, i.e., descriptivism and prescriptivism, *NESRST* slightly deviates from the theory of lexicographic functions, which advocates a more relaxed form of prescriptivism, i.e., a proscriptive approach in which a lexicographer recommends a term out of the existing variants found in the corpus.

Seen within the functional framework, *NESRST* is profiled not only for the communicative function (productive and receptive) but also pedagogical (Béjoint 2010; Hartmann 2001; Łukasik 2016) and cognitive ones (cf. Fuertes-Olivera and Tarp 2011; 2014). The first two are provided by offering standardized terminological units in Serbian created by adaptation of English terms at the level of form and content. The solutions are founded on a corpus-based analysis of the sports register, as well as the latest knowledge in the field of English–Serbian language contacts (Prčić 2019). The fact that English-based terms are brought into accordance with the linguistic standard of Serbian makes the Dictionary fit for teaching and learning ESP, especially at the tertiary level, which was corroborated by the results of a survey in 2017 (cf. Milić et al. 2018). Concerning the cognitive function, which ranks high among lexicographers (cf. Cabré 1999; Fuertes-Olivera and Tarp 2014; Łukasik 2012; 2018; Nielsen 2018), *NESRST* offers multifaceted information accessible from the levels of microstructure and macrostructure alike, by means of which the user acquires knowledge in the specialized field of sport, be it for academic studies, research and professional advancement in sports, translation or merely facilitating communication in this field. The multifunctional character of *NESRST* further entails a full-sentence definition that offers sufficient information for experts, semi-experts, and laypeople, as well as users with different levels of foreign language competence. In this sense, this lexicographic approach differs further from the theory of lexicographic functions, according to which a user/situation-focused dictionary is expected to fit into one of the ten available types of definitions, ranging from full sentences (providing extra information, e.g., active/passive voice, collocates, etc.), to substitutable ones (meaning only) (cf. Kwary 2011: 62).

The following strongly emphasized issue concerns user requirements and expectations. The first step in designing a dictionary is deciding on the potential users (Fuertes-Olivera and Tarp 2014; Yong and Peng 2007: 123–126). According to Cabré (1999: 11), there are two main groups of users — direct users or professionals and intermediaries who indirectly get in contact with a certain register. If applied to *NESRST*, the answer is that it is intended not only for university students of sports and professionals but also non-professionals who directly or indirectly get in contact with the register of sports. All this entails that dictionary entries should contain sufficient and accurate information needed for a full understanding of a term's meaning and its lexical features, which could be accessed fast and easily, regardless of whether the user is a professional or not. Besides, seen within the context of the ever-more demanding digital challenges in the 21st century (cf. Fuertes-Olivera and Tarp 2014; Gouws 2017; Knežević et al. 2021; Lew 2013; Lew and De Schryver 2014; Łukasik 2018: 196), the electronic medium is *the* strongly preferable option. Thus, lexicography is sometimes placed within the realm of information science rather than linguistics (Bergenholtz and Bothma 2011: 74).

Finally, it goes without saying that dictionary-making cannot be reduced

to its final product. Rather, it is a process-oriented activity based on validation (Granger and Paquot 2015: 17), i.e., on studying, nurturing, and monitoring the effects of its use (Prčić 2019: 247). Complying with this requirement, the first edition of the dictionary (Milić 2006) was subjected to revision aimed at publishing a new expanded edition, bearing in mind the following: the users' feedback requirement for including more sports registers (cf. Milić 2015b); the fact that the life span of the dictionary was longer than ten years after which, by default, a dictionary is liable to revision and reprint (cf. Yong and Peng 2007: 118); encouraging reviews of the dictionary model in linguistic circles (cf. Ćirić-Duvnjak 2013: 64; Vasić 2018: 209), as well as the fact that the first edition was sold out. In addition to updating the content and design of the first edition (cf. Milić et al. 2017), the past three years have been devoted to corpus-based studies and standardization of additional English-based sports registers in Serbian (Milić and Kardoš 2018; Panić Kavgić 2021). The main result of these activities is the publication of *NESRST*, which is briefly described in the following subsections.

2.1 Practical aspects of *NESRST*

NESRST represents the result of a teamwork project, involving three linguists — university lecturers with lexicographic experience, as well as twenty-two sports professionals — professors, distinguished athletes, and coaches at the national and international level. As already pointed out, *NESRST* is based on the model of the first edition of *An English–Serbian Dictionary of Sports Terms* (Milić 2006), which was published as a printed edition with approximately 1300 entries of standardized English-based terms in Serbian from the registers of the five most popular ball games in this part of Europe: basketball, football, handball, volleyball, and water polo. Owing to important innovations at the levels of macrostructure and microstructure, as well as the fact that the number of entries in *NESRST* is significantly higher, the title of the new edition has been modified by adding the adjective 'New' (Serb. *Novi*).

2.1.1 Macrostructure of *NESRST*

Seen within the typological framework, *NESRST* is: (i) synchronic, since it contains currently used sports terms; descriptive with elements of prescriptivism, as it includes standardized English-based terms in Serbian; (ii) bilingual — English–Serbian; (iii) printed, in terms of its medium; and medium-sized, when it comes to its volume. In addition to the central alphabetically ordered word list, *NESRST* includes the following secondary front-matter components, whose function is to provide an insight into its aims, theory of the lexicographic model, and user-focused lexicographic information: Table of Contents (Serb. *Sadržaj*), Symbols and Abbreviations (Serb. *Oznake i skraćenice*), Authors' Foreword

(Serb. *Reč unapred*), User's Guide (Serb. *Vodič kroz Rečnik*), and References (Serb. *Literatura*). At the end of *NESRST*, there is an Appendix (Serb. *Dodatak*) with a short description of the included sports.

The central part, i.e., the word list, comprises 5262 entries, one-fourth of which make up the inventory of the already existing dictionary. The entries belong to thirteen registers: four martial arts (boxing, judo, karate, wrestling), three individual sports (athletics, gymnastics, skiing with snowboarding), five ball games (basketball, football, handball, volleyball, water polo) and one racket game (tennis). From the morphosyntactic point of view, the headwords in *NESRST* are English terms that are either single-worded or polylexical, and predominantly nouns and verbs.

2.1.2 Microstructure of *NESRST*

A typical *NESRST* entry consists of six elements, which are described and exemplified in the User's Guide: English term (1), grammatical information (2), cross references (3), Serbian term(s) (4), gloss (5), and diatechnical label (6). The numbered elements are shown in Figure 1 below (*NESRST*, 160). To distinguish between English and Serbian terms used for exemplification in the text, the former is written in italics and the latter in small capitals, with the symbol > in between, to mark the direction of adaptation from English into Serbian.

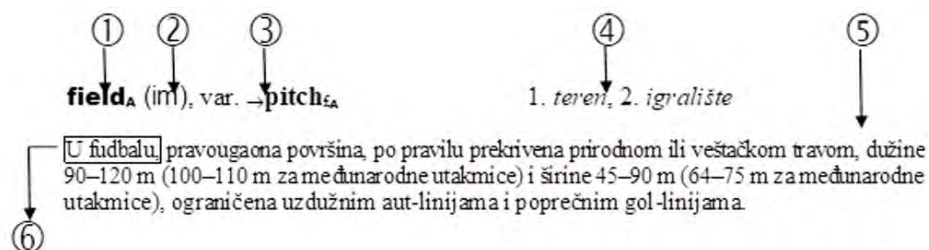


Figure 1: Microstructure of a *NESRST* entry

More details about the elements of a *NESRST* entry are given in the following text.

(1) Alphabetically ordered ENGLISH TERMS as headwords are set in Arial bold typeface, in order to be distinguished from the rest of the text in Times New Roman. Headwords sharing the same form but different grammatical functions, meaning, and/or translation equivalents are re-entered and labeled with subscript codes _A through _H, placed immediately after the headword, as illustrated by Figure 2.

guard_a (gl), sin. → **mark_a**

1. *čuvati (protivnika)*, 2. *markirati (protivnika)*

U fudbalu, košarci i vaterpolu, čuvati protivnika na malom rastojanju – ukoliko je bez lopte, sa ciljem da mu se ometa prijem lopte, pregled igre i njegovo kretanje u slobodnom prostoru ili prema голу ili košu, a ukoliko je sa loptom, sa ciljem da mu se onemogući dobar pregled igre, odigravanje pasa prema slobodnom igraču ili upućivanje lopte na gol ili koš.

guard_b (im), sin. → **guarding**, sin. → **marking**

1. *čuvanje (protivnika)*, 2. *markiranje (protivnika)*

U fudbalu, košarci i vaterpolu, čuvanje protivnika na malom rastojanju – ukoliko je bez lopte, sa ciljem da mu se ometa prijem lopte, pregled igre i njegovo kretanje u slobodnom prostoru ili prema голу ili košu, a ukoliko je sa loptom, sa ciljem da mu se onemogući dobar pregled igre, odigravanje pasa prema slobodnom igraču ili upućivanje lopte na gol ili koš.

guard_c (im)

bek

U košarci, jedan od najčešće dva igrača u petorci, koji su obično najniži, čiji je primarni zadatak da prenesu loptu u prednje polje i organizuju napad, kao i da šutiraju sa većih udaljenosti od koša ili dodaju loptu igraču koji je u najboljoj poziciji da postigne pogodak.

Figure 2: Headwords of the same form with different grammatical functions, meanings and/or translation equivalents

(2) GRAMMATICAL INFORMATION for lexical units is given in the form of an abbreviation of a word class for single-worded and phrasal units¹ alike, as follows: *gl* (verb), *im* (noun), *prid* (adjective), *pril* (adverb), and *uzv* (interjection). A smaller number of multi-word syntactic units are marked as *predl. sint* (prepositional phrase) or *isk* (statement). Listed in descending order, *NESRST* includes: 4854 nouns, 330 verbs, 30 statements, 22 adjectives, 18 prepositional phrases, eight adverbs, and one interjection.

(3) CROSS REFERENCES for synonyms, antonyms, and variants are set in Times New Roman bold. A *NESRST* synonym is a term that has the same but not necessarily identical meaning as another term (predominantly an Anglicism, as opposed to its translation equivalent), while an antonym is a term with a meaning opposite to that of another term. Language variants for British and American English are coded using subscript symbols _£ and _{\$}, respectively. Some cross references are additionally coded with a preceding subscript arrow, as a sign that a particular cross reference is also a headword. By doing so, the authors followed the lexicographic convention applied by Vasić et al. (2018) and Prčić et al. (2021).

(4) Numbered standard SERBIAN TERMS and their variants (if existing) are set in Times New Roman italic typeface, with the standard term labeled as 1. As stated above, Serbian terms have been standardized using the priority-ordered six-principle model² (preciseness, transparency, systematicity, productivity, concision, and frequency). In addition to being numbered, terms with a high frequency of use that are not fully in accordance with the principles of standardization are labeled with the sign ⊕, following the lexicographic convention introduced in Vasić et al. (2018) and Prčić et al. (2021).

(5) GLOSS is a single-sentence definition of meaning that explains the main characteristics of a concept. The definition type fits the traditional 'genus-and-differentia' model (Atkins and Rundell 2008: 436), according to which a term is defined using a hypernym which is followed by additional features that distinguish a particular meaning from those of other hyponyms. In the case of synonymous terms, the gloss is given only once for the first-entered unit, as illustrated by Figure 3.

aerial_g (im), sin. → aerial cartwheel , sin. → no-handed cartwheel , sin. → side aerial	1. ☞ <i>premet strance slobodno, čono,</i> 2. <i>zvezda bez ruku</i>
U sportskoj gimnastici, element sa fazom leta, bez upora rukama, odrazom jednom nogom.	
aerial cartwheel (im), sin. → aerial_g , sin. → no-handed cartwheel , sin. → side aerial	1. ☞ <i>premet strance slobodno, čono,</i> 2. <i>zvezda bez ruku</i>

Figure 3: A *NESRST* entry — synonymous terms

(6) DIATECHNICAL LABEL is an introductory sentence adverbial that follows the pattern "In + name of the sport / a group of related sports / sport in general, ...", which represents an integral part of the definition of a term's meaning. With this adverbial, the translation of the full definition of *field_A* (see Figure 1) reads as follows: "In football, a rectangular area, normally covered in natural or artificial grass, 90–120m long (100–110m for international matches) and 45–90m wide (64–75m for international matches), which is delimited by two side-lines and two goal-lines."

3. Problems addressed and solutions provided

Even though *NESRST* is based on the existing model, theoretically elaborated in Milić (2004) and practically realized in Milić (2006), it could not fit the substantially increased lexical inventory of the new dictionary. Thus, the inclusion of new sports registers has brought to light eight problems related to the morphosyntax of headwords, the adaptation of Japanese-based and gymnastics terms from English into Serbian, and the applied dictionary medium. In the following sections, the problems are elaborated on and exemplified by representative *NESRST* entries.

3.1 Morphosyntactic issues

The findings of previous research on word classes in the terminological system indicate the prevalence of nouns over verbs and other word classes (cf. Cabré 1999; Milić 2015a). As already stated in Section 2.1.2, the same holds for this research.

It goes without saying that word lists in terminological dictionaries are a reflection of such findings. However, recent research advocates giving more room in terminological dictionaries to the previously neglected word classes such as verbs (cf. López Ferrero 2011; Williams and Millon 2010) and even prepositions (Nielsen and Fuertes-Olivera 2009). Since this is still a disputable issue, the authors of *NESRST* have not gone beyond the conceptual framework of the sports register in terms of giving more room to word classes other than nouns. However, three questions come to the foreground in regard to the morphosyntax of the dictionary: firstly, it seems that there is no generally accepted attitude concerning the treatment of proper nouns as lexical entries; secondly, there is the question of the lexicographic status of terms extending beyond lexical limits, albeit less prominent in specialized registers; lastly, there is a problem with the proper description of gender-sensitive terms which is especially typical of Serbian as an inflectional language. The elicited problems and their solutions are elaborated on in the following text.

PROBLEM 1. PROPER NOUNS AS LEXICAL ENTRIES. From the theoretical point of view, these units could not be lexical entries for two reasons. Firstly, they do not have sense (Lyons 1977: 223), and, secondly, most have not developed an appellative function (cf. Zgusta 1971: 246). However, they do have a status of lexical entries to a limited extent in sports dictionaries (cf. Bateman et al. 2006; Milić 2006; Room 2010; Tomlinson 2010), in which they appear as the names of international sports federations and competitions, as well as in some other English–Serbian (and –Croatian) terminological dictionaries (cf. Mihajlov et al. 2004; Pritchard 2013). The situation is similar with reasonably sized general English dictionaries (Atkins and Rundell 2008: 187), most of which include proper nouns. Some authors justify their status of dictionary entries by claiming that all that is exceptional about them is their sheer number (Mignot and Philippe 2022). To establish a standardized approach to the lexicographic treatment of proper nouns in general dictionaries, Bozkurt (2019: 19) concludes that the issue necessitates a detailed study of the attitudes of dictionary users towards proper names, the ways in which they search for them in dictionaries, and the kind of reference sources they use to look them up.

SOLUTION. The way the problem has been solved in *NESRST* is exemplified and described below.

- (1) *Olympic Games* > Olimpijske igre
- (2) *Tsukahara* > 1. cukahara, 2. ☹ Cukahara³

Amidst such conflicting attitudes, the authors of *NESRST* have decided not to include proper nouns, except when they have developed appellative functions. An additional argument in favor of such a solution is that sport does not generally include concepts named by proper nouns, which means that the lexico-

graphic status of these entries is highly dependent on the register. For example, proper nouns would be necessary for dictionaries dealing with astronomy or mycology (cf. L'Homme 2020: 60). Thus, the word list of the previous dictionary was modified so that proper nouns were left out. An exception is the term *Olympic Games* (1), which refers to the biggest and oldest international sports competition and which includes, among others, all sports analyzed in this research. On grounds of this, the authors have decided that the term should be a lexical entry in *NESRST* for cultural and historical reasons. However, eponyms (2) are an exception since they have developed appellative functions. Complying with the orthographic standard of both languages, such an entry in English begins with a capitalized initial letter, as opposed to a small-case one in the Serbian equivalent. Accordingly, the correctly written Serbian term CUKAHARA is the standard unit, whereas its variant CUKAHARA is labeled with the symbol ⊗, to mark a case of unacceptable compliance with the English orthographic standard.

PROBLEM 2. SYNTACTIC UNITS AS DICTIONARY ENTRIES. Unlike the lexicographic model applied in Milić (2006), which includes lexical units only, *NESRST* has a significant share of polylexical terms that are not only lexical units but also syntactic ones, and are, according to their structure, clauses, i.e., sentences⁴. Even though not uncommon in translation manuals (cf. Joksimović 2019) and even terminological dictionaries (cf. Mihajlov et al. 2004; Pritchard 2013; 2015) the introduction of such entries further entails the challenge of defining the meaning using the customary 'genus-and-differentia' model.

SOLUTION. The problem has been resolved as exemplified and described below.

- (3) *out of bounds* > 1. IZGURAVANJE, 2. IZLAZAK IZ BORILIŠTA (Eng. prepositional phrase — Serb. noun, phrasal noun⁵).
In wrestling, the situation when competitors are outside the competition area and the mat chairman stops the fight and then signals its continuation, whereupon the competitors return to the center of the mat.
- (4) *The ball was called in.* > DOSUĐENA JE DOBRA LOPTA. (clause in both languages).
In tennis, the announcement of the chair umpire that repeats the decision of the line judge, or, rarely, the decision of the umpire himself/herself, that the ball has landed within the court boundaries, which is uttered upon a player's request for a challenge and before applying the available electronic system for monitoring the ball's trajectory.

Concerning morphosyntax, dictionary entries (3) and (4) exceed lexical limits, the former being a prepositional phrase, and the latter a clause. Besides, the morphosyntax of the above terms is not necessarily correspondent in the two

languages since a prepositional phrase in English (3) is realized as a lexical unit in Serbian (noun and phrasal noun).

The analysis of similar entries in *NESRST* indicates that they belong to the semantic field of judges' decisions and announcements, which account for a significant segment of a sports event. Encouraged by the finding that dictionaries can and should be much more phrasal than they currently are (Granger and Paquot 2008), the *NESRST* authors have decided to treat syntactic units as dictionary entries on the grounds that they ensure successful communication in certain sports. The practical realization of this innovation required the introduction of two new labels: (*predl. sint*) (Eng. *prepositional phrase*) and (*isk*) (Eng. *statement*). Additionally, adjusting a sentence-long entry to the orthographic standards of English and Serbian imposed the need to capitalize the initial letter and use a period at the end. Concerning the challenge of defining the meaning, both glosses above confirm that, despite expected difficulties, syntactic entries fit the customary 'genus-and-differentia' model.

PROBLEM 3. GENDER-SENSITIVE TERMS. More specifically, this aspect included adding feminine forms of all nominal and adjectival equivalents in Serbian that have developed a corpus-based frequency since the publication of the previous dictionary (Milić 2006).

SOLUTION. The problem and its solutions are exemplified below, where example (5) represents the entry in Milić (2006: 81) without a feminine form, while (5a) is its innovated gender-sensitive description in *NESRST* (171).

(5) *football player* > (1) FUDBALER, (2) NOGOMETAS

(5a) *football player* > 1. FUDBALER, FUDBALERKA, 2. ⊕ NOGOMETAS, NOGOMETASIKA

Complying with research findings relevant to gender-sensitive lexis in Serbian (Savić 2009), the authors have included recently established feminine derivatives of masculine bases that have consolidated their position in the Serbian lexicon. Accordingly, where possible, both gender variants are provided, following the convention in Serbian lexicography, according to which the masculine form comes first, followed by the feminine one.

3.2 Adaptation of Japanese-based terms from English into Serbian

The adaptation of Japanese-based judo and karate terms from English into Serbian has been a challenge. The first difficulty the authors encountered was the insufficient number of translated documents in Serbian, which necessitated additional help from professional athletes and professors of sports and physical education at the University of Novi Sad. Another, and even more serious obstacle was that these originally Japanese terms had first been adapted in

English at the level of content and form, whereas the corresponding Serbian terms were created by adapting the terms from English as an intermediary language. Since the semantic content of these units has remained intact in the process of lexical borrowing, it is not the subject of this analysis. Instead, the focus is shifted towards the phonological and morphosyntactic adaptation of some Japanese-based English terms that could not be carried out in full compliance with the existing linguistic standard of Serbian. The two problems are addressed in the following passages.

PROBLEM 4. PHONOLOGICAL ADAPTATION. The first attempt to standardize English-adapted Japanese terms consisted in bringing them into accordance with the valid orthographic rules in Serbian (Milić and Kardoš 2018). Accordingly, the English-written Japanese sounds *ch*, *i*, and *y* have the following orthographic and phonetic equivalents in Serbian: *ch* – *č* /tʃ/; *i* – *j* /j/ (after a vowel at the end of a word) or *dž* /dʒ/ (in other positions), and *y* – *j* /j/. However, the existing phonological standard of Serbian proved inadequate for the adaptation of the sound /i/ preceded by a vowel at the end of a word.

SOLUTION. Complying with the phonetic norms of Serbian, the following Japanese-based terms would be adapted as follows.

- (6) *chudan* > ČUDAN (Eng. *middle level* [karate])
- (7) *hajime* > HADŽIME (Eng. *start* [judo/karate])
- (8) *seoi-nage* > *SEOJ-NAGE (Eng. *shoulder throw* [judo])
- (9) *yame* > JAME (Eng. *stop* [karate])

Despite being in accordance with the valid standard of Serbian, example (8) turned out to be a problematic issue. Sports professionals were strongly in favor of the English-based form SEOJ-NAGE as unacceptable. According to them, the properly written and pronounced term should be SEOI-NAGE, i.e., the letter 'i' matches the sound /i/ in Serbian irrespective of its position in a word. The problem remained open until recently, when an acceptable solution was found based on a recent study on the adaptation of Japanese names in Serbian and Russian (cf. Marković 2018: 205). Namely, this research highlighted the need to follow the Russian phonological standard, according to which the sound /i/ preceded by a vowel at the end of the word is pronounced /i/. Thus, the correctly adapted term in example (8) is SEOI-NAGE, as exemplified by (8a).

- (8a) *seoi-nage* > 1. SEOI-NAGE, 2. RAMENSKO BACANJE (Eng. *shoulder throw* [judo])

PROBLEM 5. NOMINAL CATEGORY OF NUMBER. The problem of the non-correspondent category of number across languages is not rare. Gardani (2020: 101) demonstrates this finding with the Japanese noun *kimono*, which has a constant invariant paradigm in Italian. Despite the fact that only one lexical unit con-

firms this finding in *NESRST*, its frequency of use in compound headwords has justified the special attention it receives in this analysis.

SOLUTION. The number-category adaptation of Japanese nouns in English and Serbian is exemplified and described below.

- (10) *ashi-waza* > AŠI-VAZA (Eng. *foot or leg techniques*, Serb. NOŽNA TEHNIKA [judo])
(11) *tatami* > TATAMI (Eng. *mat*, Serb. STRUNJAČA [judo])

Focusing on example (10), the *NESRST* authors have decided to treat the term *waza* as singular in Serbian, even though it is plural in Japanese and English. An argument in favor of this solution is that it is a generic name for all judo and karate techniques performed by a certain part of the body. The argument draws on the analysis of the judo and karate register in English, which shows that the term *waza* (Eng. *techniques*) is always lexicalized as the plural-labeled second component of a semi-compound, to indicate that each technique may have several variants in performance. Even though the absence of formal correspondence in the nominal category of number is not rare across languages, the term *waza* is an exception, since most Japanese terms borrowed into English and from English into Serbian have variant number forms, e.g., *tatami(s)* (Serb. TATAMI(JI), *kimono(s)* (Serb. KIMONO/KIMONA), etc.

PROBLEM 6. SEMI-COMPOUNDS. The following question concerns language-dependent orthographic rules for writing compounds and semi-compounds. Due to the low transparency of Japanese words in Serbian, it was not possible to apply the existing orthographic rule for such units borrowed from English into Serbian, according to which the two components of a compound unit are written hyphenated if the meaning of the second component, which is a noun by default, is superordinate to the meaning of the whole (Prčić et al. 2021: 18) — e.g., *art-film* vs. *film*. The matter was additionally challenging because some polylexical semi-compounds consisted of more than two components.

SOLUTION. The problem has been resolved as exemplified and described below.

- (12) *moto-no-ichi* > MOTO-NO-IČI (Eng. *original position*, Serb. PRVOBITNA POZICIJA [karate])
(13) *ude-hishigi-ashi-gatame* > UDE-HIŠIGI-AŠI-GATAME (Eng. *leg lock*, Serb. POLUGA NOGAMA [judo])

As shown in example (12), MOTO-NO-IČI (Eng. *moto-no-ichi*) is double-hyphenated in a semi-compound consisting of three words: *moto* (Eng. *original*), *no* (Eng. *own*), and *ichi* (Eng. *position*). The same is true of example (13), which consists of four words and three hyphens. Since the existing books of orthographic rules in Serbian (Pešikan et al. 2019; Prčić et al. 2021; Vasić et al. 2018) do not offer an

acceptable solution for multi-hyphenated semi-compounds in Serbian, a Japanese language consultant was asked for help. Following her advice, such units are written single-hyphenated or multi-hyphenated due to a typological mismatch between the three languages, as well as the fact that this is the predominant orthographic practice in English as an intermediary language (cf. Iwasaki 2013).

3.3 Adaptation of gymnastics terms from English into Serbian

Despite the global domination of English, this language has not left a significant imprint on Serbian gymnastics since its register has undergone multi-language influences (cf. Milić and Kardoš 2019). Namely, the influence of Czech and German in the early, pre-WWII period of Serbian gymnastics was subsequently followed by that of Croatian and Slovenian. The influence of English might be traced to a more recent period when English established its sociolinguistic status as *the* global language of communication, which is predominantly reflected in the introduction of eponyms (e.g., *Shushuniva* > ŠUŠUNOVA⁶) and metaphoric terms (e.g., *ring jump* > PRSTEN-SKOK⁷). Such multilingual imprint on the Serbian register has resulted in the importation of rather non-transparent lexical borrowings and neologisms. On the one hand, this might be due to the fact that the first gymnastics glossary, which is still used in official documents, was created in the former Serbo-Croatian language (cf. Sedlaček 1966). On the other hand, this may also be the result of the need to imitate the source language word-formation pattern rather than its content, which has given rise to unusual calques.

PROBLEM 7. LEXICAL BORROWINGS AND NEOLOGISMS. The multilingual imprint has put the *NESRST* authors in a difficult situation during the stage of terminological standardization. On the one hand, some terms could not be brought into accordance with the Serbian standard, and on the other, these terms have established consistency of use in official sports circles.

SOLUTION. The problem addressed and the solutions provided are exemplified and described below.

- (14) *free hip handstand* > 1. dotik, 2. kovrtljaj bez dodira telom
- (15) *front uprise* > naupor usklopno
- (16) *illusion* > 1. ilužn, 2. ☹ ilužon, 3. sunce, 4. veliki točak
- (17) *trampoline* > 1. trambulina, 2. ☹ trampolina

In example (14), the standard term is the Czech-based DOTIK, on grounds of its consistent use by sports professionals, even though the English-based translation equivalent (KOVRTLJAJ BEZ DODIRA TELOM) is more transparent in wider sports circles. Similarly, the translation equivalent in example (15), which contains uncustomary compounds (NA+UPOR, U+SKLOPNO), probably created fol-

lowing the Czech compounding pattern, is given the status of the standard Serbian term due to its frequency of use and the domestic origin of the constituent lexical units. The influence of English is illustrated by example (16), where the standard term is an Anglicism ILUŽN and not any of its translated equivalents (SUNCE and VELIKI TOČAK), owing to its consistent use by professionals. The same example also shows a prescriptive intervention in giving preference to the correctly adapted term ILUŽN over its non-standard but more frequent variant ILUŽON preceded by the symbol ☹. The same holds for the English-based form TRAMPOLINA (17), labeled as a variant term of TRAMBULINA, an earlier established lexical borrowing from Italian.

3.4 NESRST and the digital revolution

Looking at lexicography from a general perspective, new technologies have left a notable imprint on dictionary compilation, which makes the question of medium an issue belonging to the theory of lexicography rather than the practical aspect of dictionary-making. As already mentioned, some authors even claim that current lexicography is a part of information science rather than linguistics (Bergenholtz and Bothma 2011: 74). Along similar lines, Nielsen (2013: 370) predicts that dictionaries of the future will be 'digital assistants', which they already are. A digital turnover in lexicography foreshadows multiple challenges for lexicographers such as metalexicography, dictionary design, publishing, marketing, etc. (cf. Bothma 2017; De Schryver et al. 2019; Fuertes-Olivera 2022; Gouws and Tarp 2017). The drift towards information science is due to, among others, abundant information presentation tools on the internet, as most recently exemplified by *ChatGPT* (De Schryver and Joffe 2023) and the newly compiled *ParCoLab* parallel online searchable corpus for linguistic research, containing original sports rulebook texts and their translations into Serbian and three other European languages (cf. <http://parcolab.univ-tlse2.fr/en/>). Besides, the digital format also commends itself to publishers and users in terms of production costs, as the market would hardly allow the full range of necessary dictionaries. This is especially true when it comes to the theory of lexicographic functions advocating the monofunctionality of lexicographic sources (cf. Fuertes-Olivera and Tarp 2014), which seems to be beyond the reach of small languages and many specialized ones, especially if they are in the printed medium. Within the context of abundant pros for digital dictionaries (Bothma 2017; De Schryver 2022; Gouws 2013; 2017; Nkomo 2015; Roche et al. 2019), some earlier claims in favor of the survival of printed dictionaries in the digital era (cf. Nielsen 2013) become outdated.

PROBLEM 8. DICTIONARY MEDIUM. In light of the above-mentioned, the authors were in two minds concerning the dictionary medium. The once foreseen digital future of lexicography has already become the present — however, regret-

tably, without adequate backup and support from information specialists in Serbia. As a result, the question of the *NESRST* medium remained open almost until the submission of the manuscript to the publisher.

SOLUTION. Having given the matter careful consideration, the authors decided to publish a print dictionary based on the following arguments: firstly, the current circumstances in the region of South-East Europe and the Balkans encourage neither the engagement of information specialists in lexicography nor their cooperation with lexicographers; secondly, *NESRST* is designed as a polyfunctional dictionary (receptive, productive, pedagogical, and cognitive); thirdly, in light of the possibility of using *NESRST* as a teaching resource for building English–Serbian contact linguistic competence⁸ (cf. Milić, Glušac and Kardoš 2018) in the conditions of technologically and technically underequipped Serbian educational institutions, giving priority to print medium was the only option. However, as the digital medium has been paving the way for becoming the must-do of lexicography, the authors hope that, at some point in the near future, the time will be ripe for publishing a digital, improved and amended, edition of *NESRST*.

4. Conclusions

This article aimed to shed light on the theoretical and practical aspects of a recently published English–Serbian dictionary of sports terms. Theoretical issues have elicited eight problems related to the morphosyntax of headwords, the adaptation of Japanese-based terms and gymnastics terms from English into Serbian, and the dictionary medium in the digital era. The problems are solved as follows: firstly, in addition to customary lexical entries that have been adapted according to the current linguistic standards of Serbian, *NESRST* also includes syntactic units — however, it does not include proper names; secondly, the adaptation of Japanese-based terms from English into Serbian involves communication-driven solutions that do not necessarily match the adaptation pattern in English and the orthographic standard of Serbian; thirdly, the existing gymnastics terms in Serbian are given the status of standard units regardless of their opacity outside the professional circles, provided that they comply with the linguistic standard of Serbian and English-based terminological standards; lastly, despite the present-day strong preference for the digital medium, *NESRST* was published as a printed dictionary due to insufficient involvement of Serbian language software specialists in the field of lexicography.

In most general terms, the inclusion of uncustomary entries, not necessarily English in origin, makes *NESRST* substantially different from other specialized bilingual dictionaries, both in terms of its corpus and lexicographic description. Namely, headwords exceeding the lexical level raise questions related to the dictionary entry, the corresponding lexicographic description, and the definition of meaning in the target language. From a broader lexicological perspective, it is significant to note that calquing and other means of

creating neologisms for non-English lexical borrowings are missing in the dominant language of global communication and, consequently, in languages importing these units via English.

Acknowledgment

The authors are grateful to the two anonymous reviewers for their constructive comments and suggestions relative to this article and future issues of *NESRST*.

Endnotes

1. According to the term 'phrasal lexeme', used by Lipka (1992: 74) and Lyons (1977: 23).
2. For more information on the standardization model see Milić (2015a).
3. In artistic gymnastics, a type of male or female vault that consists of one-fourth or a half-turn in the first fly stage, i.e., before the support on the vault table, followed by a salto in tuck, pike or layout position — named after the Japanese gymnast Mitsuo Tsukahara (English translation, *NESRST*: 459).
4. In strictly grammatical terms, a clause implies a syntactic structure, i.e., a unit in which the relation of predication has been realized. In other words, it contains a subject–predicate relationship. A sentence and a clause are often functionally congruent, in the case of a simple sentence, i.e., independent clause. When included in the structure of a complex sentence, a clause, either dependent or independent, is treated as a syntactically lower-ranked structure than the sentence. For the sake of clarification, in this article the term 'clause' will predominantly be used with the meaning corresponding to that of a 'simple sentence', which implies a syntactically independent structure containing the relation of predication.
5. According to the term 'phrasal lexeme', used by Lipka (1992: 74) and Lyons (1977: 23).
6. In artistic gymnastics, an element on the floor or the beam, which is performed by jumping up in the straddle position followed by body rotation forward to become parallel to the mat or beam with or without a half-twist, followed by a fall to a face-down position — named after the Soviet gymnast Yelena Shushunova (English translation, *NESRST*: 389).
7. In rhythmic and artistic gymnastics, a kind of jump and a dance element in exercises on the balance beam or the floor, which is performed by jumping off both feet, tossing the head backwards, and arching the body while lifting the feet to the head height (English translation, *NESRST*: 359).
8. According to Prčić (2014: 147), contact linguistic competence is "a type of linguistic knowledge related to the use of elements, i.e., words and names, from English as *the* nativized foreign language in a non-English language that regularly comes into contact with it."

References

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bateman, H., K. McAdam and H. Sargeant. 2006. *Dictionary of Sport and Exercise Science*. London: A & C Black.

- Béjoint, H.** 2010. *The Lexicography of English: From Origins to Present*. Oxford: Oxford University Press.
- Bergenholtz, H.** 2012. Concepts for Monofunctional Accounting Dictionaries. *Terminology* 18(2): 243-263.
- Bergenholtz, H. and T.J.D. Bothma.** 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.
- Bergenholtz, H. and R.H. Gouws.** 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13-45.
- Bothma, T.J.D.** 2017. Lexicography and Information Science. Fuertes-Olivera, P.A. (Ed.). 2017. *The Routledge Handbook of Lexicography*: 197-216. London: Routledge.
- Bozkurt, F.** 2019. Proper Names in General (Purpose) Dictionaries: Necessity. *Uluslararası Eğitim ve Dil Dergisi. International Journal of Education and Language* 1: 10-20.
- Cabré, M.T.** 1999. *Terminology: Theory, Methods, and Applications*. Amsterdam/Philadelphia: John Benjamins.
- Ćirić-Duvnjak, K.** 2013. Uloga anglicizama u sportskoj terminologiji srpskog jezika [The Role of Anglicisms in Serbian Sports Terminology]. *SPORT — nauka i praksa* 3(1): 63-70.
- De Schryver, G.-M.** 2022. Metalexicography : An Existential Crisis. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs and P. Storjohann (Eds.). 2022. *Dictionaries and Society: Proceedings of the 20th EURALEX International Congress, 12–16 July 2022, Mannheim, Germany*: 196-202. Mannheim: IDS-Verlag.
- De Schryver, G.-M., R. Chishman and B. da Silva.** 2019. An Overview of Digital Lexicography and Directions for Its Future: An Interview with Gilles-Maurice de Schryver. *Calidoscópico* 17(3): 659-683.
- De Schryver, G.-M. and D. Joffe.** 2023. *The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks*. Conference paper. 20th CODH Seminar, Center for Open Data in the Humanities, Research Organization of Information and Systems, National Institute of Informatics, Tokyo, Japan, 27 February 2023.
- Fuertes-Olivera, P.A.** 2022. Theoretical, Technological and Financial Challenges: Some Reflections for Making Online Dictionaries. Jackson, H. (Ed.). 2022. *The Bloomsbury Handbook of Lexicography*: 361-374. London: Bloomsbury.
- Fuertes-Olivera, P.A. and S. Tarp.** 2011. Lexicography for the Third Millennium: Cognitive-oriented Specialised Dictionaries for Learners. *Ibérica* 21: 141-162.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter.
- Furiassi, C., V. Pulcini and F. Rodríguez González (Eds.).** 2012. *The Anglicization of European Lexis*. Amsterdam/Philadelphia: John Benjamins.
- Gardani, F.** 2020. Morphology and Contact-induced Language Change. Grant, A.P. (Ed.). 2020. *The Oxford Handbook of Language Contact*: 96-122. New York: Oxford University Press.
- Gouws, R.H.** 2013. Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, V. (Ed.). 2013. *Specialised Lexicography: Print and Digital, Specialised Dictionaries, Data-bases*: 51-62. Berlin/Boston: De Gruyter.
- Gouws, R.H.** 2017. Dictionaries and Access. Fuertes-Olivera, P.A. (Ed.). 2017. *The Routledge Handbook of Lexicography*: 43-58. London: Routledge.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

- Granger, S. and M. Paquot.** 2008. From Dictionary to Phrasebook? Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, Spain, 15–19 July 2008*: 1345–1355. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Granger, S. and M. Paquot.** 2015. Electronic Lexicography Goes Local: Design and Structures of a Needs-Driven Online Academic Writing Aid. *Lexicographica* 31(1): 118–141.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Iwasaki, S.** 2013. Romanization and Text Presentation. Iwasaki, S. (Ed.). 2013. *Japanese: Revised Edition*: xvii–xix. Amsterdam/Philadelphia: John Benjamins.
- Joksimović, J. (Ed.).** 2019. *Priručnik za prevodjenje pravnih akata Evropske unije*. Šesto, izmenjeno i dopunjeno izdanje. [Manual for Translation of EU Legal Acts. Sixth, modified and amended edition]. Beograd: Republika Srbija. Vlada. Ministarstvo za evropske integracije.
- Knežević, Lj., S. Halupka-Rešetar, I. Miškeljin and M. Milić.** 2021. Millennials as Dictionary Users: A Study of Dictionary Use Habits of Serbian EFL Students. *SAGE Open* 11(2): 1–11.
- Kwary, D.A.** 2011. Towards a Typology of Definitions for LSP Dictionaries. *Journal of English Studies* 9: 55–73.
- Lew, R.** 2013. Online Dictionary Skills. Kosem, I. et al. (Eds.). 2013. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia*: 16–31. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies / Eesti Keele Instituut.
- Lew, R. and G.-M. de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341–359.
- L'Homme, M.-C.** 2005. Sur la notion de «terme». *Meta* 50(4): 1112–1132.
- L'Homme, M.-C.** 2020. *Lexical Semantics for Terminology. An Introduction*. Amsterdam/Philadelphia: John Benjamins.
- Lipka, L.** 1992. *An Outline of English Lexicology: Lexical Structure, Word Semantics, and Word-formation*. Second edition. Tübingen: Max Niemeyer.
- López Ferrero, C.** 2011. Grammatical Patterns in Spanish: Verbs of Existence and Appearance. *Corpora* 6(2): 179–199.
- Łukasik, M.** 2012. Terminological Dictionary as a Comprehensive Cognitive and Linguistic Tool. *Language in Different Contexts: Research Papers = Kalba ir kontekstai* 5(1): 98–108.
- Łukasik, M.** 2016. Specialised Pedagogical Lexicography: A Work in Progress. *Polilog. Studia Neofilologiczne* 6: 211–226.
- Łukasik, M.** 2018. Theoretical Underpinnings of Metaterminography. *Applied Linguistics Papers* 25(4): 195–213.
- Lyons, J.** 1977. *Semantics*. Cambridge: Cambridge University Press.
- Marković, Lj.** 2018. Transkripcija i adaptacija imena iz japanskog u srpski i ruski jezik [Transcribing Japanese Onomastic Terms into Serbian]. Bajić, Lj. (Ed.). 2018. *Srpska slavistika: kolektivna monografija. Radovi srpske delegacije na XVI međunarodnom kongresu slavista. Tom 1. Jezik*: 189–211. Belgrade: Savez slavističkih društava Srbije.
- Mignot, É. and M. Philippe.** 2022. Introduction: Proper Names and the Lexicon — An Exposition. *Lexis — Journal in English Lexicology* 20: 1–17.
- Mihajlov, A., M. Ilić, H. Stevanović-Čarapina, S. Tošović and A. Jovović.** 2004. *Englesko-srpski rečnik terminologije u oblasti upravljanja otpadom* [English–Serbian Dictionary of Waste Management]. Belgrade: Misija OEBS-a u Srbiji i Crnoj Gori, Sektor za politiku životne sredine i ekonomska pitanja. Available at <https://www.osce.org/files/f/documents/5/1/25358.pdf> [25 April 2023].

- Milić, M. 2004. *Termini igara loptom u engleskom jeziku i njihovi prevodni ekvivalenti u srpskom* [Ball Game Terms in English and Their Translation Equivalents in Serbian]. Unpublished M.A. Thesis. Novi Sad: Faculty of Philosophy, University of Novi Sad.
- Milić, M. 2006. *Englesko–srpski rečnik sportskih termina* [English–Serbian Dictionary of Sports Terms]. Novi Sad: Zmaj.
- Milić, M. 2015a. Creating English-Based Sports Terms in Serbian: Theoretical and Practical Aspects. *Terminology* 21(1): 1-22.
- Milić, M. 2015b. Principles of Compiling an English–Serbian Dictionary of Sports Terms in the Modern Anglo-globalized World. *ESP Today. Journal of English for Specific Purposes at Tertiary Level* 3(2): 180-195.
- Milić, M., T. Glušac and A. Kardoš. 2018. The Effectiveness of Using Dictionaries as an Aid for Teaching Standardization of English-based Sports Terms in Serbian. *Lexikos* 28: 262-286.
- Milić, M. and A. Kardoš. 2018. Adaptacija izvorno japanske terminologije borilačkih veština iz engleskog jezika u srpskom: teorijski i praktični aspekti [Adaptation of Japanese-based Terminology of Martial Arts from English into Serbian: Theoretical and Practical Aspects]. Živančević Sekeruš, I. (Ed.). 2018. *Deveti međunarodni interdisciplinarni simpozijum "Susret kultura"*. Zbornik radova: 219-231. Novi Sad: Faculty of Philosophy, University of Novi Sad.
- Milić, M. and A. Kardoš. 2019. Adaptation of Gymnastics Terms from English into Serbian. Theoretical and Practical Aspects. *Babel* 65(6): 852-874.
- Milić, M., O. Panić Kavgić and A. Kardoš. 2017. Novi englesko–srpski rečnik sportskih termina: makrostrukturne i mikrostrukturne inovacije [A New English–Serbian Dictionary of Sports Terms: Macrostructural and Microstructural Innovations]. Gudurić, S. and B. Radić-Bojanić (Eds.). 2017. *Jezici i kulture u vremenu i prostoru VI. Tematski zbornik*: 245-256. Novi Sad: Faculty of Philosophy, University of Novi Sad.
- Milić, M., O. Panić Kavgić and A. Kardoš. 2021. *Novi englesko–srpski rečnik sportskih termina* [A New English–Serbian Dictionary of Sports Terms]. Novi Sad: Prometej.
- Nielsen, S. 2013. Chapter 5.3. The Future of Dictionaries, Dictionaries of the Future. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 355-372. London/New York: Bloomsbury.
- Nielsen, S. 2018. LSP Lexicography and Typology of Specialized Dictionaries. Humbley, J., G. Budin and C. Laurén (Eds.). 2018. *Languages for Special Purposes: An International Handbook*: 71-95. Berlin/Boston: De Gruyter Mouton.
- Nielsen, S. and P. Fuentès-Olivera. 2009. Proposals for Upgrading the Lexicographical Treatment of Prepositions in Bilingual Dictionaries for Business Translation. *Revista de Lexicografía* 15: 79-98.
- Nkomo, D. 2015. Developing a Dictionary Culture through Integrated Dictionary Pedagogy in the Outer Texts of South African School Dictionaries: The Case of *Oxford Bilingual School Dictionary: IsiXhosa and English*. *Lexicography ASIALEX* 2: 71-99.
- Panić Kavgić, O. 2021. Teniska terminologija na engleskom i srpskom jeziku: strukturni, semantički i prevodni aspekti [Tennis Terminology in English and Serbian: Structural, Semantic and Translational Aspects]. *Godišnjak Filozofskog fakulteta u Novom Sadu* 46(1): 47-73.
- ParCoLab. Available at <http://parcolab.univ-tlse2.fr/en/> [25 April 2023].
- Pešikan, M., J. Jerković and M. Pižurica. 2019. *Pravopis srpskoga jezika*. Treće, izmenjeno i dopunjeno izdanje [Orthography of Serbian. Third, modified and amended edition]. Novi Sad: Matica srpska.
- Prčić, T. 2014. Building Contact Linguistic Competence Related to English as the Nativized Foreign Language. *System* 42: 143-154.

- Prčić, T.** 2018. *Ka savremenim srpskim rečnicima*. Prvo, elektronsko, izdanje [Towards Modern Serbian Dictionaries. First, electronic edition]. Novi Sad: Faculty of Philosophy, University of Novi Sad. Available at <http://digitalna.ff.uns.ac.rs/sadrzaj/2018/978-86-6065-454-2> [08 August 2022].
- Prčić, T.** 2019. *Engleski u srpskom*. Treće, elektronsko, izdanje [English in Serbian. Third, electronic edition]. Novi Sad: Faculty of Philosophy, University of Novi Sad. Available at <https://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-512-9> [08 August 2022].
- Prčić, T., J. Dražić, M. Milić, M. Ajdžanović, S. Filipović Kovačević, O. Panić Kavgić and S. Stepanov.** 2021. *Srpski rečnik novijih anglicizama*. Prvo, elektronsko, izdanje [A Serbian Dictionary of Recent Anglicisms. First, electronic edition]. Novi Sad: Faculty of Philosophy, University of Novi Sad. Available at <http://digitalna.ff.uns.ac.rs/sadrzaj/2021/978-86-6065-636-2> [08 August 2022].
- Pritchard, B.** 2013. *An English–Croatian Maritime Dictionary*. Available at https://www.ucg.ac.me/skladiste/blog_19782/objava_44640/fajlovi/pomorski%20rjecnik%20Pritchard.pdf [25 April 2023].
- Pritchard, B.** 2015. On Multiword Lexical Units and Their Role in Maritime Dictionaries. *Iranian Journal of English for Academic Purposes* 4(1): 90-117.
- Roche, C., A. Alcina and R. Costa.** 2019. Terminological Resources in the Digital Age. *Terminology* 25(2): 139-145.
- Room, A.** 2010. *Dictionary of Sports and Games Terminology*. Jefferson, NC/London: McFarland & Company. Available at <https://epdf.pub/dictionary-of-sports-and-games-terminology.html> [08 August 2022].
- Savić, S.** 2009. Uputstva za standardizaciju rodno osetljivog jezika [Instructions for the Standardization of Gender-Sensitive Language]. Bečanović, T. et al. (Eds.). 2009. *Njegoševi dani, međunarodni naučni skup, Cetinje, 27–29. jun 2008. Zbornik radova*: 301-320. Nikšić: Faculty of Philosophy.
- Sedlaček, I.** 1966. *Terminologija prostih vežbi sa terminološkim rečnikom i 342 crteža* [Terminology of Simple Exercises with a Terminological Dictionary and 342 Drawings]. Belgrade: Sportska knjiga.
- Tomlinson, A.** 2010. *A Dictionary of Sports Studies*. Oxford: Oxford University Press. Available at <https://www.oxfordreference.com/view/10.1093/acref/9780199213818.001.0001/acref9780199213818-e-477?rskey=UvB6wA&result=477> [01 August 2022].
- Vasić, V.** 2018. *Leksikologija i leksikografija u svetlu savremenih pristupa*. Zbornik naučnih radova. Stana Ristić, Ivana Lazić Konjik, Nenad Ivanović (ur.). Prikaz [Lexicology and Lexicography in the Light of Contemporary Approaches. Ristić, Stana, Ivana Lazić Konjik and Nenad Ivanović (Eds.). Review.]. *Zbornik Matice srpske za filologiju i lingvistiku* 61(1): 203-216.
- Vasić, V., T. Prčić and G. Nejgebauer.** 2018. *Du you speak anglosrpski? Rečnik novijih anglicizama*. Treće, elektronsko, izdanje [Do you speak Anglo-Serbian? A Dictionary of Recent Anglicisms. Third, electronic edition]. Novi Sad: Faculty of Philosophy, University of Novi Sad. Available at <https://digitalna.ff.uns.ac.rs/sites/default/files/db/books/978-86-6065-451-1.pdf>.
- Williams, G. and C. Millon.** 2010. Going Organic: Building an Experimental Bottom-up Dictionary of Verbs in Science. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden/Ljouwert, 6–10 July 2010*: 1251-1257. Leeuwarden/Ljouwert: Fryske Akademy.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam/Philadelphia: John Benjamins.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague/Paris: De Gruyter Mouton.

The Names of the Balkan Peoples and the Names of the Inhabitants of Balkan Countries in Lexicography (on the Example of the *Dictionary of Montenegrin National and Literary Language*)

Sanja Šubarić, *Faculty of Philology, University of Montenegro,
Nikšić, Montenegro (sanjas@ucg.ac.me)*

and

Jovana Đurčević, *Faculty of Philology, University of Montenegro,
Nikšić, Montenegro (jovanadj@ucg.ac.me)*

Abstract: Lexicography is yet another witness of the historic recurrence in the Balkans — fifty years after banning Miloš Moskovljević's *Dictionary of Contemporary Serbo-Croatian Literary Language with Language Manual* (1966), the distribution of the first volume of the *Dictionary of Montenegrin National and Literary Language* (2016) of the Montenegrin Academy of Sciences and Arts was stopped due to political reasons. Among other things, the representatives of the Albanian and Bosniak people in the Parliament of Montenegro demanded that the *Dictionary of Montenegrin National and Literary Language* be withdrawn due to the "offensive" and "discriminatory" definition of the terms *Albanian* and *Bosniak*. This has shown, once again, that Balkan ethnonyms and (or) demonyms have considerable weight in the descriptions of lexical material and that they can lead to the situation where politics defeats lexicography. That is why we will deal, primarily from a linguistic perspective, with the lexicographic definitions of ethnonyms and demonyms that are controversial in Montenegro. Starting from their foundation on ethnic or civic identity, we want to examine the possibilities and justification of the definition of these concepts bearing in mind the broader socio-political framework.

Keywords: ETHNONYM, DEMONYM, PEOPLE, INHABITANT, BALKANS, DICTIONARY, LEXICOGRAPHIC DEFINITION, REDEFINITION, ETHNIC IDENTITY, CIVIC IDENTITY

Opsomming: Die benamings van die Balkanvolke en die benamings van die inwoners van die Balkanlande in die leksikografie (met die *Dictionary of Montenegrin National and Literary Language* as voorbeeld). Die leksikografie is weer eens 'n voorbeeld van die historiese herhalende gebeure in die Balkanlande — vyftig jaar nadat Miloš Moskovljević se *Dictionary of Contemporary Serbo-Croatian Literary Language with Language Manual* (1966), verban is, is die verspreiding van die eerste volume van die *Dictionary of Montenegrin National and Literary Language* (2016) van die Montenegrynse Akademie van Wetenskap en Kuns weens poli-

tieke redes gestaak. Die verteenwoordigers van die Albanese en Bosniese volke in die Parlement van Montenegro het onder andere daarop aangedring dat die *Dictionary of Montenegrin National and Literary Language* weens die "beledigende" en "diskriminerende" definisies van die terme *Albaniër* en *Bosniër* onttrek word. Dit het weereens getoon dat Balkan-etnionieme en (of) demoniemi aansienlike invloed het op die beskrywings van leksikale materiaal en dat hierdie twee taalelemente tot die situasie kan lei waar die politiek die leksikografie die onderspit laat delf. Dit is waarom ons die leksikografiese definisies van etnionieme en demoniemi wat kontroversieel in Montenegro is, hoofsaaklik vanuit 'n linguistiese perspektief sal hanteer. Met hul basis van etniese en burgerlike identiteit as vertrekpunt wil ons die moontlikhede van en regverdiging vir die definisies van hierdie konsepte met inagneming van die groter sosio-politieke struktuur ondersoek.

Sleutelwoorde: ETNONIEM, DEMONIEM, VOLKE, INWONER, BALKAN, WOORDEBOEK, LEKSIKOGRAFIESE DEFINISIE, HERDEFINISIE, ETNIESE IDENTITEIT, BURGERLIKE IDENTITEIT

1. Introduction

According to encyclopedic sources dating back to the times of the Yugoslav state, the geographical position of the Balkan Peninsula made this peninsula a "bridge and main road" between Europe and Asia. This explains the "ethnic mix, diversity of cultural and political influences, and turbulent history" of the Balkans (*Mala enciklopedija Prosveta, opšta enciklopedija*, 1 A-Lj: 119). The same sources emphasize that "Albania, Bulgaria, Greece, Yugoslavia, European Turkey, and partly Romania" were formed as independent states in the 19th century and the first decades of the 20th century, and that their "history is also the history of the Balkans" (ibid.). From today's perspective, it is certain that the last decade of the 20th century was marked by new military conflicts that paved the way for new inter-state border changes in the Balkans in the first decade of the 21st century.

The geographical map of the Balkans changed again on 21 May 2006, when Montenegro seceded from the state union with Serbia, which led to its international recognition. According to the results of the 2011 population census, Montenegro had 620,029 inhabitants belonging to different ethnic groups (Montenegrins, Serbs, Bosniaks, Albanians, Muslims, Croats, Roma, etc.).

According to Article 13 of the Constitution of Montenegro (Parliament of Montenegro 2007): "The official language in Montenegro is the Montenegrin language. The Cyrillic and Latin alphabets are of equal status. Serbian, Bosnian, Albanian, and Croatian are also in official use". This means that all four mutually intelligible national varieties of the once unified Serbo-Croatian language (Montenegrin, Serbian, Bosnian and Croatian) are in official use in Montenegro. However, Montenegro is characterized by a discrepancy "between national and linguistic declaration" (Bugarski 2018: 33). Although Montenegrins (44.98%) are more numerous than Serbs (28.73%) in Montenegro, the 2011 population census (Monstat 2011) shows that Serbian is spoken by 42.88% of the population,

and Montenegrin by 36.97%. Bosnian is the mother tongue of 5.33% of the population in Montenegro, where Bosniak is identified as the mother tongue by only 0.59% of the population. This implies that Bosniaks, who make up 8.65% of the population of Montenegro, predominantly call their language Bosnian. The census also shows that in Montenegro the Albanian language has a larger number of speakers (5.27%) in comparison to the number of inhabitants who declare themselves ethnically as Albanians (4.91%). The situation is different when it comes to Croatian — it is the mother tongue of 0.45% of the population, although Croats make up 0.97% of the population in Montenegro. The population census also showed that 2.03% of the population still call their language Serbo-Croatian, while 3.99% of the population did not want to declare with regard to this matter.

The complex situation with national and linguistic identity is also reflected in the educational system in Montenegro (for more on the problematic naming of subjects/courses in Montenegrin schools/faculties, see Šubarić 2018). Covering the "official language" (Montenegrin) and the languages in "official use" (Serbian, Bosnian and Croatian), a subject of instruction in Montenegrin schools was given a four-part name in 2011/2012: *Montenegrin-Serbian, Bosnian, Croatian language and literature*. However, the public university (University of Montenegro) offers only Montenegrin or Serbian (within separate study programs), with Montenegrin being the only general course in other study programs at the Faculty of Philology (foreign languages) from the academic year 2017/18. Unfortunately, linguists still strongly disagree on the matter of language standardization in spite of the fact that in 2009 and 2010 the Ministry of Education and Science of Montenegro helped standardize a Montenegrin language variety which (unlike Serbian, Croatian and Bosnian) has two additional phonemes and graphemes — *ś, ź*, recognizing, therefore, iotated forms that are unjustified from a systemic point of view. Many Montenegrin linguists do not agree with the normative variety of Montenegrin as it implies "archaization and dialecticization" (Bugarski 2018: 49).

The aim of this paper is to emphasize the importance of rethinking the lexicographic model that defines identity designations such as *Albanian* and *Bosniak*, but also the fact that correct lexicographic definitions of this type require the interdisciplinary overcoming of existing terminological and conceptual inconsistencies inherent in discussions about ethnic and/or national names. Linguistic confrontation with political attitudes towards the identity feelings of ethnic minorities in today's Montenegro showed that minorities feel they belong to both their home and domicile state, which is why they expect their double identity should also be recognized lexicographically.

2. *Rječnik crnogorskog narodnog i književnog jezika (Dictionary of Montenegrin National and Literary Language)* and historic recurrence

At the beginning of April 2016, the Montenegrin Academy of Sciences and Arts

(MASA) published the first volume of *Rječnik crnogorskog narodnog i književnog jezika* (*Dictionary of Montenegrin National and Literary Language* — henceforth DMNLL) on the occasion of the tenth anniversary of the independence of Montenegro. However, members of the Albanian and Bosniak people demanded the withdrawal of the *Dictionary* due to (among other things) the "offensive" and "discriminatory" definition of the entries *Albanian* and *Bosniak*. On 9 June, 2016, a "performance" inappropriate for democratic societies of the 21st century took place in the hall of the Parliament of Montenegro — a member of the Albanian Alternative Party in the Parliament of Montenegro destroyed the *Dictionary* using a paper shredder machine. At the suggestion of the MPs of the Albanian parties, the Parliament of Montenegro adopted a *Resolution on the Dictionary of Montenegrin National and Literary Language* as early as 30 July, 2016. This adoption suspended not only the distribution of the *Dictionary* but, unfortunately, also further work on it.

At this point, we need to return to the past to remind that another lexicographical edition in the Balkans experienced a similar fate fifty years earlier. In 1966, Miloš Moskovljević's (1884–1968) *Rječnik savremenog srpskohrvatskog književnog jezika sa jezičkim savetnikom* (*Dictionary of Contemporary Serbo-Croatian Literary Language with Language Manual*) (55,000 words) was banned for political reasons. Moskovljević's *Dictionary* (as well as the DMNLL) "was anathematized and banned" (Sretenović 2008) in the same year when it was published. This was due to the negative attitude towards "socialism, [...] revolution and the establishment of socialism" (Ćorić 2008: 210). After the publication of the text by Mirko Tepavac, the then editor of the daily newspaper *Politika*, the Belgrade District Court decided to destroy the entire edition of the *Dictionary* because of only three or four words "that were disturbing for the Central Committee of Serbia" (ibid.). In an interview given on the occasion of the publication of the monograph *The Life and Work of Miloš Moskovljević, 1884–1968*, Dr. Momčilo Isiđ explained that Tepavac (in his text in *Politika* from March 6, 1966) found fault with Moskovljević's dictionary because the noun *Croat* was not given as a dictionary entry (Lakićević 2018). Nevertheless, Moskovljević's *Dictionary* had multiple editions — in 1990 and 2001.

Thus, half a century after banning Moskovljević's *Dictionary* on the territory of former Yugoslavia, history repeats itself showing, once again, that Balkan ethnonyms and (or) demonyms have considerable weight in the descriptions of lexical material and that they can lead to the situation where politics defeats lexicography. Although the edition was published by the leading research institution in Montenegro as a contribution to the modern state identity, the deputies of the Montenegrin Parliament perceived it as an attack on multi-ethnic and multi-confessional Montenegro and the collective feelings of certain minority peoples.

Lexicographic activities by MASA on the Montenegrin language were stopped one year after the publication of the first volume of the DMNLL. The first volume of the DMNLL included 12,018 entries — those starting with the

letters A, B, and V (the Cyrillic alphabet, therefore such an order). In the meantime, the MASA lexicographic team also prepared the second volume, which included lexemes beginning with G and D (10,015 entries), but this volume still has not seen the light of day. MASA's controversial publication, unfortunately, once again shed light on the ethnic, demographic and even linguistic polarization in Montenegrin society, and with the interference of the Montenegrin media, it divided the general public as well. This ultimately causes an unfortunate delay in the establishment of lexical norms in Montenegrin.

Owing to the situation with the DMNLL, and our experience as both an editor and a member of the Editorial Board of the DMNLL, we will present the definitions of the lexemes *Albanian* and *Bosniak* from this *Dictionary* and point out their semantic structure. Bearing in mind that these definitions are based on ethnic or civic identity, we will examine the possibility and justification of their redefinition in view of the wider socio-political framework. In that sense, we will analyze the lexicographic status of the terms *ethnonym* and *demonym*, as well as their grammatical or linguistic representation (in the editions published during the official use of the former Serbo-Croatian language, as well as modern editions of the languages that are now spoken in the area of former Serbo-Croatian — Bosnian, Croatian, Montenegrin, Serbian, commonly abbreviated as BCMS). With this analysis, we want to contribute to a more stable theoretical position of both terms, but also to a better conceptual understanding of the names that this paper focuses on.

In that context, we will dwell on the semantic content of the lexemes *people* and *inhabitant*, and try to shed light on the relationship between *a member of a people* and *an inhabitant of a country* from the perspective of multi-ethnic countries. Given the offered models of their lexical descriptions, the lexicographic question will actually deal with the scope and priority of semantic components — is it ethnic or territorial, i.e. civic identity that has priority in defining these names? We will use data from the last population census in Montenegro (2011), which was carried out by the Statistical Office of Montenegro¹. According to the results presented, it is clear that national and ethnic affiliation are equated in the census, which is why we will here treat these concepts as synonymous, without any further distinction between national and ethnic communities, that is, national and ethnic identity² (see Korunić 2003).

3. Linguistic terms *ethnonyms* and *demonyms*³

The critics of DMNLL (including linguists) identified the names that we deal with in this article as *ethnonyms* in their public appearances and statements.

From the linguistic perspective, this topic therefore first requires raising the question — should names as *Albanian*, *Bosnian*, *Bosniak*, *Bulgarian*, *Greek*, *Montenegrin*, *Croat*, *Serb*, *Macedonian*, *Slovene* (and their feminine forms) be defined exclusively as *ethnonyms*?

Considering the status of the term *ethnonym* in the linguistic literature, its frequent identification with the term *demonym*, and the way(s) of their lexicographic representation, we will tackle the question by examining the terms *ethnonym* and *demonym*. We will look into their meaning and semantic distance in lexicographic manuals, as well as their usage in grammatical or linguistic literature primarily belonging to the area of the former Serbo-Croatian language. That is how, bearing in mind the semantic content of the lexemes in question, we want to point out the inconsistencies in their defining, their unjustified identification, as well as the actual broader semantic scope of the term *demonym* compared to its traditional interpretation. We will also address the theoretical inconsistencies related to the nominal status of words belonging to the category of ethnonyms and demonyms, which is why they are defined both as names (proper nouns) and as appellatives (common nouns)⁴. Taking into consideration the modern orthographic standards of the BCMS language (proper nouns are capitalized), we treat these specific words as proper names or proper nouns.

At this point, we want to present the terminological distinction that we use as a theoretical background for the following study: 1. **ethnonym** — the name ascribed to a people or a member of a people; 2. **demonym** — the name denoting the inhabitants of a particular territory; the name of a person whose origin is linked to a certain territory.

3.1 *Ethnonyms and demonyms in lexicographic publications*

The terms *ethnonym* and *demonym* do not have an equal status in onomastic and word-formation systems, and the same holds true for their lexicographic interpretations within different languages. Their unequal treatment can be confirmed in the available dictionaries of English and French⁵, as well as dictionaries of Serbo-Croatian or Serbian, Croatian, and Bosnian (although the normativity of the standard language implies rules established by the orthography, grammar, and dictionary, the Montenegrin language still does not have a standard determined by the normative inventory of lexical material).

The close insight into the lexical descriptions of the above-mentioned terms in lexicographic publications of the former Serbo-Croatian language has shown the following.

Dictionaries of the Serbo-Croatian language (*Rečnik srpskohrvatskog književnog i narodnog jezika*; *Rečnik srpskohrvatskoga književnog jezika*) only offer the entry *demonym*, which is, without terminological designation, defined as the name of inhabitants of a particular place or a country. This is also the case with the one-volume *Rečnik srpskoga jezika* (*Dictionary of the Serbian Language*).

Essentially, the entry *demonym* is treated in the same way in the dictionaries of the Bosnian language, as a term (Halilović, Palić and Šehović 2010) or as a word of general meaning (Jahić 2010). According to these dictionaries, *ethnonym* is not part of the lexicon of the Bosnian language. Furthermore, one of the dictionaries of Bosnian (*Rječnik bosanskog jezika* 2007) lacks both entries.

Veliki rječnik hrvatskoga standardnog jezika (*The Great Dictionary of the Croatian Standard Language*) has both entries — marked as linguistic terms — and defines them as follows: *demonym*, "a word derived from the name of a country, region, city or some other geographical term denoting its inhabitant", and *ethnonym*, "a name of an ethnic community, group or people". However, *Introductory comments* in this dictionary designate *demonym* as a word for "a member of a people or inhabitants", explaining further that these words are listed as part of the dictionary entry with the place name adjectives. In that way, the already presented semantic peculiarity of *demonyms* and *ethnonyms* as linguistic terms becomes problematic.

On that basis, we can note that the Croatian dictionary does not offer the names of the inhabitants of state territories as separate entries. Instead, the names of state territories are given within the passage interpreting a certain place name adjective (derived from the name of a territory), followed by male and female forms of the inhabitants of the given territory. A special lexical description is therefore missing. Hence, the names such as *Albanac*, *Alбанка* (*Albanian*), *Crnogorac*, *Crnogorka* (*Montenegrin*), *Makedonac*, *Makedonka* (*Macedonian*) are qualified as "demonyms" but, quite unexpectedly, the same names or their plural forms⁶ as collective designations of ethnicity are not represented. According to the same principle, the entries *hrvatski* (*Croatian*)i *slovenski* (*Slovene*) are followed by the masculine and feminine forms of the demonyms *Hrvat*, *Hrvatica* (*Croat*), and *Slovenac*, *Slovenka* (*Slovene*), while the status of special lexical units is assigned to the forms *Hrvati* (*Croats*) and *Slovinci* (*Slovenes*) — plural forms are marked with the abbreviation *etn* — for the "specialized dictionary entry" of "ethnic communities (groups)".

The masculine and feminine forms *Bošnjak*, *Bošnjakinja* (*Bosniak*) are given the syntagmatic definition of "a member of a people", and they are also listed as part of the adjective entry *bošnjački* (*Bosniak*). In line with previous inconsistencies in the Croatian dictionary, the masculine and feminine forms *Srbin*, *Srpkinja* (*Serb*) are marked as "demonyms" within the entry *srpski*⁷ (*Serbian*), but *Srbi* (*Serbs*) has the status of a separate entry, denoting the name of a people.

Thus, although *Veliki rječnik hrvatskoga standardnog jezika* (*Great Dictionary of the Croatian Standard Language*) differentiates between demonyms and ethnonyms as special lexical units or linguistic terms, the description of lexical material exclusively relies on demonyms — instead of the term *ethnonym*, the abbreviation *etn* is used or the descriptive identification — *a member of a people*. Bearing in mind certain segments of the specific edition, the general impression is that the semantic content of the terms *demonym* and *ethnonym* and the semantic difference between the *inhabitant of a state* – *a member of a people* are not consistently and adequately represented. On the other hand, it is surprising that *Hrvatska enciklopedija* (*Croatian Encyclopedia*) presents the terms *demonym* and *ethnonym* as synonymous.

Veliki rečnik stranih reči i izraza (*Dictionary of Foreign Words and Expressions*) by Klajn and Šipka (2007), defines an *ethnonym* as "the name of a people, a

member of a people", without labeling it terminologically. The demonym, on the other hand, is marked as a linguistic term denoting "inhabitants in relation to a particular populated place or a country". The same entries, however, cannot be found in *Rječnik stranih riječi* (*Dictionary of Foreign Words*) by Domović, Anić and Klaić (2001).

3.2 *Ethnonyms and demonyms in BCMS grammar books*

In the part reserved for noun formation, *Normativna gramatika srpskog jezika* pays special attention to the so-called *ethnic suffixes*, calling the derivatives formed by this type of suffix only as *demonyms* (Piper and Klajn 2013: 223). The primary definition of the term *demonym* is categorically inconsistent — "broadly speaking, demonyms do not only designate nationality but are also derived from the names of cities, countries, provinces, continents, etc." (ibid.). Based on such a formulation, it could be concluded that the noun *Crnogorac* (*Montenegrin*) is not derived from the two-part toponym *Crna Gora* (*Montenegro*) when it designates nationality (cf. Šimunović 2009: 201). Nevertheless, it can also be concluded that the noun *crnogorizacija* (*Montenegrization*) is a derivative belonging to demonyms. Evidently, the given definition not only neglects the semantic specificity of the terms *demonym* and *ethnonym* but also brings into question the terminological and use value of ethnonyms.

In *Gramatika bosanskoga jezika* (Jahić, Halilović and Palić 2000: 314, 150), *demonyms* are defined as nouns denoting ethnic or geographical affiliation, namely "belonging to a place, people, country, etc." The semantic value of the concept *demonym* includes, therefore, what belongs to the concept *ethnonym*.

The term *demonym* is also present in *Hrvatska gramatika* (Barić et al. 2005) in a chapter dealing with noun formation/suffixation. It is semantically identified as "a word designating the inhabitant of a settlement, region, country, continent, or the appellative for men or women in relation to where they come from" (Barić et al. 2005: 313)⁸.

Gramatika crnogorskoga jezika deals with the word-formation of "nouns referring to the inhabitants and members of a people/country" (Čirgić, Pranjković and Silić 2010: 141, 144), but their terminological naming by means of internationalisms of Greek origin (*demonyms* – *ethnonyms*) is missing.

The above-mentioned grammar books of the BCMS language show, therefore, that none of them identifies the nouns designating "members of a people" as ethnonyms — instead, such meaning is attributed to demonyms in the grammar books of Serbian and Bosnian.

3.3 *Ethnonyms and demonyms in onomastic descriptions*

The fact that the names of peoples and inhabitants are neglected from a theoretical perspective is expressed in the manual of lexicology by Šipka (2006).

Regardless of the time of publication and the linguistic circumstances at the time, the manual was made with the aim of "building a coherent and simple model of a general and Serbo-Croatian lexicon" (Šipka 2006: 7). In the chapter dedicated to onomastics as a "highly autonomous lexicological discipline" that is focused on the proper names, ethnonyms and demonyms are not identified as onomastic categories (Šipka 2006: 141-149, 245).

On the other hand, in support of the frequent semantic differentiation of the terms *ethnonym* and *demonym* in the Croatian scientific context, we will mention *Uvod u hrvatsko jezikoslovlje* by Šimunović (2009). Ethnonyms and demonyms are singled out and defined in this truly unique contribution to the onomastics of Slavic languages as special anthroponymic categories in a separate subsection (5.2.4). Their definitions are, however, more precisely given in the previous part of the book (Šimunović 2009: 75): *demonym* — "the name of the inhabitant of a place, region", *ethnonym* — "the name of a people, nationality, ethnic group".

To a certain extent, Šimunović's interpretation of demonyms and ethnonyms is in accordance with his view that "the singular form of demonyms and ethnonyms has the characteristics of appellatives instead of names, although they retain some features of proper names in terms of their associative content," that is, their singular form does not identify a person as an individual but as a member of a particular collective/group (Šimunović 2009: 200-201) (cf. also Peti 1997). Šimunović (2009) also defines demonyms as "names" when they designate a particular "inhabitant of a place". Contrary to that, he (ibid.) does not define an ethnonym as a separate designation — as a name of a member of "a particular ethnicity/people/nationality". With regard to his definition of ethnonyms, we need to point out the (previously mentioned) theoretical disagreements in the interpretation of the nominal status of ethnonyms and demonyms (proper or common nouns), which is the reason why different lexicographic editions of Serbo-Croatian and BCMS offer different grammatical forms of ethnonyms: plural nominative naming ethnos as a group (*Rečnik srpskohrvatskoga književnog jezika* (Dictionary of the Serbo-Croatian Literary Language), *Veliki rječnik hrvatskoga standardnog jezika*⁹ (Great Dictionary of the Croatian Standard Language, etc.)) or singular nominative naming a member of a particular ethnos (e.g. *Rečnik srpskohrvatskog književnog i narodnog jezika* (Dictionary of Serbo-Croatian Literary and Vernacular Language), *Rečnik srpskoga jezika*¹⁰ (Dictionary of the Serbian Language), *Rječnik crnogorskog narodnog i književnog jezika* (Dictionary of Montenegrin National and Literary Language, etc.)).

Based on the above-mentioned, it can be concluded that the term *demonym* has a longer tradition than *ethnonym* and, therefore, a more stable theoretical position in the linguistic literature of the former Serbo-Croatian language: it always appears as a lexical unit in dictionaries, and mainly as a category of word-formation in current grammar books. Although their lexicographic definitions are generally homogenous, demonyms often have incomplete definitions (they are explained as words denoting people in relation to their residence, not origin).

However, from the perspective of word formation (dominantly incomplete in the previous sense) they are sometimes interpreted as ethnonyms in terms of their semantic content. The internationalism *ethnonym* is unjustifiably absent from most of the consulted lexicographic editions, and its terminological value is neglected in the word-formation classifications of nouns in BCMS or descriptions of their grammatical systems.

On the basis of a broader insight into the literature related to the former Serbo-Croatian language, and considering the editions consulted in this article, it can be concluded that the terminological delineation of the terms *ethnonym* and *demonym* has a relatively stable theoretical status in the Croatian literature — they have the status of autonomous linguistic units, and they are lexicographically defined and categorically positioned in the onomastic system. However, the lexical descriptions of the modern Croatian language show us that a theoretical demarcation is not always easily presented in lexicographic practice.

Considering the above, we can conclude that defining *ethnonyms* and *demonyms* as linguistic terms should be established and consistent (in dictionaries, grammar books, and onomastic descriptions). Ethnonyms should not be reduced to collective designations (peoples), because they also designate individuals as members of a collective (a people). As for demonyms, on the other hand, their definition should not exclude the territorial affiliation based on origin as these names identify individuals in relation to a particular place — the place of residence or origin. In the continuation of the paper, we will show that considering certain political and legal resolutions, their semantic content is wider.

3.4 The relationship between the terms *a people* / *a member of a people* — *an inhabitant of a country* from the perspective of multi-ethnic countries

Observing the terminological peculiarities of the internationalisms *demonym* and *ethnonym* as they are mainly presented in linguistic literature and dictionary editions, we will also dwell on the lexicographic definitions and semantic content of the concepts *a people* and *an inhabitant*.

As a lexical unit, the noun *narod* (a people) has several meanings. In accordance with our topic, we will only interpret the meaning identified as primary in lexicographic definitions. It is according to the primary meaning that this noun refers to a large community of people "with the same ethnic name", formed usually "on the shared origin, language, territory, tradition, culture, religion, social life, etc." (*Rečnik srpskoga jezika* 2007). Given the previously explained interpretations of the terms *ethnonym* and *demonym*, our impression is that their interpretations do not always take into account the semantic specificity of the concepts *a people*, *nation*, *nationality*. The fact that they (and other concepts related to them) are not terminologically differentiated can be proved by the results of the last population census in Montenegro, which showed that national affiliation and ethnic affiliation were perceived as one thing ("national or ethnic affiliation"¹¹) (Monstat 2011). Although the relation-

ship between *a people* and *a nation* is perceived as a complex one in the literature of various social disciplines, the space constraints of the article limit the discussion thereof. Certainly, these are not categories of the same value for social science experts, and the modern age has shed light on their political and legal distinction. It is also certain, however, that lexical descriptions of their meanings are often confused (see the entry *nation* in Vujaklija 2007), or that the dichotomy of *a people* – *a nation* does not have a precise lexicographic form (cf. *Rečnik srpskoga jezika* 2007), and is not, therefore, consistently present in the definition of individual identity designations. In that context, it is important to say that adequate semantic demarcation of the concepts of *a people* and *a nation*, from today's perspective, is important for lexicographic interpretations of the individual names of peoples or nations, but also that correct interpretations of that kind can be expected only after modern social sciences solve conceptual and terminological problems that usually follow discussions of "ethnic and national identity" and social identity in general (see Korunić 2003: 163).

Unlike the lexeme *people*, the lexical scope of the noun *inhabitant* is unambiguous and defined as *a person who is permanently or temporarily residing somewhere*.

The semantic interpretation of the lexeme *people* and *inhabitant* also raises the question of the distinction between the name of a people from the name of the inhabitants of a country, considering that the members of the people who have their own country are mostly the inhabitants of that country.

The statement that a member of a people is the same as "a resident of the country (if one refers to it as a state)" (Vuković 2007: 172; cf. 173-174) is not realistic from the perspective of demographic reality and multi-ethnic states. In fact, the given claim is denied by the fact that representatives of different peoples usually live in one state, but also that certain peoples do not have their own state territory. This claim can be opposed with data on the "national, or ethnic affiliation" of the Montenegrin population according to the 2011 census. Of the total population (620,029), 44.98% are Montenegrins, 28.73% Serbs, 8.65% Bosniaks, 4.91% Albanians, 3.31% Muslims, 1.01% Roma, 0.97% Croats, 0.15% Macedonians, 0.15% Russians, and 0.07% Bosnians. Interestingly, some declared themselves as Yugoslavs (0.19%), but there were also people declared as Serbs–Montenegrins, Montenegrins–Serbs, Muslims–Bosniaks, Bosniaks–Muslims, Muslims–Montenegrins, and Montenegrins–Muslims.

4. Lexemes *Albanian* and *Bosniak* in the DMNLL

In societies with a diverse ethnic/national mosaic, the reality is heavily burdened with collective feelings. What best confirms this is the question addressed to the MASA after the publication of the first volume of the DMNLL — "Does MASA advocate ethnically pure states when it defines ethnonyms according to the state they allegedly come from?"¹² (Čirgić 2016: 788). It could be concluded that such a question resulted from the lexicographic definitions of the entries

*Albanian, Austrian, Bosnian, Bosniak, Bulgarian*¹³ and their feminine forms in Montenegrin. However, in the case of the DMNLL, both professionals and laymen problematized only the definition of the entries *Albanian* and *Bosniak*.

The lexeme *Albanian* in the DMNLL is defined as "an inhabitant of Albania; one having their origin from Albania" (see Appendix A). The same type of definition is also used for the forms such as *Bosnian, Bulgarian, etc.*

When it comes to the lexeme *Bosniak*, two meanings are identified: "1. A member of the Bosniak people, a Muslim who lives in Bosnia or one having their origin from Bosnia 2. A member of the South Slavic people who are followers of Islam, mainly residing in Bosnia" (ibid.) (see Appendix B).

The gender-specific word pairs are defined in the same way: Albanac – Albanka (Albanian), Bošnjak – Bošnjakinja (Bosniak) or Bosanac – Bosanka (Bosnian), and Bugarin – Bugarka (Bulgarian), etc.

Unlike *Bosniak*, the entry *Albanian* is defined as a demonym in the DMNL Dictionary (both categories — inhabitants and origin imply geographic affiliation). However, the same dictionary does not define the lexeme *Albanian* as an ethnonym. It is evident that the lexicographic content in question lacks a part which identifies the Albanian people as a cultural-historical category or a part by which nationality is primarily determined on the basis of perceived origin — on the so-called spiritual level, i.e. as a member of something "which is only a concept in linguistics" (Vuković 2007: 174). In that sense, it can be concluded that the lexicographic description is not complete. On the other hand, the noun *Albanian* as a derivative representing demonym (based on its constituent elements) has a proper definition from a linguistic perspective — this is confirmed by the ways in which word formation is treated in the grammatical literature of the BCMS language. In that regard, it is worth mentioning that the lexeme *Albanian* is, for example, defined in the DMNLL in the same way as in *Fjalor i shqipes së sotme* (Dictionary of Modern Albanian) of the Albanian Academy of Sciences and Arts (2002). Although perfectly aware of that, Albanian politicians in Montenegro remained resolute that the DMNLL definition was discriminatory.

Political representatives of the Albanian and Bosniak people in Montenegro, but also some linguists, indicated strong disapproval of the definitions of *Albanian* and *Bosniak* in the MASA Dictionary because they did not recognize these two peoples as autochthonous constituents of Montenegrin society, and according to Albanian politicians, the definition of *Albanian* lacked yet another important segment — the fact that "Albania is the mother country of all Albanians, wherever they live" (Cafe del Montenegro 2016). After the publication of the DMNLL, political interpretations of the problematic definitions emphasized the fact that these are minority peoples who represent an autochthonous part of Montenegrin society, but that Montenegro is not their mother country. Regardless of the political and legal scope of the concept *mother country*, it is certain, however, that the linguistic scope of defining an *Albanian* as "a person of Albanian origin" does not exclude the sense of belonging of Albanians born

in Montenegro to Albania as their home country. Exactly the same feeling is present in the definition in the *Dictionary of Modern Albanian*. In this regard, it is important to understand the lexeme *origin*: its semantic content implies *belonging or background of a person acquired by birth – belonging to a family, nation, state, class, etc.* In this case, it would imply belonging to a country acquired by ethnicity, not by place of birth.

Unlike the noun *Albanian*, as we have shown, the entry *Bosniak* in the DMNLL is primarily a designation of ethnic-religious identity; its "spiritual" dimension is partially visible in the geographical designation, and its semantic content is interpreted similarly in *Rječnik bosanskoga jezika*¹⁴ (Halilović, Palić and Šehović 2010). Thus, the DMNLL does not primarily link the lexeme *Bosniak* with the meaning derived from its form, based on which Jahić (2020) sees it as an ethnic designation semantically identical with the form *Bosanac* (*Bosnian*) and *Bošnjakin*, *Bošnjanka* ("formed from the root *Bos-* and the South Slavic suffix *-(n)jak*").

In relation to Jahić's interpretation, it is now interesting to look at the fact that in *Rečnik srpskohrvatskoga književnog jezika* (*Dictionary of the Serbo-Croatian Literary Language*) (1967–1976), a synonymous definition is given for the entry *Bošnjak* (*Bosniak*) – "a. *Bosanac*" (*Bosnian*), whereby the lexeme *Bosanac* (*Bosnian*) is explained as "a person from Bosnia". In fact, according to the lexicographic interpretation of the time, these two lexical units designated spatial affiliation to the "province of Bosnia" instead of ethnic/national affiliation (that is why they are offered in the singular form, unlike other peoples of Yugoslavia – *Macedonians*, *Slovenes*, *Serbs*, *Croats* and *Montenegrins*, who were represented in the plural form¹⁵). The modern perspective sheds light on the "socio-historical principle" (Ristić 2006: 97), which is why the same dictionary marks the secondary meaning of the noun *Bosniak* — "a Muslim from Bosnia" as "obsolete" — therefore, a component of the current semantic content of the lexeme *Bosniak* (see *Bošnjak* in the *Rečnik srpskoga jezika* 2007) was *out of date* (*no longer in use*) in the seventies of the XX century.

Although renominations and lexical redefinitions reflect the functioning of language in a certain social environment, a linguistic fact cannot be ignored in this particular case — the country of Bosnia makes "the formal and semantic basis" of the name *Bosniak* (Jahić 2020). Lexical description of the entry *Bosniak* should thus be based on it, regardless of the demographic profile of a certain area.

Concerning the lexical definition of the lexeme *Bosniak* in the DMNLL, we have linguistic responsibility to pay attention to the elements that were ignored by the critics of the *Dictionary*: the semantic components reflecting the inclusion or hyponymic relationship are represented as different — the meaning of a *Bosniak* as a "member of the Bosniak people" is singled out as the primary and special one compared to the meaning expressing a *Bosniak* as a "member of the South Slavic people ...". In doing so, the principle of a circular definition was applied: the entry *Bosniak* is defined with the relative adjective *bošnjački* (*Bosniak*) which has a grammatical definition ("relating to *Bosniaks*") and whose

lexical meaning should be revealed by the definition of the basic word — *Bosniak*. Circular lexicographic definitions are, however, also present in the analysis of this type of entry in the other contemporary lexicographical editions. For example, *Rječnik bosanskog jezika* (*Dictionary of the Bosnian Language*) (2007), defines the lexeme *Bošnjak* (*Bosniak*) by the adjective *bošnjački* (*Bosniak*) – as a "member of the Bosniak nation" (cf.: adj. *Bosniak* – "relating to Bosniaks"). Although the specific definition does not convey the specifics of the Bosniak people, the ethnic/national feelings of their members were not hurt — or the reactions were not recognized.

4.1 Designations of national and territorial affiliation and their political-legal contextualization in Montenegro

We believe that the criticism of the given lexicographic definitions needs to be contextualized with the following data.

Unlike the *Constitution* from 1992, which designates Montenegro as the national state of the Montenegrin people, the *Constitution* from 2007 (the latest one) defines Montenegro as a civic state, in which the bearer of sovereignty is the "citizen with Montenegrin citizenship", and the preamble of the *Constitution* stipulates that "Montenegro is inhabited by Montenegrins, Serbs, Bosniaks, Albanians, Muslims, Croats and the others who are committed to democratic and civic Montenegro"¹⁶. The current constitution, therefore, does not define Montenegro as a national state of Montenegrins (not even in the preamble of the above-mentioned nations); it is the "multinational state" or "national neutrality" (Stojanović 2021: 94) that is set as one of its basic principles. Of course, all residents of Montenegro with Montenegrin citizenship are Montenegrins: even if they are not part of the Montenegrin people, they are Montenegrins according to their civic identity — and their demonym is *Montenegrins*. In other words, all Montenegrins as citizens of Montenegro do not belong to the majority of Montenegrin people. Namely, Serbs, Bosniaks, Albanians, Muslims, Croats, etc. are members of ethnic groups in Montenegro and are united with the Montenegrin people by the demonym *Montenegrins*. However, as shown by the interpretations given within the 2011 population census, the demonym *Montenegrin* also belongs to foreign citizens and stateless persons with permanent or temporary residence in Montenegro (Monstat 2011). Even more broadly, the demonym *Montenegrin* also belongs to the foreign citizens who live outside of Montenegro but who have obtained Montenegrin citizenship (eg, so-called economic citizenship obtained by investing money in the local economy). This actually suggests that the contemporary meaning of demonyms (designating territorial affiliation) is actually broader in comparison to its constituent elements that we first presented. The demonym *Montenegrin*, for example, does not only designate a person who lives in Montenegro (with or without Montenegrin citizenship) and one having their origin from Montenegro — it also

designates an individual with Montenegrin citizenship who does not live in Montenegro.

In view of the consequences that followed after the publication of the first volume of the DMNLL (2016), and for the sake of the future lexicographic activities in Montenegro, a question needs to be addressed — does the concept of Montenegro as a civic state imply that civic identity has priority over ethnic identity in lexicographic descriptions of the designations such as *Montenegrin*, *Serb*, *Bosniak*, *Albanian*, *Muslim*, *Croat*, etc.? The affirmative answer would imply that these names are primarily identified as demonyms — designating the inhabitants of the state territory they are named after (if they are derived from the name of a particular state territory) / designating the individuals who come from that territory / designating citizens of Montenegro — which also implies the harmonization of linguistic and non-linguistic criteria in the arrangement of their components of meaning, and then as ethnonyms — as names of peoples specified linguistically (language family) and geographically (home country). With such a model, minority ethnic communities would (in line with their expectations after the publication of the DMNLL) be identified both according to the domicile and home country. In that sense, however, Muslims and Serbs in Montenegro should be given special lexicographic attention. For Montenegrin Muslims (unlike Albanians, Bosniaks and Croats), Montenegro is their "only homeland" and "their mother tongue is Montenegrin" (Tomović 2011). On the other hand, Serbs in Montenegro are divided when it comes to their home state (Serbia/Montenegro), which is yet another confirmation that the ethnic peculiarities of the Balkans are difficult to include in lexicographic models of national and territorial affiliation.

Concerning the ethnic image of Montenegro, we need to add that the last census (2011) in Montenegro showed that 427 persons (0.07% of the population) identified themselves as *Bosanci* (*Bosnians*). In theory, however, this particular name should not be identified with national identity as it belongs to the political and legal framework and as such exclusively refers to the identity of the citizens of Bosnia and Herzegovina. This type of divergence represents precisely one of the issues and topics within the sphere of social sciences, and lexicographic descriptions depend on the interpretations in this field. On the other hand, the political-legal simplification of the name *Bosnian* as a designation of exclusively civic identity is contrary to philological or linguistic interpretations which emphasize that in linguistic terms "there is no difference" between the designations *Bošnjak* (*Bosniak*) and *Bosanac* (*Bosnian*), or *Bošnjakin*, *Bošnjak*, *Bošnjak* and *Bosanac* as they all "stem from *Bos-*, designating the specific Bosnian space, country and its historical, cultural and linguistic being" (Jahić 2020). In light of this divergence, lexicography faces an important question — how to overcome the disparity of linguistic and political-legal interpretation of certain linguistic designations, as well as the fact that collective identity feelings in the Balkans cannot always fit into the theoretical framework of the social sciences. While answers to this type of question require interdisciplinary engagement, lexico-

graphic definitions of ethnic/national identity designations (based on the meaning of their constituent elements) presuppose both political and legal basis and demarcation of certain conceptual categories.

5. Conclusion

Although the problematic definitions from the DMNLL are more or less standard in dictionaries of other languages of the Balkans and monolingual dictionaries of world languages, it is obvious that today's situation in the Balkans imposes their reconsideration and redefinition. The DMNLL has shown that Balkan societies are "still heavily burdened by collective feelings" (Vukčević 2015: 20), which is why the lexical description of the names of Balkan peoples and the names of the inhabitants of Balkan countries is a very sensitive issue that requires careful examination by experts from various fields. After the publication of the DMNLL, dictionary ethnic and civic identifications were not within the framework of objective scientific postulates and lexicographic principles and were generally "defined" by the media coverage of politicians who articulated the collective feelings of their voters. Nevertheless, Montenegrin lexicography is confronted with specific expectations of unsatisfied ethnic/national groups — their collective identity feelings are based on their affiliation to both the "domicile" and "mother country".

Aware of the complex identity issue that stopped the lexicographic endeavors of the highest-ranking scientific institution in Montenegro, this article has tried to present the issues that exceed linguistic competencies and are important for the establishment of a proper model for the lexicographic definitions of the identity designations — a model that would reflect what is known as an ideal type of identity in the sociological literature (the identity that includes ethnic-cultural and territorial-political/civic elements). Certainly, such a model of the definition implies a difference between categories *a member of a people* and *an inhabitant of a country*, but our study has shown that it should also include the category of *a citizen of a country* (as citizenship is a legal relationship that is not exclusively conditioned by the place/country of residence). Along with the precise and consistent linguistic demarcation of the terms *ethnonym* and *demonym*, the definition should also be expanded to include demonyms as a derivative motivated by the name of a country (a person who belongs to a particular country as a place of residence, origin, or citizenship — regardless of the place of residence). Apart from raising important questions and offering insights for further research, this article has also demonstrated that linguistics needs to be complemented by other disciplines when it comes to the lexicographic descriptions of the Balkan names and the priority of their semantic components. It is only the interdisciplinary engagement that can, therefore, solve issues of this kind and prevent any potential negative reactions to future lexical descriptions.

Endnotes

1. [https://www.monstat.org/userfiles/file/popis2011/saopstenje/saopstenje\(1\).pdf](https://www.monstat.org/userfiles/file/popis2011/saopstenje/saopstenje(1).pdf), accessed on July 17, 2022.
2. At this point, we need to remind that different social science theories distinguish between two types of nations — *ethnic* (Eastern, German, cultural) and *civic* (Western, French, political, state) models of the nation. According to this classification, the geographical area of Central and Eastern Europe is characterized by an ethnic model — a nation is an *ethnos*, representing a community with a shared cultural heritage, ancestry, history, and tradition. Unlike the civic model (Western Europe, the USA), where the concept of nation is equated with the state and its territory (nationality = citizenship), the ethnic model is centered around the national community: "the state also includes national minorities that do not constitute a nation", which is why "terms *nation* and *state* are not synonymous" (Nedeljković 2007: 19; cf. Tamir 2019; see also Bugarski 2018: 16). However, in addition to ethnic and civic models of the nation, some theoreticians also recognize "plural" type of nation/nationalism (Smith 2003: 215).
3. Instead of the term *ethnic*, here we use the term *demonym* because the English noun *ethnic* refers only to a member/person belonging to an ethnic group, not also to the residents of a particular territory (or a state).
4. For the dichotomy "ime – naziv" in BCMS, here we use the dichotomy "name – appellative" (which is grammatically analogous to the dichotomy between proper nouns and common nouns/appellatives).
5. The semantic content of the terms *etnonim* and *etnik* in the BCMS language is expressed in English with the forms *ethnonym* and *demonym* (Merriam-Webster and the OED defines both terms, *Cambridge Dictionary* defines only the term *demonym*. Cf. in French: *etnique* ("3. linguistique"): *Dictionnaire de L'Académie Française*.
6. Ethnonyms are given in their plural form (plural nominative) in this *Dictionary* (discussed further in the following text). However, the names of the Balkan peoples as separate entries are selectively represented: e.g. *Croats*, *Slovenes*, and *Serbs*, for example, have the status of an entry, but that is not the case with *Montenegrins*, *Bosniaks*, *Macedonians*, *Albanians*, *Greeks*, *Bulgarians* ...
7. The entry *srpski* (Serbian) includes the masculine and feminine forms *Srbin*, *Srpkinja* (Serb), but the official name of the state associated with the demonym is missing. Based on the qualification of the forms *Srbin*, *Srpkinja* (Serb) as a demonym, it can be assumed they are semantically equivalent to the demonyms *Srbijanac*, *Srbijanka* (Serb) — although such semantic equivalence is not specified (cf. *srpski* with *srbijanski* and *hrvatski*).
8. In comparison to other similar publications, only this grammar book defines demonyms and as words designating people according to their origin.
9. For example: *Slovenici* (Slovenians) "ethn South Slavic peoples mainly living in today's Slovenia ..."
10. For example: *Slovenac* (Slovenian) ... "a member of the Western group of South Slavic peoples mainly living in today's Slovenia".
11. The form *or*, in this case, has the meaning of an explicative word — *to be precise, namely* ... (some linguists believe that usage is wrong; cf. Klajn 2009: 183 and *Dictionary of the Serbian Language* 2007).

12. Although the question was signed by a linguist, we cannot ignore its imprecision — the definitions here do not identify the origin/etymology of ethnonyms; instead, they identify the origin and/or civic affiliation of the bearers of the concrete names. It is also obvious that the terminological, semantic and lexicographic value of demonyms is neglected by the question.
13. Although the linguistic difference between ethnonyms and demonyms was ignored in the question addressed to the MASA, we have left out the determinants *American, African, Australian, Balkan* (represented in Volume I of the DMNLL because they are lexemes whose basic meaning exclusively designates territorial affiliation).
14. In the Bosnian dictionary, this part of the lexical description is marked with the abbreviation *etn.* — according to the *Dictionary List of Abbreviations*, *etn.* refers to the noun *etnik* (BCMS word for a *demonym*) or adjective *etnički* (*ethnic* in BCMS). The same dictionary does not define Bosniaks as members of the Islamic religion: "2. *etn.* A member of the South Slavic people, mostly settled in Bosnia and Herzegovina."
15. Muslims received the status of one of the constituent nations in the Socialist Federal Republic of Yugoslavia in 1971, although they already had an entry (singular form) designating a member of a people in the III volume of the same dictionary in 1969.
16. Clearly, the Roma are not recognized in the preamble (1.01% of the total population).

References

Dictionaries and encyclopedias

- Akademia e Shkencave e Shqipërisë.** 2002. *Fjalor i shqipes së sotme*. Tiranë: Instituti i Gjuhësisë dhe i Letërsisë, Botimet Toena.
- Cambridge Dictionary.** 2022. <https://dictionary.cambridge.org/>. Accessed on July 17, 2022.
- Dictionnaire de L'Académie Française.** 2022. <https://www.dictionnaire-academie.fr/>. Accessed on July 17, 2022.
- Merriam-Webster.** 2022. <https://www.merriam-webster.com/>. Accessed on July 17, 2022.
- Domović, Ž., Š. Anić and N. Klaić.** 2001. *Rječnik stranih riječi*. Belgrade: Evro.
- Halilović, S., I. Palić and A. Šehović.** 2010. *Rječnik bosanskoga jezika*. Sarajevo: Filozofski fakultet u Sarajevu.
- Hrvatska enciklopedija, mrežno izdanje.** <http://www.enciklopedija.hr/>. Accessed on July 17, 2022.
- Institut za jezik.** 2007. *Rječnik bosanskog jezika*. Sarajevo: Institut za jezik.
- Jahić, Dž.** 2010. *Rječnik bosanskog jezika 2 (D–F)*. Sarajevo: Bošnjačka asocijacija.
- Klajn, I.** 2009. *Rečnik jezičkih nedoumica*. Novi Sad: Prometej.
- Klajn, I. and M. Šipka.** 2007. *Veliki rečnik stranih reči i izraza*. Novi Sad: Prometej.
- Matica srpska, Matica hrvatska.** 1967–1976. *Rečnik srpskohrvatskoga književnog jezika 1–6*. Novi Sad: Matica srpska (1–3, Zagreb: Matica hrvatska).
- Matica srpska.** 2007. *Rečnik srpskoga jezika*. Novi Sad: Matica srpska.
- Montenegrin Association of Sciences and Arts.** 2016. *Rječnik crnogorskog narodnog i književnog jezika (Montenegrin Dictionary of Vernacular and Literary Language)*. Podgorica: Montenegrin Association of Sciences and Arts.
- Oxford English Dictionary.** <https://www.oed.com/>. Accessed on July 17, 2022.

- Prosveta. 1970. *Opšta enciklopedija: Mala enciklopedija Prosveta 1 A–Lj*. Belgrade: Prosveta.
- Serbian Association of Sciences and Arts. 1959–(2020). *Rečnik srpskohrvatskog književnog i narodnog jezika 1–(21)*. Belgrade: SANU.
- Školska knjiga. 2015. *Veliki rječnik hrvatskoga standardnog jezika*. Zagreb: Školska knjiga.
- Vujaklija, M. 1996/97. *Leksikon stranih reči i izraza*. Belgrade: Prosveta.

Other literature

- Barić, E., M. Lončarić, D. Malić, S. Pavešić, M. Peti, V. Zečević and M. Znika. 2005. *Hrvatska gramatika*. Zagreb: Školska knjiga.
- Bugarški, R. 2018. *Govorite li zajednički?* Belgrade: Biblioteka XX vek.
- Cafe del Montenegro. 2016. Đeljošaj traži da CANU hitno povuče Rječnik crnogorskog jezika. 17 May. Available: <https://www.cdm.me/drustvo/deljoshaj-trazi-da-canu-hitno-povuce-rjecnik-crnogorskog-jezika/>. Accessed on July 17, 2022.
- Čirgić, A. 2016. Ova bruka osta do vijeka. *Matica* 66: 784-789.
- Čirgić, A., I. Pranjković and J. Silić. 2010. *Gramatika crnogorskoga jezika*. Podgorica: Ministarstvo prosvjete i nauke.
- Čorić, B. 2008. *Tvorba imenica u srpskom jeziku*. Belgrade: Društvo za srpski jezik i književnost Srbije.
- Jahić, Dž. 2020. Bošnjani, Bošnjak, Bosanac su sinonimski nazivi za matični narod Bosne. *INS*, December 3, 2020. <https://www.ins.ba/bs/article/11922/dzevad-jahic-bosnjanin-bosnjak-bosanac-susinonimski-nazivi-za-maticni-narod-bosne>. Accessed on July 17, 2022.
- Jahić, Dž., S. Halilović and I. Palić. 2000. *Gramatika bosanskoga jezika*. Zenica: Dom štampe.
- Korunić, P. 2003. Nacija i nacionalni identitet: uz porijeklo i integraciju hrvatske nacije. *Zgodovinski časopis* 57(1–2): 163-208.
- Lakićević, M. 2018. Interview Momčilo Isić: Miloš Moskovljević — The Author of the Dictionary whose Word Was Taken Away. *Novi magazin*, 27 February 2018. Available: <https://novimagazin.rs/vesti/169307-intervju-momcilo-isic-milos-moskovljevi---autor-renika-kome-je-oduzeta-re>. Accessed on July 17, 2022.
- Monstat (Montenegro Statistical Office). 2011. *Census of Population of Population, Households and Dwellings in Montenegro 2011*. Available: [https://www.monstat.org/userfiles/file/popis2011/saopstenje/saopstenje\(1\).pdf](https://www.monstat.org/userfiles/file/popis2011/saopstenje/saopstenje(1).pdf). Accessed July 17, 2022.
- Nedeljković, S. 2007. *Čast, krv i suze — ogledi iz antropologije etniciteta i nacionalizma (Honor, Blood and Tears — Experiments from Anthropology of Ethnicity and Nationalism)*. Belgrade: Odeljenje za etnologiju i antropologiju Filozofskog fakulteta.
- Parliament of Montenegro. 2007. *Constitution*. Available: <https://www.skupstina.me/me/ustav-crne-gore>.
- Peti, M. 1997. Koje su imenske riječi etnici i etnonimi? *Folia onomastica Croatica* 6: 99-112.
- Piper, P. and I. Klajn. 2013. *Normativna gramatika srpskog jezika*. Novi Sad: Matica srpska.
- Ristić, S. 2006. *Raslojenost leksike srpskog jezika i leksička norma*. Belgrade: Institut za srpski jezik SANU.
- Smith, D.A. 2003. *Nacionalizam i modernizam: Kritički pregled suvremenih teorija nacija i nacionalizma*. Prevela: Mirjana Paić Jurinić. Zagreb: Fakultet političkih znanosti.
- Sretenović, M. 2008. Sudbina jednog rečnika. *Politika*, 22 January 2008. Available: <https://www.politika.rs/sr/clanak/11666/Sudbina-jednog-recnika>. Accessed on July 17, 2022.

- Stojanović, M.** 2021. Neobičan slučaj kontinuiteta državnih institucija Crne Gore. *Studia iuridica Montenegrina* 3(1): 75-99.
- Šimunović, P.** 2009. *Uvod u hrvatsko imenoslovlje*. Zagreb: Golden Marketing — Tehnička knjiga.
- Šipka, D.** 2006. *Osnovi leksikologije i srodnih disciplina*. Novi Sad: Matica srpska.
- Šubarić, S.** 2018. Rod u gramatičkom opisu imenica (prilog stabilnosti kategorije imeničkog roda). *Folia linguistica et litteraria — časopis za nauku o jeziku i književnosti* 20: 21-44.
- Tamir, Y.** 2019. Not So Civic: Is There a Difference Between Ethnic and Civic Nationalism? *Annual Review of Political Science* 22: 419-434.
- Tomović, P.** 2011. Popis u Crnoj Gori dijeli i Bošnjake i Muslimane. *Radio Slobodna Europa*, 21 January 2011. Available: https://www.slobodnaevropa.org/a/crna_gora_popis_stavnovnistva_bosnjaci_muslimani/2283323.html. Accessed on July 17, 2022.
- Vukčević, D.K.** 2015. Tri misli o identitetu (Three Thoughts on Identity). Vukčević, D.K. 2015. *O identitetu*: 13-23. Podgorica: CANU.
- Vuković, S.** 2007. Onomastička terminologija. *Čakavska rič* XXXV(1): 139-184.

Appendix A: Dictionary scanned text — lexeme *Albanac* (*Albanian*)

Албанац м. [іен. јг. Алб́анца, вок. јг. Алб́анче, инсѝр. јг. Алб́анцем, іен. мн. Алб́анѝца] сѝановник Албаније; онај који је ѝоријеклом из Албаније.

албаниза́ција ж. [іен. јг. -ѝ] намеѝање албанској језика, кулѝуре и обичаја друѝим народима. — Питајуѝи се нудио је и неке одговоре — на основу искустава издѝеглица православног живља из Албаније које је било изложено систематској албанизацији, нарочито у времену послѝје II свјетског рата 1945. године. (Лак. II)

албанизи́рање с. іл. им. од албанизираѝи.

албанизи́рати сврш. и несврш. [ѝрез. албанѝзѝрам, -ајѝ/ албанѝзѝрајѝ, ѝрил. сад. албанизи́рајѝѝи] в. албанизовайѝи. — Наиме, иако су Малисори албанизирани, ѝихов говор се знатно разликује од стандардног албанског ... (Ник. В. II)

албанизовѝње (албанѝзовѝње, изг.) с. іл. им. од албанизовайѝи.

албанизовати (албанѝзовѝати, изг.) сврш. и несврш. [ѝрез. -ујѝм] I ѝреобраѝиѝи/ѝреобраѝаѝи у Албанце; намеѝнуѝи/намеѝаѝи албански језик, кулѝуру и обичаје друѝим народима. II ~ се ѝосѝа(ја)ѝи Албанац; доби(ја)ѝи обиљежје Албанца.

Албанија ж. [іен. јг. -ѝ] држава у заѝадном дијелу Балканској ѝолуосѝрва.

Албѝнка ж. [іен. јг. -ѝ, даѝи. јг. -ки] сѝановница Албаније; она која је ѝоријеклом из Албаније.

албанолог м. (+ ж.) [вок. јг. албаноложе, ном. мн. албанолози] сѝручњак за албанологију. (Даш. II)

албанологија (албанологија, изг.) ж. [іен. јг. -ѝ] наука која се дави изучавањем албанској језика, књижевносѝи и кулѝуре. — Чак 29 универзитетских професора и научника из Албаније држало је предавања на постдипломским студијама Приштинског универзитета из албанологије и историје. (Лак. I)

Appendix B: Dictionary scanned text — lexeme *Bošnjak* (*Bosniak*)

Бòшњāк м. [іен. јд. Бошњāка, вок. јд. Бòшњāче, ном. мн. Бошњāци] **1.** ѝриѝадник бошњачкої народа, муслиман који живи у Босни или је ѝоријеклом из Босне. **2.** ѝриѝадник јужнословенскої народа исламске вјероисѝовијесѝи, уїлавном насѝањен у Босни.

Бошњāкиња ж. [іен. јд. -ѝ] **1.** ѝриѝадница бошњачкої народа, муслиманка која живи у Босни или је ѝоријеклом из Босне. **2.** ѝриѝадница јужнословенскої народа исламске вјероисѝовијесѝи, уїлавном насѝањена у Босни.

Бòшњāчкѝ, -ā, -ō који се односи на Бошњаке: ~ кулѝура.

KIU (Kiswahili–Italiano–UniOr): The UniOr online Dictionary for Italian L1 Swahili Learners

Flavia Aiello, *Department of Asian, African and Mediterranean Studies,
University of Naples 'L'Orientale', Italy (faiello@unior.it)*

and

Rosanna Tramutoli, *Department of Asian, African and Mediterranean
Studies, University of Naples 'L'Orientale', Italy
(rosanna.tramutoli@gmail.com)*

Abstract: KIU is an online bilingual Swahili–Italian dictionary with about 6000 entries aimed primarily at Italian L1 Swahili learners and which has been developed at the University of Naples 'L'Orientale'.

The project was started in 2003 by M. Toscano and developed with the collaboration of language experts and young researchers until 2009 with the aim of offering online lexical resources from Swahili to Italian for learners of the language. After a long interruption, the work was resumed in 2019 by the authors of this article in cooperation with M. Toscano and a team of expert IT technicians. The current work consists of the development ex novo of the dictionary software, which had become obsolete, along with a redesign of some lexicographic features.

In this report we will show how the upgraded version of the dictionary software has been implemented, with relevant learner-oriented features, by taking into consideration the standard lexicographic characteristics of Swahili–Italian bilingual dictionaries. This dictionary represents a valuable support for L2 learners and is the only on-line Swahili–Italian dictionary expressly built for university students and Italian users at large.

Keywords: SWAHILI, ITALIAN, ONLINE DICTIONARY, LEARNERS, LEXICOGRAPHY

Opsomming: KIU (Kiswahili–Italiano–UniOr): Die UniOr aanlyn woordeboek vir Italiaanse L1-Swahili-aanleerders. KIU is 'n aanlyn tweetalige Swahili–Italiaanse woordeboek met ongeveer 6000 inskrywings wat hoofsaaklik gerig is op Italiaanse L1-Swahili-aanleerders en wat deur die Universiteit van Napels 'L'Orientale' ontwikkel is.

Die projek is in 2003 deur M. Toscano van stapel gestuur en is met die samewerking van taalkundiges en jong navorsers tot 2009 ontwikkel met die doel om aanlyn leksikale hulpbronne uit Swahili na Italiaans vir aanleerders van die taal aan te bied. Ná 'n lang onderbreking is die werk in 2019 deur die outeurs van hierdie artikel in samewerking met M. Toscano en 'n span kundige IT-tegnici hervat. Die werk bestaan tans uit die ex novo-ontwikkeling van die woordeboeksagteware wat verouderd geraak het asook uit 'n herontwerp van sommige leksikografiese kenmerke.

In hierdie verslag sal ons aantoon hoe die opgegradeerde weergawe van die woordeboeksagteware, met inagneming van die standaard leksikografiese eienskappe van Swahili–Italiaanse tweetalige woordeboeke, geïmplementeer is met relevante aanleerder-georiënteerde kenmerke. Die

woordeboek verleen waardevolle steun aan L2-aanleerders en is die enigste aanlyn Swahili-Italiaanse woordeboek wat spesifiek vir universiteitstudiante en vir Italiaanse gebruikers oor die algemeen geskep is.

Sleutelwoorde: SWAHILI, ITALIAANS, AANLYN WOORDEBOEK, AANLEERDERS, LEKSIKOGRAFIE

1. Overview of the existing lexicographic resources for Italian L1 Swahili learners

Swahili lexicography, including bilingual dictionaries, began in the last decades of the 19th century, and during colonial times was practised by foreigners who came to East Africa either as missionaries or colonial administrators (Chuwa 1996: 324).¹

After independence was gained by East African states, Swahili lexicography continued to grow in a context of development of Swahili in a wide geocultural area and in multiple contexts of language use. In Tanzania this was especially so. There, the language became the dominant medium of communication in the public sector (basic education, administration, courts, national assembly, media etc.) thanks to the commitment of experts from the National Swahili Council (Baraza la Kiswahili la Taifa, BAKITA) and the Institute of Swahili at the University of Dar es Salaam (Taasisi ya Uchunguzi wa Kiswahili², TUKI). These institutions, besides producing general dictionaries, monolingual and bilingual (SWA-ENG and ENG-SWA), created and disseminated terminologies for many domains of language use, such as law, science, and IT (Sewangi 2007; Aiello and Toscano 2017). Bilingual dictionaries were also produced over the years in countries where Swahili is taught at university level (such as France, Germany, Russia, etc.).³ Throughout the history of Swahili lexicography, most new dictionaries were based on their predecessors. From the 2000s, however, innovative corpus-based dictionaries also started to appear (Wójtowicz 2016: 402), such as the Swahili-Finnish dictionary (Abdulla et al. 2002) and the Swahili-Polish dictionary by Wójtowicz (2003), which is available also online (see Wójtowicz and Bański 2012 for more details).

Since the 1990s, indeed, several bilingual electronic dictionaries for Swahili L2 learners have been developed, the majority of them targeting English speakers. A community-based SWA-ENG online dictionary was started in 1994 at Yale University under the name of Internet Living Swahili Dictionary, later run by the association Kamusi Project International with the aim of incorporating other languages (www.kamusi.org). The SWA-ENG and ENG-SWA dictionaries produced by TUKI in print form are also accessible online as basic databases listing the entries by alphabetical order via the links <https://swahili-dictionary.com/swahili-english> and <https://swahili-dictionary.com/english-swahili> respectively. The TshwaneDJe Swahili-English Dictionary (<https://tshwanedje.com/dictionary/swahili/>) is a corpus-driven dictionary of Swahili, downloadable upon purchase. It is based on web-based corpus data, with over 16,000 entries

and phrases and over 36,000 translation equivalents. Presently, the most popular SWA-ENG / ENG-SWA dictionary available for free online is the *Glosbe — Swahili-English dictionary* (<https://it.glosbe.com/sw/eng>) which is part of the Glosbe multilingual dictionary (6000 languages) created by a Polish team and which offers translations created by users or automatically generated on the basis of a database of translated examples.

All these numerous lexicographical works, both printed and online, can be used by the students of Swahili L2 of the University of Naples 'L'Orientale' (henceforth UniOr), who can find many bilingual dictionaries in our library and have free Wi-Fi access to the Internet. Many Italian students, though, find it difficult to use a bilingual dictionary which is not in their native language for their exercises (drills, comprehension tests, translations, oral production, etc.), especially during the first years of their studies when they do not have a perfect mastery of English or other foreign languages. Also, advanced students of Swahili have a great need of resources in Italian as the translation of literary texts and other specialist writings (for instance essays or newspaper articles on politics) is one of the focusses of the teaching of Swahili at UniOr. This is the way it has been since the teaching was established in 1969 by E. Bertoncini Zúbková, an internationally renowned scholar of Swahili language and literature, whose educational activity was closely connected with research on the Swahili lexicon, resulting in the production of lists of words and vocabularies. This huge work has remained largely unpublished, apart from a small vocabulary (*Vocabolario swahili-italiano e italiano-swahili*, Opera Universitaria I.U.O., Naples 1977), which is now out of print.

Concerning bilingual lexicographical works aimed at Italian learners, three dictionaries have been published so far. These are Vittorio Merlo Pick's *Vocabolario kiswahili-italiano e italiano-kiswahili* (EMI, Turin 1961, re-edited in 1978, currently out of print), Maddalena Toscano's pocket-size *Dizionario swahili. Swahili-italiano, italiano-swahili* (Vallardi, Milan 2004) and Gianluigi Martini's *Dizionario swahili. Swahili-italiano, italiano-swahili* (Hoepli, Milan 2016). Furthermore, a terminological work, namely a Swahili-Italian linguistic glossary by Rosanna Tramutoli (*Kamusi ya isimu Kiswahili-Kiitaliano*, TUKI, Dar es Salaam 2018), was published by the University Press of the Institute of Swahili at the University of Dar es Salaam within the sphere of a longstanding cooperation agreement with UniOr. All these works for Italian speakers are, however, in print format and only partly available in the library of UniOr. With regard to online resources for Italian learners of Swahili, the following dictionaries are those available on the web:

(1) *Glosbe — Dizionario Swahili-italiano* (<https://it.glosbe.com/sw/it>), also IT-SWA, as part of the above-mentioned Glosbe multilingual dictionary, which contains many entries, but is not always reliable. It presents a number of incorrect results, such as in fig. 1, where the second translation of the word *utumishi*, 'battuta' (setback) is completely wrong, as well as the Italian translation of the example.⁴

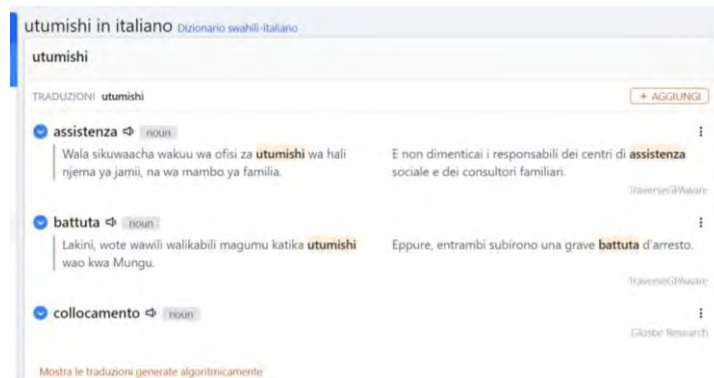


Fig. 1: Incorrect translation in *Glosbe* — *Dizionario Swahili-italiano*

In some instances, when the translation is missing, *Glosbe* proposes some 'hypotheses' by using an algorithm and the user is warned to be cautious. This is the case, for instance, when we look for grammatical forms such as possessives, demonstratives, etc. with their class concord as in fig. 2. Instead of a direct translation of *yako* (second-person possessive *-ako* + class 9 [singular] verbo-pronominal concord *y-*), that is 'tuo, tua' (your, second-person singular possessive referring to a masculine or feminine noun), there is a list of misleading results as the Italian translations are derived from the English ones, and 'your' can be translated into Italian as 'tuo, tua' but also as 'tue, tuoi' (your, second-person singular possessive referring to a masculine or feminine noun in the plural form) and 'vostro, vostra, vostre, vostri' (your, second-person plural possessive referring to a masculine or feminine noun in the singular or plural form).



Fig. 2: Grammatical forms in *Glosbe* — *Dizionario Swahili-italiano*

(2) *Online Swahili Italian Dictionary* (<http://www.etranslator.ro/swahili-italian-online-dictionary.php>) (also IT-SW), which provides automatic translations through English and seems to be even less reliable, as we see in fig. 3 where *yako* (see the explanation above) has been wrongly translated into Italian as 'il vostro' (your/yours, second-person plural possessive referring to a masculine noun in the singular form).

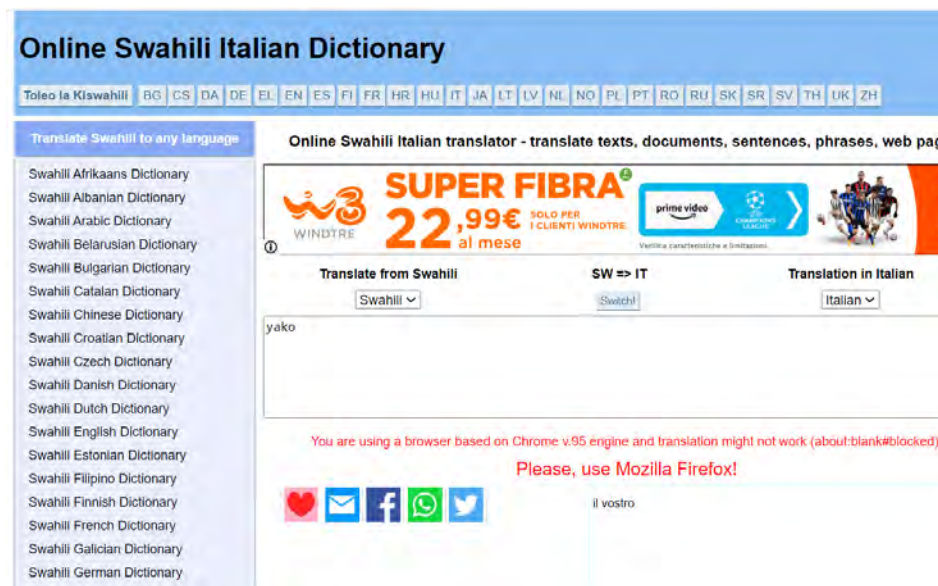


Fig. 3: *Online Swahili Italian Dictionary*

(3) *17minute-languages — Dizionario gratuito online: Italiano–swahili* (<https://www.17-minute-languages.com/it/dizionario-di-swahili/>), which allows the search for SWA–IT and IT–SWA and works as an automatic translator offering a list of the sentences containing (in the Italian translation) the searched word or form. The results are sometimes incorrect or lacking the main, basic translation as, for instance, in fig. 4, showing the result for the word 'amore' (love). Here we don't find the main Swahili translations for 'amore' (*mapenzi*, *mahaba*, *upendo*) but two examples: (1) *barua la* (sic!) *kimapenzi* = la lettera d'amore (love letter), where the Swahili word *mapenzi* is used in a particular way, namely with a ki- prefix with an attributive function; (2) *sikitika* = soffrire le pene d'amore (to suffer the pains of love). Actually, the Swahili verb *sikitika* means 'to be sorry, to be sad', which clearly was translated into Italian as 'soffrire le pene d'amore' in a specific context.

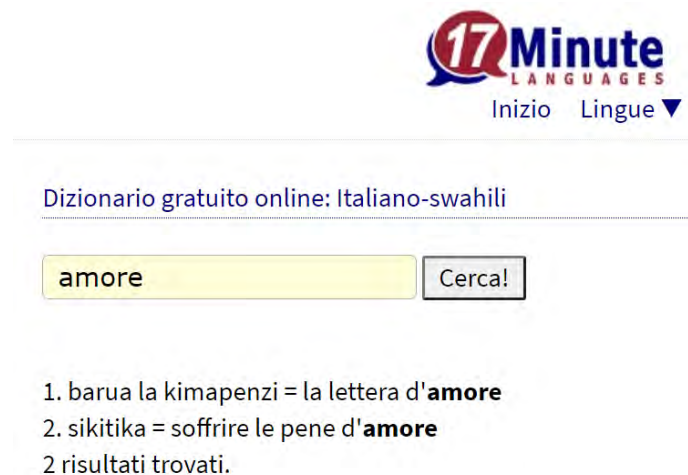


Fig. 4: 17minute-languages — Dizionario gratuito online Italiano–swahili

Similarly, when looking for the Italian translation of *mapenzi* (see fig. 5), the first translation given in the Swahili–Italian printed dictionaries, 'amore', is not given, instead we see two examples: (1) *barua la kimapenzi*, the same expression as commented on above, which, furthermore, is grammatically incorrect, as it presents *barua* as a noun belonging to class 5/6, whereas it belongs to class 9/10; 2) *fanya mapenzi* (to make love).

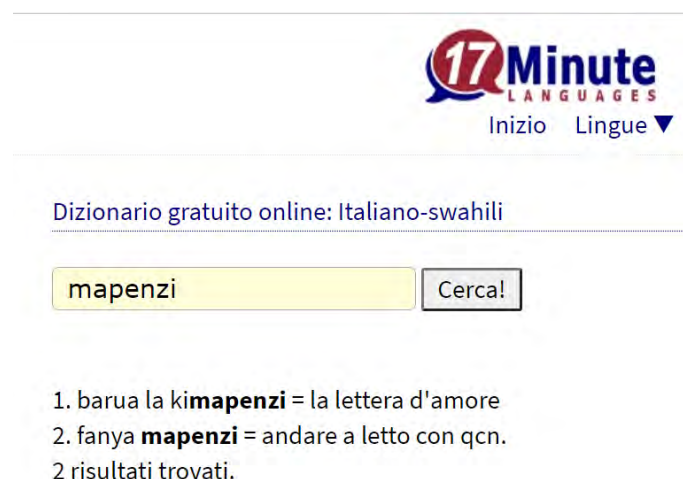


Fig. 5: Incorrect translation of *mapenzi* in Dizionario gratuito online

(4) *Vocabolario Italiano-Swahili*, on the webpage 'Changamano Onlus/Karibu!' by Nino Vessella (<https://swahili.it/glossword/index.php?a=index&d=1>; latest update in 2008). This dictionary is a useful resource, but it only allows the search IT-SWA. The entries are mainly based on the above-mentioned Vittorio Merlo Pick's *Vocabolario kiswahili-italiano e italiano-kiswahili*, for instance, the following entry 'amore' is very similar in the two dictionaries (see fig. 6 and 7):



Fig. 6: 'amore' in *Vocabolario Italiano-Swahili*

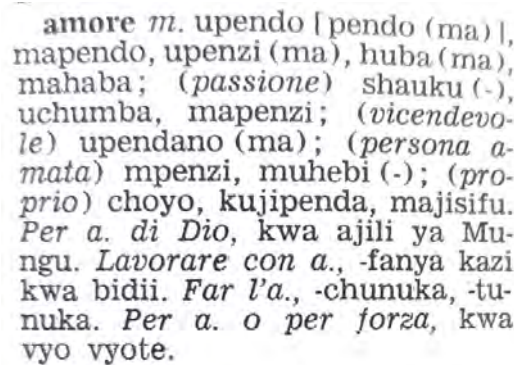


Fig. 7: 'amore' in Vittorio Merlo Pick's IT-SWA dictionary

Here we find also some additions related to new lexicon, for instance AIDS (in Swahili UKIMWI, the acronym of Upungufu wa Kinga Mwilini, literally 'deficit of protection in the body'), though other contemporary vocabulary is missing, like words related to IT (e-mail, Internet, web, etc.).

From this overview, it appears that online lexicographical resources for Italian learners of Swahili are not wholly satisfying, especially with regard to the direction SWA-IT. This reality inspired the idea of developing a lexicographical tool accessible through the Internet in order to help our students in their exercises and translations when they are far from the university. This need for online resources also emerged vividly with the outbreak of the Covid-19 pandemic, since the UniOr libraries were closed, and all the teaching activities were based on digital platforms.

2. KIU: the updated Swahili-Italian online dictionary for learners

The development of an online lexicographical resource for Italian learners (UWAZO) was started at the beginning of the 2000s by M. Toscano in a context of renewal of the courses of Swahili at UniOr and was encouraged by the introduction of new digital technologies, which have revolutionised the teaching of foreign languages. The Swahili-Italian online dictionary UWAZO was based on T.E.I. guidelines⁵ and was developed at UniOr between 2003 and 2009.

The old software used for UWAZO had become obsolete and so in order to meet the increasing need of updated digital learning resources for Swahili students, especially beginners, updated lexicographic software, namely KIU (Kiswahili-Italiano UniOr), has been designed. This work was done from 2020 to 2022 by a team of experts on the basis of the lexicographical indications received by the authors of the present article, in collaboration with M. Toscano. The dictionary project aimed to create updated software that would be developed in accordance with recent lexicographic practices.

The Swahili-Italian dictionary is designed primarily as a didactic tool for Swahili students, but at the same time it is suitable for a wider audience, including people working for Italian NGOs, staff of cultural associations and institutions, embassy staff, tourists, businessmen and anyone interested in Swahili language and culture. While there are very few lexical resources available for Italian L1 Swahili students, this online dictionary will serve as a language-learning support and supplement other teaching materials and paper-dictionaries used in class. The dictionary KIU will be published online on a dedicated website linked to the UniOr website,⁶ and will be accessible for free to university students and the general public.

The website of the lexicographic project, apart from being a dictionary interface, contains an introduction, information on the language, a dictionary user guide, a grammatical sketch, and information on how to quote the lexicographic work. The interface of KIU will be accessible through a general webpage which will include a description of the lexicographic research project, academic publications of the lexicographic team, information about the software, and all contacts and credits of the subjects and institutions involved in the project.⁷

Unlike printed dictionaries, it will be possible to search KIU without limitations, and it is easy to maintain and expand. The database contains about 6,000 headwords. Most of them were selected from a Swahili frequency list (Bertoncini Zúbková 1973) and from various other sources, e.g. Merlo-Pick 1961 and Toscano 2004 (see par. 1 about other Swahili–Italian dictionaries). The sources also include the lexicon used in the teaching materials of Swahili courses by prof. E. Bertoncini Zúbková. Moreover, beside complete sets of inflected forms, closed sets of words such as days of the week, months, and general lexicon, the dictionary also contains some specific vocabulary collected by students and researchers who worked on chosen sets they found useful in their studies like body parts or immigration.

Following T.E.I. guidelines for printed dictionaries, the KIU database structure is based on various groups which include a fixed list of elements with free position. Sub-class elements are also available. The main T.E.I. groups used in KIU include: *Gruppo grammaticale* (Grammatical Group), *Traduzione* (Translation), *Esempio* (Example), *Esempio Tradotto* (Translated Example), *Confronta* (Cross-reference), *DictScrap* (additional notes regarding grammatical indications or specific usage). Also, open lists of labels, to be set by the operator, are possible.

It is freely accessible online by users and supports two different levels of access depending on the role. Administrators have the highest level of access to the database and can implement and edit data, and set up and manage the accounts of students and/or collaborators. Students and learners will have full access to database and software tools except for data publishing which needs the approval of the administrators.

Moreover, in addition to the Swahili–Italian dictionary, the new software has been designed to be extended to other Bantu languages, in particular, a Zulu–Italian lexical database (including a collection of body vocabulary) is under development by R. Tramutoli and will be accessible through the general webpage of the lexicographic project.

The following sections briefly outline the microstructure and macrostructure of the dictionary as well as the software tools used to implement learner-oriented features.

3. Dictionary entries in KIU

Bantu languages are characterized by a complex noun class system, each class being expressed by a noun prefix; classes are generally paired as singular and plural, e.g. cl. 1 is singular, + human; cl. 2 is plural, + humans. Apart from classes 12 and 13 (not productive anymore in the language) and classes 11–14 which have merged into one class, the Swahili noun class system includes 15 noun classes. Swahili words are composed in a main, basic order, i.e. *Morpheme(s)+Stem*⁸:

- nouns: class prefix + noun stem (noun classes: cl. 1/2, 3/4, 5/6, 7/8, 9/10, 11–14/10; verbal nouns: class 15; 'locative' classes: 16, 17, 18)⁹ (see Table 1);
- adjectives: adjectival concords (cl. 1...18) + adj. stem (nominal prefixes and adjectival concords are usually identical, except for the concord *my*-class 4 which occurs with adjectives beginning with the vowels e- or o-, and for class 11–14 which follows the same adjectival agreement of nouns in class 3);
- pronominals: pronominal concords (cl. 1...18) + stem; stem + pr. conc. (1...18) (see Table 2);
- verbs: subject class concords (cl. 1...18) + tense marker + (relative conc. cl.1...18) + (object concords cl. 1...18) + verbal stem + final vowels/ extensions (see Table 3).

Table 1: Noun class prefixes

cl.	sing.	pl.
1/2	m-, mu-, w- ¹⁰	wa-, w-
3/4	m-, mu-, mw-	mi-
5/6	Ø, ji-, j-	ma-, m-
7/8	ki-, ch-	vi-, vy-
9/10	Ø, ¹¹ n-, ny-, m-, etc.	n-, ny-, m-, etc.
11/10	u-, w-	n-, ny-, m-, etc.
12/13 ¹²	ka-	tu-
15	ku-, kw-	

Table 2: Pronominal Concords (possessives and demonstratives)

Class	Possessive Concord	Demonstrative Concord
1/2	w-/w-	yu-/wa-
3/4	w-/y-	u-/i-
5/6	l-/y-	li-/ya-
7/8	ch-/vy-	ki-/vi-
9/10	y-/z-	i-/zi- y-/z-
11/10	w-/z-	u-/zi-
15	kw-	ku-
16	p-	pa-
17	kw-	ku-
18	mw-	mu-

Table 3: Subject Concord (SC), Object Concord (OC), Relative Concord (RC)

Class	SC	OC	RC
cl. 1/2 (1pers.)	ni-/tu-	-ni-/ -tu-	-ye-/ -o-
cl. 1/2 (2pers.)	u-/m-	-ku-/ -wa- -eni	-ye-/ -o-
cl. 1/2 (3pers.)	a-/wa-	-m-/ -wa-	-ye-/ -o-
cl. 3/4	u-/i-	-u-/ -i-	-o-/ -yo-
cl. 5/6	li-/ya-	-li-/ -ya	-lo-/ -yo-
cl. 7/8	ki-/vi-	-ki-/ -vi	-cho-/ -vyo-
cl. 9/10	i-/zi-	-i-/ -zi-	-yo-/ -zo-
cl. 11/10	u-/zi-	-u-/ -zi-	-o-/ -zo-
cl. 15	ku-	-ku-	-ko-
cl. 16	pa-	-po-	-po-
cl. 17	ku-	-ko-	-ko-
cl. 18	mu-	-mo-	-mo-

Swahili is an agglutinative language, which means that morphemes are juxtaposed to form words. Within the Swahili lexicographic tradition, the accepted lemmatisation strategy is to list nouns in their singular full forms with class prefixes, whereas the concords of verbs, numerals, pronouns and inflected adjectives are ignored, and the stems alone are listed (Wójtowicz 2016: 410). Thus, in standard printed dictionaries, we find the following types of Swahili entries:

- nouns: singular form (moyo→mioyo; moyoni, mioyoni¹³);
- adjectives: stem (-baya→mbaya, wabaya, mibaya, baya, mabaya, kibaya, vibaya, kubaya, pabaya, mubaya);
- pronouns: stem (h→huyu, hawa, huu, hii, hili, haya, hiki, hivi, hizi, huku, hapa, humu; huyo, hao, etc.);
- verbs: stem (-soma→somea, -someka, -somesha, -somwa, -somana, etc.);
- invariable forms (some prepositions, conjunctions, adverbials, etc.).

The problem of how to compile dictionaries for elementary learners is often presented in Bantu lexicography as a debate between 'stem lemmatization' (often considered as the 'right size') vs. 'word lemmatization' (typically the 'wrong size',

or at least looked down upon, especially by linguists) (De Schryver 2010: 162; see also Prinsloo 2009: 152-153).

The debate on the proper arrangement of the Bantu lexicon has been clearly summarized by Bennett (1986: 3-4), who highlights how the complex derivational system of Bantu languages poses difficulties:

If items are alphabetized by prefix [...] a verb will be listed far from its nominal derivations, however transparent these may be. [...] A competing school arranges the lexicon by stem or root; this usefully groups related items and saves on cross-referencing. Unfortunately, in such a system the user must be able to identify the stem, which given the sometimes complex morphophonemics of Bantu languages may not be easy.

Considering that Swahili students are used to the lexicographic 'stem tradition', in the online dictionary KIU we have decided to choose solutions found in nearly all Swahili dictionaries with regard to the process of lemmatization. Therefore, we have listed the stems alone for verbs, numerals, and inflected adjectives, ignoring agreement concords. However, it has been argued that the so-called 'stem tradition' in dictionary-making of Bantu languages is inadequate for young learners, who fail to isolate stems (see De Schryver 2010). In some cases, therefore, we have decided to reject traditional lexicographic solutions usually adopted in printed dictionaries and have responded to beginner learners' needs by opting for the lemmatization of full words for closed grammatical sets such as pronouns (including the stems also as separate entries).

Thus, each dictionary entry includes the following types of information:

- Headword
- [Variant/variants of the headword]
- [Full plural of nouns]
- PoS (part of speech), that is the grammatical category
- Noun class
- Translation (gloss and/or description)
- Examples (quotations, collocations, idioms, proverbs, etc.)
- Translations of examples
- [Cross-reference]

KIU (dizionario online Kiswahili-Italiano UniOr)

Ricerca veloce di un lemma

In questa sezione è possibile cercare termini italiani che si trovano nella traduzione dalla lingua target. La ricerca produce la lista delle voci che contengono il termine cercato. È una ricerca fruibile da chi ha una conoscenza della lingua di livello almeno iniziale-intermedio.

Risultati per: moyo

moyo

1. **moyo** *n. cl.3/4 (m-/mi-)* cuore; animo; coraggio

moyo, (pl. **mioyo**) (*n. cl.3/4 (m-/mi-)*) cuore; animo; coraggio **kwa moyo**, volentieri; a memoria **-jipa moyo/ -(ji)piga moyo konde**, farsi coraggio **-fa moyo**, essere scoraggiato
Cfr. **fuadi**

Fig. 8: Screenshot of the entry search 'moyo' in KIU

All entries contain an indication of the grammatical category (PoS) and at least one simple gloss. Most of the items are divided into sub-items and completed with examples of use. The inflected grammatical forms in particular contain various examples of use referring to the different meanings and functions. Most of the examples were taken from the UniOr digital corpus of Swahili texts, presently available only offline, which has been collected by M. Toscano in collaboration with a number of students and researchers. It is a small Swahili untagged raw corpus (1 million words) made mainly of about fifty full texts, mostly contemporary written literature, with the addition of some oral narratives and non-literary works (socio-political essays, handbooks about agriculture, media studies, informatics). The corpus is at the disposal of researchers and MA students for their thesis research on Swahili language, literature and linguistics.

Each entry is also categorized according to the type:

- form (nouns or invariable entries);
- grammatical stem (verbs, pronouns, adjectives, variable entries in general);
- morpheme: e.g. noun prefixes, subject prefixes, object markers, derivational suffixes, etc.;
- compounds.

Given Swahili word structure and the elements of the dictionary entries, it is evident that, since we are working with a Bantu language, we have to address problems not experienced by lexicographers working with European languages.

These problems are connected primarily to two issues: the form of headwords and the presentation of the numerous derivatives of a single root (Wójtowicz 2016: 410). In the following sections, we will explore challenges and difficulties regarding the design of a new Swahili online dictionary as a learning/teaching language tool.

4. How to transfer specific grammatical knowledge into an online dictionary

KIU is a learner dictionary, thus, differently from standard dictionaries, it has among its scopes, the aim of supporting Swahili students in autonomous language learning. It does this not only by building an up-to-date and comprehensive lexicon, but also by guiding students through a search process based on a system of interrelated grammar skills. In the construction of the online Swahili-Italian dictionary we faced several challenges which highlight the difficult process of transferring all grammatical knowledge and linguistic descriptions useful for Swahili students at beginner level into a lexicographic database. Without a doubt an online dictionary is much more comprehensive and allows for more types of search for entries compared to a printed dictionary, thus it represents a valuable tool for Swahili learners at a beginner level. Nevertheless, before presenting the new features of the dictionary (see par. 5), we should elucidate the limitations and difficulties experienced in producing a Swahili dictionary for learners.

The Swahili noun class system is quite standardized and homogeneous. Each noun class is expressed by prefixes which mark all elements of a Swahili sentence and thus encode the grammatical information necessary for grammatical agreement. Nevertheless, in some cases, apart from knowledge of class prefixes, other semantic and grammatical skills, which can hardly be included in entries in standard printed dictionaries, are required in order to build a correct Swahili sentence. For instance, Swahili students at beginner level face huge difficulties while applying the rule of animate noun agreement, that is, those nouns whose meaning refers to animate entities generally follow the same grammatical agreement as these latter, even if not belonging to class 1/2. Moreover, this rule does not apply to possessives with nouns in class 9/10 indicating human beings, which, only in this specific case, follow the possessive agreement of nouns in class 9/10:

<i>*rafiki wangu</i>	class 1 agreement
<i>rafiki yangu/rafiki zangu</i>	class 9 agreement

Apart from lexical entries, the dictionary headword list also includes grammatical morphemes, such as noun class prefixes, and, since it is not possible to translate the morphemes into Italian, their function in Swahili is explained. However, class 9/10 and class 5, include a huge number of loanwords, characterized by a zero prefix, thus, in this case, no entry can be inserted in the data-

base since no morpheme corresponds to the noun prefix. Similarly, in class 5, we generally find a zero prefix before deverbatives, e.g. *ombi* 'prayer', from the verb *kuomba* 'to pray'; the class 5 prefix *ji-* is found only before most of the monosyllabic roots, e.g. *jiwe* 'stone'.

In addition to the loanwords, class 9/10 contains invariable nouns, historically characterized by the nasal prefix N- which occurs before nominal and adjectival stems, and often undergoes phonological changes according to the general Bantu rules, for instance, the morpheme of class 9 *n-* changes to *ny-* before vowel-initial stems. Swahili learners looking for entries such as '*nyumba*, pl. *nyumba*'; '*ndizi*, pl. *ndizi*' or '*njia*, pl. *njia*' in the dictionary will need to have full knowledge of the corresponding phonological Bantu rules and some examples in order to remind them of the correct grammatical agreement according to the phonological changes of prefix *n-* + consonant: e.g. *njia ndefu* (**njia nrefu*); and *n-* + vowel: e.g. *ndizi nyingi* (**ndizi ningi*).

Furthermore, Swahili locative classes (16, 17, 18) behave differently from others, and grammatical rules applying to them can't be easily represented in a dictionary system (see also Toscano and Sewangi 2005: 274-275). For instance, the noun prefix of class 16 *pa-* only occurs with very few nouns (e.g. *pahala* 'place'), while the noun prefix of class 17 (*ku-*) is only used as adjectival concord and in copula constructions; similarly, the locative prefixes *mu-* and *mw-* (cl. 18) only occurs as a noun prefix in the term *mwahala*, which is often used as plural form of *pahala* 'place' (Bertoncini 2009: 186).

Some classes, which are still present in other Bantu languages, are not productive in Swahili, and, although we can still see traces in the language, this information is not transparent from the description in a dictionary entry. For instance, class 11–14 is conventionally indicated with a hyphen to point out that we are not dealing with a class pair (class 14 is not the plural class of 11), but rather with two different classes, originally separated, and which have merged into one in a later development of Swahili language, while remaining independent in other Bantu languages (such as in Zulu or Xhosa). In Swahili, class 11 includes mostly nouns referring to parts of a mass (e.g. *unywele* 'hair'), long objects (e.g. *ufunguo* 'key') and nominal deverbatives (e.g. *wimbo* 'song'). Class 14 is the class of abstract nouns (e.g. *utu* – humanity) and does not have a plural, while nouns of class 11 have the plural in class 10.

In most cases, singular and plural classes in the Swahili noun system are organized in pairs (for example, class 1 singular – class 2 plural; class 3 singular – class 4 plural, etc.). However, in a number of cases, Swahili nouns can have the plural in a different class (like in the case of nouns in class 11–14 which have plural in class 10) or even in two classes, for instance, some Swahili nouns can have plural both in class 6 and in class 10 (e.g. *rafiki/rafiki* or *rafiki/ marafiki*¹⁴). Also, a few nouns, especially loanwords, are assigned to class 5 or to class 9 in different dictionaries. For example, the words *kamusi* 'dictionary'; or *dawa* 'medicine' are indicated as belonging to cl. 5 in some dictionaries and to cl. 9 in others.

Thus, we have shown that there are crucial issues in transferring Swahili grammatical knowledge into a dictionary. Although there are some limitations,

we have tried to overcome most of them by creating an up-to-date online dictionary specifically designed to address the needs of learners, as illustrated in the next paragraph.

5. KIU as a learning support

Since the basic aim of KIU is to support Italian students in their autonomous learning of Swahili, it provides extended information beyond what is minimally necessary in a normal dictionary. Indeed, it is assumed that learner dictionaries (e.g. see also the Swahili–Polish dictionary presented in Wójtowicz 2016) aim to provide help not only in the process of text reception but also in text production, that is, the dictionary is not exclusively centred on translation. Rather, detailed information on grammar or lexical usage and information on morphological and syntactic structures are included, in order to support learning activities such as grammar exercises (drills and text comprehension), translations, and oral production/comprehension. The dictionary also aims at providing updated lexical information, i.e. loanwords from English (e.g. *skrini* 'screen') or neologisms such as those related to technology and informatics (for instance *simu ya mkononi*, 'mobile phone'; *tovuti* 'website'; *mtandao* 'internet') or those related to the COVID-19 pandemic (*barakoa* 'mask', *Korona* 'Corona virus').

Apart from the simple search for a lemma in the alphabetical order, it is also possible to search for Italian words in the translation of Swahili examples. Although KIU is not an IT–SWA dictionary, this search option is still helpful both for beginner Swahili students working on short text production or the translation of simple sentences and for advanced students dealing with oral and written production.

Grammatical stems (e.g. possessives, demonstratives) are searchable also in all inflected forms according to the noun class grammatical agreement. In this sense, the dictionary offers reliable support for Swahili learners at beginner level with specific entry features which help users to familiarize themselves with the Bantu noun class system and its morphological and phonological rules. Moreover, some irregular plural forms are entered immediately after the class prefixes, and the user has the possibility to search for both headwords and the full plural form (e.g. *jicho*, pl. *macho*; *uso*, pl. *nyuso*; *ulimi*, pl. *ndimi*).

This method is very convenient for learners at beginner level who do not have sufficient knowledge of grammar to enable them to identify easily singular and plural forms which carry different noun prefixes and are hidden in the entries of the singular form (Kiango 2005: 264). Derivatives, such as some extended verbs, are inserted as searchable entries and linked with the basic verb through a mechanism of cross-entry references, showing both sides of the derivational process (derivative→ root and root→ derivative) (Wójtowicz 2016: 411).¹⁵ Through the cross-reference system, users can also search for different phonological variants of the same word (e.g. *asante*, *ahsante*; *blanketi*, *blangeti*; *santuri*, *senturi*).

Furthermore, in order to enrich the learning tools and support beginner learners, the design of the updated Swahili–Italian dictionary provides a tool for adding grammatical comments or usage notes where appropriate (*dictScrap*). This option allows the quality of information contained in the entry to increase with the aim of supporting the acquisition of grammatical rules and expanding vocabulary. This is achieved through the addition of the following.

- Indications on the correct grammatical agreement for more complex cases: e.g. the agreement of animate nouns from non-human classes (e.g. *kijana* 'young man' cl. 7; *bibi*, cl. 9 'grandmother, lady'); possessive agreement with animate nouns in class 9/10 referring to close relationship (e.g. *bibi yangu* 'my grandmother': the possessive agrees in class 9 and not in class 1 **bibi wangu*), etc.
- Notes on the semantic features of a term in order to disambiguate meanings and facilitate the appropriate choice or use of a term in translations and oral and written production.¹⁶ For this purpose, semantic explanations can also be accompanied by a number of labels indicating status (formal, informal, slang, derogative, euphemism, vulgar, colloquial etc.); register (literary, familiar, popular, etc.); semantic field (biology; zoology; military; music; legal; medicine; religion, etc.), frequency of use (common, rare); figurative or extended meaning.

In addition, cultural terminology, referring to untranslatable concepts and things can be supported by images accompanied by a definition. The possibility of adding images and descriptions to an entry is particularly helpful in order to clarify cultural terms which can't be easily translated because they are not part of Italian culture or in cases in which the Italian gloss or definition is not exhaustive enough to explain the concept (e.g. *ugali*: 'typical Swahili food similar to polenta'; *kanga*: 'coloured women's textile'; *chapati* 'Indian unleavened flatbread'). This additional information is generally avoided in printed dictionaries and can't be included due to printing size restrictions.

Finally, unlike printed dictionaries, learners can rely also on audio recordings of difficult pronunciations, e.g. the sound 'j' in the word *jicho* or the sound 'gh' in *ghali*.

6. Endnotes

1. These are the main bilingual publications: Krapf, *Swahili English Dictionary* (1882); Sacleux, *Dictionnaire Français–Swahili* (1891) and *Dictionnaire Swahili–Français* (1939); Madan, *English–Swahili Dictionary* (1894) and *Swahili–English Dictionary* (1903); Johnson, *A Standard Swahili–English Dictionary* (1939) and *A Standard English–Swahili Dictionary* (1939). For more details see Chuwa 1996.
2. In 2009 the Institute was renamed TATAKI (Taasisi ya Taaluma za Kiswahili — Institute of Swahili Studies).

3. Some examples are A. Lenselaer, *Dictionnaire Swahili-Français* (partiellement adapté du Standard Swahili-English Dictionary de Frederick Johnson avec l'autorisation d'Oxford University Press), Karthala, Paris 1983; G. Mertens, *Dictionnaire Kiswahili-Français et Français-Kiswahili*, Karthala, Paris 2006; K. Legère, *Wörterbuch Deutsch-Swahili*, Köppe, Cologne 2006; A.I. Kutuzov, *Kratkij Suakhili-Russkij i Russko-Suakhili Slovar / Kamusi fupi ya Kiswahili-kirusi na Kirusi-kiswahili*, Izdatel'stvo Sovetskaja Enciklopedija, Moscow 1965; B.V. Trivedi, *Gujarati-Swahili Shabdapothi* (Gujarati-Swahili Dictionary), Mombasa 1955.
4. 'Lakini, wote wawili walikabili magumu katika utumishi wao kwa Mungu' has been translated as 'Eppure, entrambi/e subirono una grave battuta d'arresto' (Yet, both suffered a serious setback), instead of 'Eppure, entrambi/e affrontarono difficoltà nel loro servizio a Dio' (Yet, both faced hardships in their service to God). Utumishi in this context means 'servizio, servire (Dio)', i.e. 'service, to serve (God)'.
5. T.E.I. Text Encoding Initiative: <http://www.tei-c.org/index.xml>
6. www.lessicafrica.it. The work is in progress, and so the link is not yet active. It will be activated by the end of 2023.
7. The project is funded by the Department of Asian, African and Mediterranean Studies at UniOr and by ISMEO (International Association for Mediterranean and Oriental Studies); see the UniOr webpage KIU (Kiswahili-Italian-Unior): the online Swahili dictionary for Italian-speaking learners.
8. Except for the pronominal stem H-, followed by the vowel and the concord, e.g. *kitabu h-i-ki* 'this book', cl. 7 (stem + morpheme).
9. Swahili locative classes currently only include the nouns *mahala/pahala* 'place, location' (cl. 16) and the plural *mwahala* (cl. 18).
10. The last prefixes indicated in the table usually precede stems beginning with vowels.
11. The original prefix *n-* (class 9/10) often disappears or undergoes phonetical changes when occurring before certain initial sounds of the Bantu stems, modifying to *ny-* (before vowel) or *m-* etc.
12. Classes of diminutives used in some Tanzanian Bantu languages and imported in some Swahili non-standard varieties that are spoken in interior regions.
13. *Moyo/mioyo* (cl. 3/4): 'heart'; *moyoni*: 'in the heart'; *-ni* is a locative suffix that expresses the locative relations indicated in Italian through prepositions, such as 'in/at/from' (cf. Bertoncini Zúbková 2009: 7).
14. Cfr. *Rafiki/rafiki* (Bertoncini Zúbková 2009: 30); *rafiki/marafiki* (BAKITA 2015: 878)
15. A fundamental aspect of the Bantu languages is the possibility of extending the basic form of the verb (e.g. *-penda* 'to love'), by adding different suffixes to the root, thus modifying the meaning (and often also the morphosyntactic features) of the verb, e.g. *-pend-ana* 'loving each other', *-pend-wa* 'to be loved', *-pend-elea* 'to prefer', etc.
16. A cross-reference system is also exploited to link synonymous entries, such as *kinywa* – *mdomo* 'mouth'.

7. References

- Abdulla, A., R. Halme, L. Harjula and M. Pesari-Pajunen. 2002. *Swahili-Suomi-Swahili Sanakirja*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Aiello, F. and M. Toscano.** 2017. On Some Swahili ICT Terms. Batic, G. and R. Leger (Eds.). 2017 *Studia Africana — Papers in Honour of Sergio Baldi*: 19-32. Cologne: Rüdiger Köppe Verlag.
- BAKITA.** 2015. *Kamusi Kuu ya Kiswahili*. Nairobi/Kampala/Dar es Salaam: Longhorn.
- Bennett, P.R.** 1986. Grammar in the Lexicon, Two Bantu Cases. *Journal of African Languages and Linguistics* 8(1): 1-30.
- Bertoncini Zúbková, E.** 1973. A Tentative Frequency List of Swahili Words. *AION* 33: 297-363.
- Bertoncini Zúbková, E.** 1977. *Vocabolario swahili-italiano e italiano-swahili*. Naples: Opera Universitaria I.U.O.
- Bertoncini Zúbková, E.** 2009. *Kiswahili kwa Furaha. Corso di lingua swahili*. Rome: Aracne.
- Chuwā, A.R.** 1996. Problems in Swahili Lexicography. *Lexikos* 6: 323-329.
- De Schryver, G.-M.** 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.
- Kiango, J.G.** 2005. Problems of Citation Forms in Dictionaries of Bantu Languages. *Nordic Journal of African Studies* 14(3): 255-273.
- Kutuzov, A.I.** 1965. *Kratkij Suakhili-Russkij i Russko-Suakhili Slovar / Kamusi fupi ya Kiswahili-kirusi na Kirusi-Kiswahili*. Moscow: Izdatel'stvo Sovetskaja Enciklopedija.
- Legère, K.** 2006. *Wörterbuch Deutsch-Swahili*. Cologne: Rüdiger Köppe Verlag.
- Lenselaer, A.** 1983. *Dictionnaire Swahili-Français* (partiellement adapté du Standard Swahili-English Dictionary de Frederick Johnson avec l'autorisation d'Oxford University Press). Paris: Karthala.
- Martini, G.** 2016. *Dizionario swahili. Swahili-italiano, italiano-swahili*. Milan: Hoepli.
- Merlo Pick, V.** 1961. *Vocabolario kiswahili-italiano e italiano-kiswahili*. Turin: EMI.
- Mertens, G.** 2006. *Dictionnaire Kiswahili-Français et Français-Kiswahili*. Paris: Karthala.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.
- Sewangi, S.** 2007. Uundaji wa Istilahi za Kiswahili Kileksikografia na Athari zake. *Nordic Journal of African Studies* 16(3): 333-344.
- Toscano, M.** 2004. *Dizionario swahili. Swahili-italiano, italiano-swahili*. Milan: Vallardi.
- Toscano, M. and S. Sewangi.** 2005. Discovering Usage Patterns for the Swahili *amba-* Relative Forms cl. 16, 17, 18: Using Corpus Data to Support Autonomous Learning of Kiswahili by Italian Speakers. *Nordic Journal of African Studies* 14(3): 274-317.
- Tramutoli, R.** 2018. *Kamusi ya isimu Kiswahili-Kiitaliano*. Dar es Salaam: TUKI.
- Wójtowicz, B. (Ed.).** 2003. *Studencki Słownik Suahilijsko-Polski*. Available online at <https://www.mimuw.edu.pl/~jsbien/BW/SSSP/SSSP.pdf>.
- Wójtowicz, B.** 2016. Learner Features in a New Corpus-based Swahili Dictionary. *Lexikos* 26: 402-415.
- Wójtowicz, B. and P. Bański.** 2012. Swahili Lexicography in Poland: Its History and Immediate Future. *Studies in African Languages and Cultures* 46: 59-72.

Evaluating the Usefulness of the Learning Tools in Monolingual Online Dictionaries for Learners of English: Gauging the Preferences of Polish Students of English

Bartosz Ptasznik, *University of Warmia and Mazury, Olsztyn,
Poland (bartosz.ptasznik@uwm.edu.pl)*

Abstract: The aim of the report is twofold. First, the learning tools available in monolingual online dictionaries for learners of English are described. Second, an evaluation of the usefulness of the learning tools in online dictionaries is provided. To meet the aims of the present contribution, a survey was administered on 318 Polish students of English. The respondents, who participated in a lecture devoted to the topic of learning tools in online dictionaries, were instructed to complete a questionnaire. A mixed-question format was adopted. In the first part of the questionnaire, the participants had to rate the usefulness of the features of the *Macmillan Dictionary* on a semantic differential scale of 1–7. In the second part, there were two open-ended questions. The students were asked to name the most and least useful learning tools of the *Macmillan Dictionary* and explain their choices. The results suggest that English majors studying at a Polish university accord high priority to consulting online learning tools which give them valuable information on collocations, synonyms and semantically related words.

Keywords: LEXICOGRAPHY, DICTIONARY USE, ONLINE DICTIONARY, MONOLINGUAL LEARNERS' DICTIONARY, ONLINE LEARNING TOOLS, SURVEY

Opsomming: Die evaluering van die bruikbaarheid van die aanleerdershulpmiddels in eentalige aanlyn woordeboeke vir aanleerders van Engels: Die bepaling van die voorkeure van Poolse studente van Engels. Die doel met hierdie verslag is tweërlei. Eerstens word die aanleerdershulpmiddels wat in eentalige aanlyn woordeboeke vir aanleerders van Engels beskikbaar is, beskryf. Tweedens word 'n evaluering van die bruikbaarheid van die aanleerdershulpmiddels in aanlynwoordeboeke verskaf. Om aan die doelwitte van hierdie bydrae te voldoen, is 'n opname met 318 Poolse studente van Engels gedoen. Die respondente, wat deelgeneem het aan 'n lesing gewy aan die onderwerp van aanleerdershulpmiddels in aanlyn woordeboeke, is versoek om 'n vraelys te voltooi. 'n Formaat van gemengde vrae is gebruik. In die eerste deel van die vraelys moes die deelnemers die kenmerke van die *Macmillan Dictionary* op 'n semanties gedifferensieerde skaal van 1–7 plaas. In die tweede deel was daar twee oop vrae. Die studente is versoek om die mees en mins bruikbare aanleerdershulpmiddels van die

Macmillan Dictionary te lys en om redes te gee vir hul keuses. Die resultate dui daarop dat studente met Engels as hoofvak wat aan 'n Poolse universiteit studeer, hoë prioriteit verleen aan die raadpleeg van aanlyn aanleerdershulpmiddels wat aan hulle waardevolle inligting oor kollokasies, sinonieme en semanties verwante woorde verskaf.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEKGEBRUIK, AANLYN WOORDEBOEK, EENTALIGE AANLEERDERSWOORDEBOEK, AANLYN AANLEERDERSHULPMIDDELS, OPNAME

1. Introduction

In today's digital world, online dictionaries present us with more information than ever before. Electronic dictionaries are more than just dictionaries. Apart from giving dictionary users the meanings of words, these reference resources can be perceived as repositories of extensive knowledge about the language. This knowledge is shared with users online via different kinds of learning tools, such as videos, language games and quizzes, thesauruses (integrated into dictionaries), blogs, translator tools, exam practice exercises and word of the day or word of the week features. No doubt, learning a foreign language outside a context restricted to book learning can be a more pleasant experience for the language learner. The wide array of online learning tools made available to users on the internet presents ample opportunity for students to further their linguistic skills. But do learners actually consult all of the different tools accompanying present-day online dictionaries? And assuming they do, how often do they decide to use them? Another question pertaining to the present report is how useful are these learning tools for students of English?

2. Learning tools in online dictionaries

The principal source of lexicographic information in an online monolingual learners' dictionary is the dictionary page. Learners of a language consult dictionaries primarily for meaning (Summers 1988: 113-114; Nuccorini 1992: 89-90; Lew 2010: 291-292; Ptasznik 2022: 236-237). In dictionaries, meanings of words are supplemented with example sentences, collocational and grammatical information (grammar codes and grammar patterns, word forms, syntactic class information), pronunciation and frequency information, helpful sense-navigation devices in the form of signposts and menus (appearing with the most polysemous words), synonyms and related words, as well as etymological and derivational information. Additionally, different English monolingual learners' dictionaries have their own unique features, appearing on the dictionary page under specific entries. For example, the *Macmillan Dictionary* incorporates *Metaphor boxes* (for words and phrases appearing with their metaphorical meanings), *Get It Right! boxes* (which illustrate correct usage of words, grammatical

patterns, provide examples of grammatically incorrect sentences, etc.) and *Expressing yourself boxes* (for example, they show what types of phrases one can use when suggesting something politely in a formal context). From the *Longman Dictionary of Contemporary English*, users can advance their knowledge of grammar patterns from *Grammar notes* and *Common Errors notes*, whereas the *Oxford Advanced Learner's Dictionary* provides users with extra *Vocabulary Building* sections.

Nowadays online dictionaries are equipped with an integrated thesaurus. For example, the *Macmillan Thesaurus* gives synonyms, antonyms, and related words which form a lexical set of words for specific words and concepts. In the *Merriam-Webster.com Dictionary*, the *Merriam-Webster Thesaurus* presents language information in the form of synonyms and similar words, antonyms and near antonyms, the *Synonym Chooser* feature (which explains the differences between selected synonyms and similar words, for example, *eager* and *anxious*), as well as phrases containing the searched word (*eager beaver* for *eager*) and articles related to the searched word ('*When Pigs Fly*' and *Other Barnyard Idioms* for the word *buy*).

In the light of the fact that correct use of collocations¹ creates a serious challenge to language learners (Bahns and Eldaw 1993: 101; Herbst 2010: 225; Lew and Radłowska 2010: 43; Chan 2012: 69), dictionaries may additionally incorporate corpus-based collocations dictionaries tailored to suit learners' productive needs. To give an example, the *Macmillan Collocations Dictionary* of the *Macmillan Dictionary* supplies dictionary users with a wide range of strong collocates for core vocabulary. Entries, which are organized by grammar and meaning, include collocations for different kinds of grammatical relations (*adjective + noun*, *verb + noun*, *noun + verb*, *adverb + adjective*, etc.) and may also provide valuable usage notes (see entries for *recognize*² and *vanish*). Importantly, the *Macmillan Collocations Dictionary* has been geared to meet the needs of all types of learners, regardless of their proficiency levels.

To increase their attractiveness, publishers supply online dictionaries with translation tools. For example, the *Cambridge Dictionary* is accompanied by a *Translator*³ tool. This multilingual translation learning tool allows one to type the text in the box and get instant translations for selected texts. And there is a broad range of possibilities: translations are performed for a fairly high number of different languages, including Arabic, Catalan, Filipino, Hindi, Korean, Polish and Thai. In the *Translator* tool of the *Collins Online Dictionary*, more than 30 languages are available.

The provision of video presentations in dictionaries for language learning purposes is standard practice nowadays. Learners have the opportunity to listen to native speakers of English elaborate on a chosen topic. Video material with a special focus on *The Schwa*, *Lay vs. Lie*, *Sneaked vs. Snuck*, *On Contractions of Multiple Words*, *A Look at Uncommon Onomatopoeia*, *What is an Eggcorn?*, *How a Word Gets into the Dictionary*, *Ending a Sentence with a Preposition* and a plethora of others can be accessed from the *Merriam-Webster.com Dictionary* website. The

videos that have been made available at *MacmillanDictionary.com* are concerned with *Real Grammar* (e.g. *Split Infinitives*), *Real Vocabulary* (e.g. *Uninterested vs. Disinterested*) and *Real Word English* (e.g. *Politeness*) topics. The *Collins Online Dictionary* provides audiovisual material under the *Learn English*, *Video pronunciations* and *Build your vocabulary* sections.

Language games and quizzes make dictionaries more appealing. *Merriam-Webster.com Dictionary* features various quizzes to test learners' knowledge of herbs and spices, flowers, weather words, colors, fashionable words, Game of Thrones words, monsters, obscure shapes, finance words, words of snow and ice and farm idioms, to name just a few, while the *Macmillan Dictionary* develops students' knowledge through language quizzes (e.g. breeds of dog, royal words, figures of speech), games (e.g. irregular verb wheel) and puzzles (e.g. American culture wordsearch, Halloween wordsearch).

Another frequently-used learning tool in dictionaries is the language blog. For example, the *Macmillan Dictionary Blog* incorporates diary-style posts for specific words, including supplementary information with regards to their meaning, usage (grammar), etymology and spelling.

By and large, publishers also endeavor to develop a closer bond between the user and lexicographer. The *Macmillan Dictionary* includes the *Open Dictionary* feature, which enables users to participate in the dictionary-making process, by creating the opportunity for learners to suggest words that could be added to the dictionary, given their frequently-recurring usage in the language. Whether the word or phrase will eventually be incorporated into the dictionary or not is a decision solely made by lexicographers.

Other learning tools include: practice tests and exercises, word lists, buzzwords, trending words, the word of the day feature, English grammar explanations, text checkers, reference materials, etc.

3. Aims of the survey

The aims of the present survey are twofold. First, selected types of online learning tools accompanying English monolingual learners' dictionaries will be described (see section devoted to the learning tools in online dictionaries). Second, the usefulness of these online learning tools will be explored by examining the preferences and habits of English majors at the University of Warmia and Mazury in Olsztyn, Poland.

4. Method, participants and procedure

The online version of the *Macmillan Dictionary* was selected for the evaluation of the usefulness of the learning tools in online dictionaries. This decision rested on two premises. First, in a previous survey of dictionary preferences (Ptasznik 2022), examining the lexical resources that students of English choose

to consult, it was found that the *Macmillan Dictionary* is the second most frequently consulted British English monolingual learners' dictionary among the students participating in the survey (the respondents of the previous survey were provided with five different answer options, which also included the *Cambridge Dictionary*, *Oxford Advanced Learner's Dictionary*, *Collins Online Dictionary* and *Longman Dictionary of Contemporary English*). In order to answer the purpose of the survey, it was assumed that using this particular reference source would increase the reliability and validity of the evaluation of specific online learning tools, given the students' familiarity with this digital resource. Second, the *Macmillan Dictionary* presents dictionary users with a comprehensive range of learning tools available online at the click of a mouse. Apart from using the dictionary page, users can learn English by accessing the *Macmillan Collocations Dictionary*, *Macmillan Thesaurus*, *Macmillan Dictionary Blog*, *Language Quizzes*, *Games and Puzzles*, *Videos* (Real Grammar, Real Vocabulary and Real World English videos), *Buzzwords* and *Trending Words*. Importantly, the above-mentioned language learning tools share the common features of the learning tools in the remaining monolingual learners' dictionaries that are available online. All things considered, it was the *Macmillan Dictionary* that was deemed to be the most suitable dictionary for the evaluation of the usefulness of the learning tools in online dictionaries.

318 English majors from the University of Warmia and Mazury participated in the survey. The participants were males and females, who were native speakers of Polish. Their age varied between 19–24. The respondents were first-year, second-year, third-year, fourth-year and fifth-year students of English (full-time studies). Their English proficiency level ranged from upper-intermediate to advanced (level B2 and C1 by the Common European Framework of Reference for Languages standards). The students made a voluntary decision to participate in the survey. The protection of sensitive personal information was achieved through data anonymization.

To meet the aims of the research, it was divided into two parts. In the first part, the students were asked to attend a one-hour lecture devoted to the topic of learning tools in online dictionaries. Several lectures were given, with a turnout of approximately 40–70 students for each separate lecture. The lectures were delivered by the researcher at the Faculty of Humanities of the university. As mentioned above, the *Macmillan Dictionary* was selected for the evaluation of the usefulness of the learning tools in online dictionaries. During the lecture, it was explained in detail to the study participants what kinds of learning tools have been made available to dictionary users by *Macmillan Dictionary* publishers: Dictionary page, *Macmillan Collocations Dictionary*, *Macmillan Thesaurus*, *Macmillan Dictionary Blog*, *Language Quizzes*, *Games and Puzzles*, *Videos* (Real Grammar, Real Vocabulary and Real World English videos), *Buzzwords*, *Open Dictionary* (entries⁴) and *Trending Words*. All of these learning tools were discussed in as much detail as possible within the required one-hour time frame, with specific learning tools being displayed to the participants on screen. Taking

into account the fact that Polish students of English are familiar with this specific digital product (Ptasznik 2022), the lecture merely served as a reminder of what specific learning tools can be accessed from the dictionary. Examples of similar learning tools from the remaining monolingual learners' dictionaries were provided during the course of the lecture. For example, the following features from the *Merriam-Webster.com Dictionary* were presented to the students: *Merriam-Webster Thesaurus*, *Games & Quizzes*, *Videos*, *Word of the Day*, *Words at Play*. At the end of the lecture, the students were allowed to ask questions.

Given the usefulness of surveys in dictionary-user research (see Kosem et al. 2019) and the advantages of designing a questionnaire (see Jackson and Furnham 2000: 5; Mackey and Gass 2005: 94-96; Blaxter et al. 2006: 79; Debois 2019), the students were asked to fill out a questionnaire in the second part of the research. A pencil-and-paper format was selected for the questionnaire. To meet the aims of the survey, a mixed-question format was adopted (quantitative and qualitative analysis). In the questionnaire, in order to assess the students' opinions, attitudes and experiences, the participants had to rate the following features of the *Macmillan Dictionary* on a semantic differential scale of 1–7 (lower scale points indicated that a specific learning tool was *less useful*, higher scale points indicated that a specific learning tool was *more useful*, see Appendix A and Appendix B): Dictionary page, *Macmillan Collocations Dictionary*, *Macmillan Thesaurus*, *Macmillan Dictionary Blog*, *Language Quizzes*, *Games and Puzzles*, *Open Dictionary*, *Videos*, *Buzzwords*, *Trending Words* and *Red Words and Stars*. Open-ended questions were used sparingly (two open-ended questions were included). The students were supposed to name the most and least useful features of the *Macmillan Dictionary*, and explain their choices. To increase comprehension of the survey questions, the ten questionnaire items and two open-ended questions were drafted in the students' native language. To avoid confusion, technical language was not used in the questionnaire (Lew 2002, 2004). The students were given twenty minutes to complete the task (for more on the optimum duration of surveys see Macer and Wilson 2013, as cited in Brace 2018: 49-50). They had five minutes to rate the learning tools of the *Macmillan Dictionary* and an additional fifteen minutes to answer two open-ended questions.

5. Results, findings and discussion

The results of the questionnaire are given in Table 1. In the table, the number of responses of the respondents are provided. Additionally, the results are expressed in percentage terms (scale points are numbered from 1–7, scale points with a higher value indicate that a specific learning tool is *more useful*, scale points with a lower value indicate that a particular learning tool is *less useful*.).

Table 1: Results of the questionnaire

LEARNING TOOL	RATING SCALE						
	1	2	3	4	5	6	7
<i>Dictionary page</i>	4 (1%)	8 (3%)	6 (2%)	26 (8%)	90 (28%)	78 (25%)	106 (33%)
<i>Macmillan Collocations Dictionary</i>	0 (0%)	2 (1%)	2 (1%)	32 (10%)	44 (14%)	104 (33%)	134 (42%)
<i>Macmillan Thesaurus</i>	4 (1%)	6 (2%)	8 (3%)	36 (11%)	52 (16%)	76 (24%)	136 (43%)
<i>Videos</i>	20 (6%)	32 (10%)	44 (14%)	66 (21%)	62 (19%)	56 (18%)	38 (12%)
<i>Macmillan Dictionary Blog</i>	30 (9%)	44 (14%)	78 (25%)	84 (26%)	48 (15%)	22 (7%)	12 (4%)
<i>Language Quizzes, Games, Puzzles</i>	10 (3%)	28 (9%)	28 (9%)	48 (15%)	40 (13%)	80 (25%)	84 (26%)
<i>Open Dictionary</i>	6 (2%)	18 (6%)	24 (8%)	50 (16%)	106 (33%)	84 (26%)	30 (9%)
<i>Buzzwords</i>	8 (3%)	10 (3%)	28 (9%)	90 (28%)	86 (27%)	66 (21%)	30 (9%)
<i>Trending Words</i>	16 (5%)	18 (6%)	38 (12%)	60 (19%)	76 (24%)	64 (20%)	46 (14%)
<i>Red Words and Stars</i>	4 (1%)	18 (6%)	34 (11%)	68 (21%)	76 (24%)	60 (19%)	58 (18%)

The findings invite some general conclusions. The data obtained from 318 participants suggest that upper-intermediate and advanced students of English highly value learning tools that provide them with information about collocations and synonyms (and related words). In the present survey, both the *Macmillan Collocations Dictionary* and *Macmillan Thesaurus* received the highest ratings from the respondents. As many as 43% (136 students) of the students gave the *Macmillan Thesaurus* a rating of 7, while the *Macmillan Collocations Dictionary* received the same rating from a slightly lower number of students (134 students). The survey participants were satisfied with the Dictionary page. More than three quarters (86%) of the students held the view that the Dictionary page was useful, by giving it a rating within the range of 5–7. Given that dictionary users consult dictionaries primarily for meaning, and taking into account various user-friendly features of the dictionary, such as entry navigation devices in

the form of menus⁵ (used for the most polysemous words, for example, under the entry *see*), hyperlinking, recordings of sounds and noises (see Lew 2010: 297; Jackson 2018: 545; Heuberger 2020: 407), customization (e.g. show/hide further synonyms and/or related words), grammar patterns, *Metaphor Boxes*, and the incorporation of example sentences, this finding does not come as a surprise. In addition, it appears that English majors from Polish universities appreciate different types of language games and quizzes that can be accessed from an online dictionary. The data show that the students rated *Language Quizzes, Games, Puzzles* as a more useful learning tool than all the remaining features, that is the *Macmillan Dictionary Blog, Videos, Buzzwords, Open Dictionary, Red Words and Stars* and *Trending Words*. 51% of the students gave *Language Quizzes, Games, Puzzles* a rating of either 6 or 7. By comparison, this figure came close to reaching the ratings of the Dictionary page, with 58% of the respondents giving the Dictionary page a rating of 6 or 7. The participants found the *Open Dictionary* entries quite appealing, with more than 50% of the respondents giving it a rating of 5 or 6. The students rated the *Macmillan Dictionary Blog* as the least useful learning tool of the *Macmillan Dictionary*. More than half of the students (51%) gave it a rating of 3 or 4.

To elicit more detailed responses on the usefulness of the learning tools in the *Macmillan Dictionary*, the students were asked to answer the following questions (two open-ended questions):

- (1) Which features of the *Macmillan Dictionary* are the most useful learning tools in your opinion? Why?
- (2) Which features of the *Macmillan Dictionary* are the least useful learning tools in your opinion? Why?

A general consensus was reached by the survey participants on the perception of the usefulness of the *Macmillan Collocations Dictionary* and *Macmillan Thesaurus*. The majority of the students answered that the *Macmillan Collocations Dictionary* and *Macmillan Thesaurus* were the most useful learning tools, given their significance in language production tasks. Most of the students mentioned that they would especially resort to using these tools when completing their writing assignments, such as: paragraph and essay writing, as well as writing BA and MA dissertations. In addition, the participants reported that they would consult the *Macmillan Collocations Dictionary* and *Macmillan Thesaurus* when preparing an oral presentation, as both learning tools have considerable potential for vocabulary building. "Getting ready for a job interview at multinational companies" was also high on their list. No doubt, using collocations correctly can be a mammoth task for language learners (Bahns and Eldaw 1993: 101; Herbst 2010: 225; Lew and Radłowska 2010: 43; Webb and Kagimoto 2011: 260-261; Chan 2012: 69; Daskalovska 2015: 130-131). Students can sense that acquiring more knowledge about collocations plays an important role in the process of language learning. It is essential that they learn to use words naturally in productive mode. Without sufficient collocational knowledge, production becomes

seriously affected. Similarly, using synonyms correctly in the target language is equally important.

Overall, the Dictionary page was credited with consistency (of how lexicographic information is organized within entries) and transparency of the presentation of information. Many of the students lauded it for lexicographic content. The students appreciated the incorporation of entry menus for polysemous words, hyperlinking of words within definitions, recordings of sounds and noises and Metaphor Boxes (for an example, see the Metaphor Box under the entry *illness*). Some of the participants, however, believed that fewer example sentences can be accessed from this reference source than from the remaining monolingual online dictionaries for learners of English, such as the *Longman Dictionary of Contemporary English* or *Collins Online Dictionary*, which provide users with a plethora of supplementary corpus examples in the extra sections of entries. Others, however, put forward the argument that the number of examples included in *Macmillan Dictionary* entries is sufficient, and that incorporating more lexicographic data could lead to learners becoming overwhelmed with an excess of information. Such an argument has also been presented in the literature (L'Homme and Cormier 2014: 333; Gouws and Tarp 2017: 394; Frankenberg-Garcia 2020: 32). A handful of the participants also said that recorded pronunciations of headwords for both British and American English are not as easily accessible as in, for example, the *Oxford Advanced Learner's Dictionary*, where both types of pronunciation variants appear at the top of a single entry. To get access to comprehensive information on different pronunciation variants, one needs to obtain this type of information from an appropriate link located at the bottom of the *Macmillan Dictionary* page, or change the default settings for the dictionary.

The *Macmillan Dictionary Blog* was assessed by the students as the least useful learning tool of the *Macmillan Dictionary*. The primary reason for this appears to be the fact that dictionary users want to acquire information as quickly as possible. Taking into account the fact that the pool of students selected for the study were digital natives, growing up with various technologies and learning tools available and easily-accessible in their digital-rich environment, it is perhaps not surprising that a fairly high number of the students rated this specific feature of the dictionary as being less beneficial than the remaining online learning tools. When learning the language online, students are not necessarily keen on scanning through larger fragments of text. The average dictionary user wants to build on his knowledge with the minimum of effort. Nowadays information is available on the internet at the click of a mouse. Dictionary users value their time and ascribe much importance to faster dictionary consultation (Bogaards 1998: 561; Chen 2010: 292; Chan 2012: 87; Knežević et al. 2021: 7), and that is why learners are likely to launch less time-consuming information search strategies rather than engage in longer reading.

As for the remaining learning tools, in their open-ended feedback the students appeared to be excited that modern online dictionaries give them access

to information about neologisms or regional varieties, as well as words that are being increasingly used by the speakers of the language, but which have yet to be incorporated into the dictionary. Hence, some of the students were full of praise for *Buzzwords* and the *Open Dictionary* entries that have been made available to the users. This is proof that such features of dictionaries can generate an avid interest in learning the target language. There is no denying that these learning tools unleash creativity and build learners' vocabulary which could be particularly important when communicating in more informal contexts. Especially more advanced learners strive for such knowledge. *Videos* and *Language Quizzes, Games, Puzzles* were assessed as being useful teaching materials for the EFL classroom, given that selected students had been planning to embark on a teaching career after the completion of their studies. Quite a few of the students wondered why more videos had not been uploaded to *MacmillanDictionary.com* (for a few possible explanations, see Heuberger 2020: 408). Others suggested that additional topics be covered in the *Videos*, such as verb complementation patterns (for example, intransitive, monotransitive and ditransitive verbs, reflexive and reciprocal verbs, verbs followed by *that*-clauses, verbs used with prepositional and adverbial phrases), verbs commonly used in the passive, conditional sentences, etc. According to the students, video clips with a specific focus on more challenging verbs in the English language from the point of view of language production (e.g. suggest, recommend, demand, explain) would be a valuable addition to the feature. *Red Words and Stars* were regarded by many as a positive feature of the *Macmillan Dictionary*. Some of the students, however, admitted that they never paid attention to frequency information in a dictionary, as they consulted dictionaries solely for meaning and collocations. Given that highlighting core vocabulary (*red words* vs. *black words*) in dictionaries raises learners' awareness about which words are more important from a productive, as well as receptive perspective, it must be contended that it is the role of the teachers to make their students more aware of the practical significance of the presence of this feature in monolingual learners' dictionaries.

6. Conclusion

The aim of the report was to gauge the preferences and habits of Polish students of English regarding their use of the learning tools accompanying monolingual online dictionaries. The survey participants were asked to indicate their degree of satisfaction with selected online learning tools on a seven-point scale and answer two open-ended questions. Considering the findings of the present work, it may be concluded that Polish learners of English representing an upper-intermediate and advanced level think highly of online learning tools designed to supply users with pertinent information on collocations, synonyms and related words. This finding implies that English majors studying at Polish universities place much importance on language production (specifically, academic writing) in the process of foreign language learning.

Endnotes

1. See Laufer (2011), Laufer and Waldman (2011) and Chen (2017) for more on collocations in production. See Herbst (1996) and Dziemianko (2014) for more on encouraging collocational awareness among learners.
2. From the entry *recognize* in the *Macmillan Collocations Dictionary*, one learns that there is a marked preference for passivization when the verb *recognize* is used in an *adverb + verb* combination.
3. See: <https://dictionary.cambridge.org/translate/>.
4. For more on crowdsourcing and user-generated content see Rundell (2012: 80-81).
5. For more information on menus in dictionaries see Rundell (2007), Lew and Tokarek (2010), Ptasznik (2015).

References

Online dictionaries

- Cambridge Dictionary*. <https://dictionary.cambridge.org/>. Accessed 24 November, 2022.
- Collins Online Dictionary*. <https://www.collinsdictionary.com/dictionary/english>. Accessed 24 November, 2022.
- Longman Dictionary of Contemporary English*. <http://www.ldoceonline.com/>. Accessed 24 November, 2022.
- Macmillan Dictionary*. <http://www.macmillandictionary.com>. Accessed 24 November, 2022.
- Merriam-Webster.com Dictionary*. <https://www.merriam-webster.com/>. Accessed 24 November, 2022.
- Oxford Advanced Learner's Dictionary*. <https://www.oxfordlearnersdictionaries.com/>. Accessed 24 November, 2022.

Paper dictionaries

- Rundell, M. (Ed.). 2007. *Macmillan English Dictionary for Advanced Learners*. (MED2). Second edition. Oxford: Macmillan.

Other references

- Bahns, J. and M. Eldaw. 1993. Should We Teach EFL Students Collocations? *System* 21(1): 101-114. [https://doi.org/10.1016/0346-251X\(93\)90010-E](https://doi.org/10.1016/0346-251X(93)90010-E). Accessed 17 November, 2022.
- Blaxter, L., C. Hughes and M. Tight. 2006. *How to Research*. Third edition. Maidenhead: Open University Press.
- Bogaards, P. 1998. Scanning Long Entries in Learner's Dictionaries. Fontenelle, T., P. Hilgsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *EURALEX '98 Actes/Proceedings*: 555-563. Liège: English and Dutch Departments, University of Liège.
- Brace, I. 2018. *Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research*. Fourth edition. London: Kogan Page.

-
- Chan, A.Y.W.** 2012. Cantonese ESL Learners' Use of Grammatical Information in a Monolingual Dictionary for Determining the Correct Use of a Target Word. *International Journal of Lexicography* 25(1): 68-94. <https://doi.org/10.1093/ijl/ecr014>. Accessed 28 January, 2023.
- 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. <https://doi.org/10.1093/ijl/ecq013>. Accessed 16 January, 2023.
- Chen, Y.** 2017. Dictionary Use for Collocation Production and Retention: A CALL-based Study. *International Journal of Lexicography* 30(2): 225-251. <https://doi.org/10.1093/ijl/ecw005>. Accessed 5 January, 2023.
- Daskalovska, N.** 2015. Corpus-based versus Traditional Learning of Collocations. *Computer Assisted Language Learning* 28(2): 130-144. <https://doi.org/10.1080/09588221.2013.803982>. Accessed 9 April, 2023.
- Debois, S.** 2019. 10 Advantages and Disadvantages of Questionnaires. <https://surveyanyplace.com/blog/questionnaire-pros-and-cons/>. Accessed 7 April, 2023.
- 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. <https://doi.org/10.1093/ijl/ecu012>. Accessed 23 September, 2022.
- Frankenberg-Garcia, A.** 2020. Combining User Needs, Lexicographic Data and Digital Writing Environments. *Language Teaching* 53(1): 29-43. <https://doi.org/10.1017/S0261444818000277>. Accessed 21 December, 2022.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415. <https://doi.org/10.1093/ijl/ecw030>. Accessed 22 February, 2023.
- 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. <https://doi.org/10.1093/ijl/9.4.321>. Accessed 17 December, 2022.
- Herbst, T.** 2010. Valency Constructions and Clause Constructions or how, if at all, Valency Grammars might Sneeze the Foam off the Cappuccino. Schmid, H.-J. and S. Handl (Eds.). 2010. *Cognitive Foundations of Linguistic Usage Patterns*: 225-256. Berlin/Boston: De Gruyter Mouton.
- Heuberger, R.** 2020. Monolingual Online Dictionaries for Learners of English and the Opportunities of the Electronic Medium: A Critical Survey. *International Journal of Lexicography* 33(4): 404-416. <https://doi.org/10.1093/ijl/ecaa018>. Accessed 18 January, 2023.
- Jackson, H.** 2018. English Lexicography in the Internet Era. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 540-553. New York: Routledge.
- Jackson, C.J. and A. Furnham.** 2000. *Designing and Analysing Questionnaires and Surveys. A Manual for Health Professionals and Administrators*. London: Whurr Publishers.
- Knežević, L., S. Halupka-Rešetar, I. Miškeljin and M. Milić.** 2021. Millennials as Dictionary Users: A Study of Dictionary Use Habits of Serbian EFL Students. *SAGE Open* 11(2): 1-11. <https://doi.org/10.1177/21582440211008422>. Accessed 29 March, 2023.
- 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. <https://doi.org/10.1093/ijl/ecy022>. Accessed 15 February, 2023.

-
- Laufer, B.** 2011. The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. *International Journal of Lexicography* 24(1): 29-49. <https://doi.org/10.1093/ijl/ecq039>. Accessed 14 January, 2023.
- Laufer, B. and T. Waldman.** 2011. Verb–Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61(2): 647-672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>. Accessed 13 February, 2023.
- 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 12–17, 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: Motiwex.
- Lew, R.** 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306. <https://doi.org/10.5788/20-0-144>. Accessed 1 January, 2023.
- 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.
- Lew, R. and P. Tokarek.** 2010. Entry Menus in Bilingual Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications*: 193-202. Louvain-la-Neuve: Cahiers du CENTAL.
- L'Homme, M.-C. and M.C. Cormier.** 2014. Dictionaries and the Digital Revolution: A Focus on Users and Lexical Databases. *International Journal of Lexicography* 27(4): 331-340. <https://doi.org/10.1093/ijl/ecu023>. Accessed 22 November, 2022.
- 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 7 April, 2023.
- Mackey, A. and S.M. Gass.** 2005. *Second Language Research: Methodology and Design*. Mahwah, New Jersey: Lawrence Erlbaum.
- Nuccorini, S.** 1992. Monitoring Dictionary Use. Tommola, H., K. Varantola, T. Salmi-Tolonen and J. Schopp (Eds.). 1992. *EURALEX '92 Proceedings. Papers Submitted to the 5th EURALEX International Congress, Tampere 1992*: 89-102. Tampere: Department of Translation Studies, University of Tampere.
- Ptasznik, B.** 2015. *Signposts and Menus in Monolingual Dictionaries for Learners of English*. Poznań: Wydawnictwo Naukowe UAM.
- Ptasznik, B.** 2022. After the Digital Revolution: Dictionary Preferences of English Majors at a European University. *Lexikos* 32: 220-249. <https://doi.org/10.5788/32-1-1721>. Accessed 31 January, 2023.
- Rundell, M.** 2012. It Works in Practice but Will it Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the XV EURALEX International Congress, 7–11 August, 2012, Oslo*: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)
<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1799> (Lexicosurvey)

122 Bartosz Ptasznik

Summers, D. 1988. The Role of Dictionaries in Language Learning. Carter, R. and M. McCarthy (Eds.). 1988. *Vocabulary and Language Teaching*: 111-125. London: Longman.

Webb, S. and E. Kagimoto. 2011. Learning Collocations: Do the Number of Collocates, Position of the Node Word, and Synonymy Affect Learning? *Applied Linguistics* 32(3): 259-276. <https://doi.org/10.1093/applin/amq051>. Accessed 9 April, 2023.

APPENDIX A: Original text of the questionnaire

ANKIETA

Dziękuję za udział w ankiecie. Ankieta składa się z dwóch części i jest anonimowa. Proszę udzielić odpowiedzi na następujące pytania:

CZĘŚĆ 1

1. Jak oceniasz użyteczność *Dictionary page* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
2. Jak oceniasz użyteczność *Macmillan Collocations Dictionary* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
3. Jak oceniasz użyteczność *Macmillan Thesaurus* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
4. Jak oceniasz użyteczność *Videos* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
5. Jak oceniasz użyteczność *Macmillan Dictionary Blog* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
6. Jak oceniasz użyteczność *Language Quizzes, Games, Puzzles* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
7. Jak oceniasz użyteczność *Open Dictionary* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
8. Jak oceniasz użyteczność *Buzzwords* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
9. Jak oceniasz użyteczność *Trending Words* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny
10. Jak oceniasz użyteczność *Red Words and Stars* w skali 1–7?
Mniej użyteczny – 1–2–3–4–5–6–7 – Bardziej użyteczny

CZĘŚĆ 2

1. Które cechy słownika *Macmillan Dictionary* są według Ciebie najbardziej pożytecznymi do nauki języka angielskiego? Dlaczego?
2. Które cechy słownika *Macmillan Dictionary* są według Ciebie najmniej pożytecznymi do nauki języka angielskiego? Dlaczego?

APPENDIX B: English translation of the questionnaire

QUESTIONNAIRE

Thank you for taking the time to fill out the questionnaire. The questionnaire consists of two parts and is anonymous. Please answer the following questions:

PART 1

1. How do you rate the usefulness of the *Dictionary page* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
2. How do you rate the usefulness of the *Macmillan Collocations Dictionary* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
3. How do you rate the usefulness of the *Macmillan Thesaurus* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
4. How do you rate the usefulness of *Videos* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
5. How do you rate the usefulness of the *Macmillan Dictionary Blog* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
6. How do you rate the usefulness of *Language Quizzes, Games, Puzzles* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
7. How do you rate the usefulness of the *Open Dictionary* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
8. How do you rate the usefulness of *Buzzwords* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
9. How do you rate the usefulness of *Trending Words* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful
10. How do you rate the usefulness of *Red Words and Stars* on a scale of 1–7?
Less useful – 1 – 2 – 3 – 4 – 5 – 6 – 7 – More useful

PART 2

1. Which features of the *Macmillan Dictionary* are the most useful learning tools in your opinion? Why?
2. Which features of the *Macmillan Dictionary* are the least useful learning tools in your opinion? Why?

John Considine. *Sixteenth-Century English Dictionaries*. 2022, xiv+482 pages. ISBN 978-0-1988-3228-7 (Hardback). Oxford: Oxford University Press. Price £90.00.

This volume is the first in the trilogy *Dictionaries in the English-speaking World, 1500–1800* published by Oxford University Press. The titles of the three volumes in this series — all written by John Considine — are rather straightforward: *Sixteenth-century English Dictionaries*, *Seventeenth-century English Dictionaries*, and *Eighteenth-century English Dictionaries*. Each of the three books examines the lexicographical landscape of one century as their titles indicate. In a previous work, Considine divides the development of English lexicography into four phases, and he calls the fifteenth and sixteenth centuries a 'prologue' (2020: 90). He expands this 'prologue' into a book-length study, in which lexicographical works and traditions of the sixteenth century play a major role, but those of the fifteenth century and earlier periods are inevitably and necessarily mentioned.

Chapter 1 'The medieval inheritance' brings up the curtain on medieval lexicography with a range of lexicographical works available in the British Isles at the beginning of the sixteenth century — though most of them were compiled in or before the fifteenth century. 'The medieval inheritance' includes a collection of Latin wordlists in the library at Syon Abbey; some very early English glosses in Latin dictionaries and further copies that formed the *Medulla grammaticae* tradition; the *Catholicon Anglicum* and the *Promptorium parvulorum*, which reversed the *Medulla* tradition and provided Latin equivalents to the English headwords; and representatives of the native Irish and Welsh lexicographical traditions.

Chapter 2 turns from incunabular glossarial manuscripts to 'The first printed dictionaries of English, French, and Latin', in which the medieval traditions were largely maintained. The first English wordlists can be found in the booklets printed by William Caxton, and the circulation of these and related wordlists reflects 'a late fifteenth-century enjoyment of linguistic ornament ... and the needs of a fraction of late fifteenth-century English society' (p. 28). This chapter also reveals how the publication of different editions of the *Promptorium parvulorum* and the *Ortus vocabulorum* was involved in the competition between Wynkyn de Worde and Richard Pynson, the two leading printers in early sixteenth-century England. In the rest of the chapter, Considine mentions various smaller and specialized wordlists, including pedagogical, medical, and herbal ones.

Chapter 3 starts with John Palsgrave's *Lesclarcissement de la langue francoyse*, the first large lexicographical work compiled in the sixteenth century, introducing French grammar but written in English. The sense of using English as a metalanguage was widely developed since the 1520s, with four more examples discussed in this chapter: John Rastell's *Exposiciones terminorum legum Anglorum* (c. 1523) combined legal French and English, though the latter was put in a secondary position as a translation of the former; Nicholas Udall's *Floures*

for *Latine spekynghe* (1534) furnished Latin drama phrases (in roman type) with idiomatic English translations (in black letter); two thin books by William Turner, namely *Libellus de re herbaria nouus* (1538) and *Names of herbes* (1548), pioneeringly used English to explain English names of plants; *Breuiary of helthe* (1547) by Andrew Boorde was within an ace of being the earliest medical dictionary with English as its defining language, because it provided too much encyclopaedic information. Strictly speaking, none of 'Palsgrave and some contemporaries' mentioned in Chapters 3 can be treated as dictionaries, but they do contain some features of dictionaries and valuable lexical information.

As a matter of fact, the word 'dictionary' in English has never appeared to define a type of scholarly work until 1538, when 'The *Dictionary* of Sir Thomas Elyot' (the subject of Chapter 4) was published. Elyot's *Dictionary* is an abridged translation of Ambrogio Calepino's *Dictionarium*, an alphabetized, comprehensive, and humanistic dictionary of Latin (also mentioned in Chapter 1). Considine suggests that Elyot's remarkable achievement of English lexicography was 'to carry out this demanding and very useful translation project quickly and accurately, and to see the results into print' (p. 85; see also Stein 2014 for a book-length study on Elyot). All of Elyot's books, including the *Dictionary* of 1538, the 1542 and 1545 editions with the new title *Bibliotheca Eliotae: Eliotis librerie*, and successive editions in 1548, 1552, 1559 published after Elyot's death, were printed by Thomas Berthelet. The same printer ensured continuity in printing. For example, Berthelet set the Latin words in roman type and the English in black letter for Elyot's *Dictionary*, just as he had done for Udall's *Floures* in 1534; he also used the same ornamental title page border for all editions as a marketing strategy, though Considine notices a subtle difference, that is, only the 1545, 1548, and 1552 editions had a hand-coloured border (for more information on typography in sixteenth-century English dictionaries, see Stein 2017).

The next two chapters introduce a diversity of multilingual and bilingual dictionaries. All of the 'Polyglot dictionaries' discussed in Chapter 5 included English text but followed different continental European traditions. Some notable examples are: the thematically arranged *Sex linguarum dictionarius* (1537) in the *Introito e porta* tradition; *Colloques ou dialogues avec un dictionnaire en six langues* (1576) in the Barlement tradition, having an alphabetical wordlist after dialogues; the subject-ordered *Nomenclator* (1567) of Hadrianus Junius; and polyglot dictionaries in the Calepino tradition, which frequently gave English words with equivalents in other languages, ancient and modern.

Chapter 6 gives an account of 'Bilingual dictionaries of vernacular languages in the 1540s and 1550s'. To begin with, William Salesbury's *A dictionary in Englyshe and Welsh* (1547), with over 7000 entries, was the one and only dictionary of Welsh printed in the sixteenth century, and also the first book-length dictionary of English and another vernacular language. *Principal rules of the Italian grammer, with a dictionarie for the better vnderstandyng of Boccace, Petrarcha, and Dante* (1550), also by a Welshman called William Thomas, was said to be 'the first bilingual English dictionary to respond to the bilingual lexicography

of the continental European Renaissance' (p. 123). Short wordlists in other languages, such as Spanish, Breton, Romani, and Taino, are briefly discussed in the latter part of Chapter 6, and they indeed widened the scope of English lexicography.

Chapter 7 is concerned with 'Latin dictionaries of the 1550s', most of which were for pedagogical use and thus in a size appropriate for schoolboys — smaller than weighty folios but larger than polyglots. A noteworthy feature of this chapter is that the mid-16th-century dictionaries discussed are cross-referenced to other parts of the book. For instance, further editions of Jean Véron's *Dictionariolum puerorum* (1552), Richard Howlet's *Abcedarium Anglicolatinum* (1552), and John Withals' *Shorte dictionarie for yonge begynners* (1553) will be discussed in Chapter 10; Laurence Nowell's annotations on Howlet's *Abcedarium* can be found in Chapter 15; the whole story of John Barrett's *Alvearie* (1574) will be told in Chapter 10, but the 1550s was when the seeds of this dictionary germinated.

As illustrated in Chapter 4, Elyot had trodden a path to Latin–English lexicography, but as time went by discontinuities occurred in the series of editions. Chapter 8 focuses exclusively on 'The *Thesaurus* of Thomas Cooper' — the first English book under the title *Thesaurus* — its disputed authorship, its copious material, its mixed reputation, and in particular, the contrast between the 1559 *Bibliotheca* and the 1565 *Thesaurus*.

In Chapter 9, Considine returns to the subject of vernacular dictionaries after an interval of two chapters, which also indicates a lull in the vernacular dictionary market. It was not until 1570 that a batch of dictionaries of vernacular languages came onto the market, including those of French, of Irish, of Spanish, of Russian, and of exotic languages of the New World, Africa, the Arctic, the Caribbean, and South East Asia. All of the 'Elizabethan dictionaries of vernacular languages before Florio' mentioned in Chapter 9 narrate the adventures of 'the discovery by Anglophones of the linguistic diversity of the world' (p. 203).

In Chapter 10, Considine continues the story by giving an account of new, revised, or translated versions of the works which originated in the 1550s. The 'Dictionaries of Latin from 1565 to 1580' in this chapter include: four editions of Withals's *Shorte dictionarie* since 1568, printed by Thomas Purfoote and edited by Lewys Evans; Peter Levins's *Manipulus vocabulorum* (1570); John Higgins's revision turning Howlet's *Abcedarium* into a triglot; John Barrett's first edition of *Alvearie* (1574), in which he used the so-called 'box rules' to make the dictionary 'readable', 'harmonious', and 'distinctive' (p. 224); a new edition of Véron's *Dictionariolum* by Ralph Waddington published in 1575; an English-language version of Simon Pelgrom's *Synonymorum sylua* (1580); and the second edition of Barrett's *Alvearie* by Abraham Fleming (1580). These lexicographical works, together with the ones discussed in Chapter 9, present the readers with a fuller picture of Elizabethan lexicography.

Chapters 11–13 deal with dictionaries in the last several decades of the

sixteenth century. The last fragment of the story of Latin dictionaries in the sixteenth-century Isles can be found in Chapter 11, accompanied by a brief account of Greek dictionaries from 1581 onwards. New editions of Hadrianus Junius's *Lexicon Graecolatinum* and *Nomenclator* were undertaken by Edward Grant and John Higgins respectively, while Abraham Fleming provided a new index to the latter. Recognized to be 'the first professional lexicographer in England' twice in the book (pp. 229, 258), Fleming's lexicographical publications — remarkably done within six years — include the indexing of the Higgins–Junius *Nomenclator*, as well as the making of Barrett's *Alvearie*, Guillaume Morel's *Verborum Latinorum*, Withals's *Shorte dictionarie*, and Waddington's version of Véron's *Dictionariolum*. The 1580s also saw a major change in England's dictionary publishing. Presses at universities were established and began to take a share of the dictionary market. Thomas Thomas's Latin–English *Dictionarium linguae Latinae et Anglicanae* (1587) was attributed to Cambridge, while its English–Latin counterpart, John Rider's *Bibliotheca scholastica* (1589), was an Oxford book.

The following chapter is concerned with 'Wordlists with Hebrew, Arabic, and Armenian'. Wordlists of these so-called 'learned languages' were rarely published in the Isles in the first half of the sixteenth century, 'a century when most of the wordlists undertaken in the Isles were directed at an insular readership' (p. 297). Whereas several Armenian and Hebrew dictionaries are touched on in this chapter, dictionaries of Arabic were not published in the sixteenth century, for instance William Bedwell's dictionary work remained in manuscript.

The story of the sixteenth-century dictionaries of English with vernacular languages come to an end in Chapter 13. Two remarkable examples given here are 'The dictionaries of Florio and Minsheu'. The Italian–English *World of wordes* of John Florio and the Spanish–English *Dictionarie in Spanish and English* of John Minsheu are remarkable in their size — both of them comprised over 40,000 entries. It can be concluded from chapters 6, 9, and 13 that English speakers of the sixteenth century had access to a wide range of dictionaries of the major European vernaculars. In comparison, wordlists of some minor European languages available to them were fewer and shorter, and those of some exotic languages had not been developed into the form of dictionaries yet.

Each of the final three chapters examines kinds of wordlists produced for specific purposes. Chapter 14 is devoted to 'Specialized wordlists of English after the 1530s', which appeared as glossaries appended to the books of specific subjects. Therefore this chapter is both chronological and subject-ordered: it begins with glossaries of medical vocabularies, which were 'in many ways at the vanguard of English dictionary-making in the mid to late sixteenth century' (Tyrkkö 2011, as cited in Considine 2022: 330); legal wordlists; followed by those of rhetorical terms, geometrical and cosmographical terms, proper names, cant words; then the vocabularies of logic, herb-lore, hunting and war; and finally glossarial entries for the realia of ancient Rome.

The ultimate two chapters turn to wordlists made for non-technical words.

Chapter 15 looks at 'Lists of old words', to some extent indicating the sixteenth-century speakers' interest in older varieties of English, i.e. Old English and (early and late) Middle English. Laurence Nowell and John Joscelyn were the leading figures in 'Saxon' lexicography. Wordlists which included late Middle English, in particular Chaucerian glossaries, were among the early sources of English historical lexicography. Chapter 16 then provides a detailed overview of 'Lists of hard words, and of words in general use'. The first part of this chapter mainly focuses on the production of lists of hard words, appearing in the Isles in an attempt to support English readers of the Bible and other Christian texts. The second focus of Chapter 16 is on a variety of general wordlists for special purposes: for children learning, for spelling reform, for secret writing, and for poetic composition.

Although it is not clearly stated in this volume, the order of the chapters can be worked out. The whole volume is primarily chronological, and the relevant dictionaries are thematically arranged and discussed together. This may lead to the fragmentation of chapters, especially in the middle part of the book. For instance, chapters 4, 7, 8, 10, and 11 tell a coherent story of Latin dictionaries in the Isles, but other chapters distract readers from this topic for a while. Likewise, Withals's *Shorte dictionarie* and its later editions are scattered in different parts of the volume, as is the case in chapters 7, 10, and 11. Even within a single chapter, this kind of interruption can be detected, for example, the discussion on the two editions of Barrett's *Alvearie* in Chapter 10 is temporarily broken off by looking at several other dictionaries. Causing confusion and discontinuity, however, is only part of such an arrangement of chapters. On the other hand, more creative alternatives can be offered to readers. A similar approach can be found in *The Cambridge World History of Lexicography* (2019), edited by the same author, which allows readers to search through the book in at least three ways.

On the whole, this is a thoughtfully structured and well-written monograph as a 'prologue' to the fascinating history of lexicographical works. Considine's book not only tells us the story of dictionaries, but also the story of people. In almost all chapters, there are biographical notes about those dictionary related people, including but not limited to lexicographers, printers, publishers, translators, and readers. Also frequently occurring in many chapters are the references to the sequel to this volume. As Considine reveals in the current volume, an enlarged edition of the *Shorte dictionarie*, the revised *Riders dictionarie*, more publications by Sturtevant, and Bedwell's Arabic dictionary, will be discussed in his *Seventeenth-century English Dictionaries*. Readers will certainly be looking forward to reading the second volume in the series.

References

- Considine, J. (Ed.). 2019. *The Cambridge World History of Lexicography*. Cambridge: Cambridge University Press.

- Considine, J.** 2020. A Dictionary Ecosystem: Four Centuries of English Lexicography. Ogilvie, S. (Ed.). 2020. *The Cambridge Companion to English Dictionaries*: 89-100. Cambridge: Cambridge University Press.
- Stein, G.** 2014. *Sir Thomas Elyot as Lexicographer*. Oxford: Oxford University Press.
- Stein, G.** 2017. *Word Studies in the Renaissance*. Oxford: Oxford University Press.
- Tyrkkö, J.** 2011. "Halles Lanfranke" and Its Most Excellent and Learned Expositive Table. Timofeeva, O. and T. Säily (Eds.). 2011. *Words in Dictionaries and History. Essays in Honour of R.W. McConchie*: 17-40. Amsterdam/Philadelphia: John Benjamins.

Ai Zhong
College of Foreign Languages
Shanghai Maritime University
Shanghai
China
(zhongai@shmtu.edu.cn)

Annette Klosa-Kückelhaus and Ilan Kernerman (Eds.). *Lexicography of Coronavirus-related Neologisms*. 2022, vi + 306 pages. ISBN 978-3-1107-9556-1 (Hardcover); 978-3-1107-9808-1 (ePDF); 978-3-1107-9831-9 (ePUB). Lexicographica. Series Maior 163. Berlin/Boston: Walter de Gruyter. Price €109.95 (Hardcover), online version (ePDF/ePUB).

Neology has been an integral component of language since its inception, serving as a driving force and motivating factor for the field of lexicography. The lexicons of all languages are constantly developing. The same is the case with coronavirus-related neologisms. Coronavirus-related neologisms have emerged in response to the unprecedented pandemic situation, revealing the adaptability of language to reflect new social realities. These words not only focus on a specific period and topic but also display distinctive tendencies in grammar and semantics. They provide a rich area of study for lexicographers interested in word formation, the relationship between language and discourse communities, and the identification of neologisms associated with the pandemic.

The Lexicography of Coronavirus-related Neologisms presents a comprehensive and cutting-edge exploration of coronavirus-related neologisms. It contains 13 research papers that investigate the study of neology and lexicography in an extensive variety of languages, including English, Korean, Hungarian, German, Italian, Portuguese, Spanish, Croatian, and the New Zealand Sign Language. These papers cover a wide range of topics related to lexicography, with a particular focus on research concerning neologisms that have emerged from the COVID-19 pandemic, as well as innovative approaches to handling such neologisms in various types of dictionaries.

This collection of papers showcases the cross-cultural perspectives of lexicographers from diverse linguistic backgrounds. In the introduction, the editors divide the papers into four groups, which discuss various lexical resources. The first group examines coronavirus-related neologisms in English, German and Korean. The second group focuses on coronavirus-related neologisms in the Spanish language. The third group investigates lesser-used languages. The final group explores COVID-19 neology in relation to collaborative editing in Wiktionary and New Zealand Sign Language.

The first group's research investigates the impact of the COVID-19 pandemic on lexicography, as well as the strategies employed to overcome the challenges presented by the rapid influx of neologisms. Based on their findings, the authors analyze the effects of the pandemic on lexicographic methods and the emergence of new neologisms in English, German and Korean.

The paper 'The Oxford English Dictionary and the language of Covid-19' examines how the editorial team of the Oxford English Dictionary (OED) responded to the lexical adjustments necessitated by the global health crisis. The research explains how the OED utilized the Oxford Linguistics' Monitor English Corpus and other textual databases to monitor the development of coronavirus-related vocabulary. Furthermore, it explains how the detection and

processing of new words and words with new meanings were accomplished by revising existing entries and introducing new ones.

The adoption of both retrospective and proactive approaches to record coronavirus-related neologisms is explored in the paper 'German Corona-related neologisms and their lexicographic representation'. It investigates how the Leibniz Institute for the German Language (IDS) has implemented these approaches to capture and document coronavirus-related neologisms. This integration of approaches allows for comprehensive coverage of COVID-19 neologisms in the German language, providing information on new lexemes, meanings, and usages throughout the pandemic's development.

The Korean researchers have contributed a paper titled 'The emergence and spread of Korean COVID-19 neologisms in news articles and user comments and their lexicographic description', which focuses on the occurrence and spread of Korean COVID-19 neologisms in news articles and user comments, as well as their dictionary definitions. It examines the frequency of occurrence and usage trends of these neologisms, and discusses their implications for lexicographic research. The findings indicate that comment data is invaluable for the lexicographic description of neologisms, as it provides pragmatic information on and the socio-cultural background of neologisms. Comments reflect the emotions and stances of a multitude of commenters on relevant neologisms, offering dictionary users and future generations fresh, raw examples of real-life language for neologisms.

The papers in the second group analyze coronavirus-related neologisms in the Spanish language from various perspectives, such as identification, classification, inclusion in dictionaries, and their usage in technical fields.

'Lexicographic detection and representation of Spanish neologisms in the COVID-19 pandemic' presents a comprehensive analysis of coronavirus-related neologisms in the Spanish language. This paper shows the diverse range of COVID-19 neologisms in Spanish, some of which are fleeting and transitory, referred to as "non-dictionarizable neologisms" (p. 80), while others gain acceptance and become part of the standard language through lexicalization, known as "dictionarizable neologisms" (p. 84). It emphasizes the critical role that lexicography plays in the development of neologisms. On the one hand, it helps to fix the usage of neologisms in the language, on the other hand, it serves as a criterion for lexical evolution.

This discussion is further developed in the paper titled 'Spanish neologisms during the COVID-19 pandemic: Changing criteria for their inclusion and representation in dictionaries'. It investigates the inclusion of coronavirus-related neologisms in various dictionaries, including general language dictionaries of Spanish, a bilingual English-Spanish dictionary, and a Spanish neologism dictionary. It delineates the criteria for the inclusion and treatment of the neologisms in these dictionaries and examines the limitations and advantages of exclusively gathering data from written press sources.

The final paper in this group, titled 'Specialized voices in the 23rd edition

of the *Diccionario de la lengua española*: Analysis of the COVID-19 field and its neologisms', focuses on the incorporation of technical neologisms related to COVID-19 into the authoritative Spanish language dictionary the *Diccionario de la Lengua Española* (DLE). The paper discusses the rapid expansion of COVID-19 vocabulary and its implications for lexicography and terminology research. This expansion is viewed as an opportunity for researchers to propose consistent solutions for including scientific and specialized words in the DLE and other Spanish dictionaries. The authors suggest abandoning ambiguous standards for the inclusion or exclusion of scientific vocabulary, technical definitions that are not readily understandable by a general readership, imprecisions in conceptualization, and erratic ascriptions of thematic designations.

The third group of papers investigates lesser-used languages, such as Hungarian, Croatian and Portuguese in Portugal and Brazil, uncovering analogous issues that exist in more commonly used European languages.

'How the COVID-19 pandemic is changing the Hungarian language: Building a domain-specific Hungarian/Italian/English dictionary of the COVID-19 pandemic' describes the process of compiling the Trilingual COVID-19 Dictionary (TCD), which includes Hungarian, Italian and English. The process of creating the TCD covers various aspects, including corpus selection, terminological extraction, a draft dictionary, and formatting. The TCD also contains morphological analysis of neologisms, which provides insight into the trends and patterns in neologism formation and frequency of use in Hungarian. The analysis of Hungarian coronavirus-related neologisms indicates that compounding and derivation are the most frequent methods of neologism formation, but syntagms, blending and semantic extension are also used. It explores the role of English in neologism creation, as well as loanwords, calques, and adaptations in Hungarian and Italian.

In 'Coronavirus-related neologisms: A challenge for Croatian standardology and lexicography', the discussion centers on the question of determining which coronavirus-related neologisms, collected from media corpora and online sources, should be included in general language dictionaries for Croatian. In this paper, the researchers distinguish between Croatian neologisms and loanwords/loan translations, while stressing the importance of providing prescriptive information in response to the high number of user questions regarding orthography, morphology, word formation, usage in sentences, and meaning for all types of neologisms.

The formation and incorporation of coronavirus-related neologisms in Portuguese media discourse, as well as their inclusion in dictionaries, are explored in the other papers of this group.

Focusing on neologisms that incorporate the components "coronavirus", "COVID-19", "pandemic" and the prefix "tele-", 'The neologisms of the COVID-19 pandemic in European Portuguese: From media to dictionary' provides various examples of how these neologisms are integrated into novel morphological structures. The researchers investigate the emergence of coronavirus-related

neologisms in press and social media, evaluate their inclusion in European Portuguese dictionaries, and affirm the crucial role of lexical neology in the COVID-19 domain.

A corpus-based investigation of Covid-19 terminology in Brazilian Portuguese is presented in the paper 'COVID-19 terminology and its dissemination to a non-specialized public in Brazil'. It analyzes and discusses the characteristics of the terminology, including the frequent use of the adjective "novo" (meaning "new"), the pluralization of terms, and the resemantization of some terminological units. This research underscores the need to compile a dictionary for non-specialized readers in the medical field.

'Neoterm or neologism? A closer look at the determinologisation process' examines the formation, categorization, and lexicographic description of new lexical units in Portuguese media discourse. The paper specifically investigates words created with the prefix "covid" and raises the question of whether these words should be classified as neoterms. It takes a further look at the challenges of incorporating new terms into the lexicon, including determining their definitions and appropriate domain labels.

The examination of COVID-19 neology in relation to New Zealand Sign Language and collaborative editing in Wiktionary are the topics of the fourth group of papers, which extend beyond mainstream lexicography.

'Neologisms in New Zealand Sign Language: A case study of COVID-19 pandemic-related signs' analyzes the neologisms that arose in New Zealand Sign Language as a response to the pandemic. The paper investigates how these neologisms can be incorporated into the *Online Dictionary of New Zealand Sign Language*. According to the research findings, the newly formed signs were created through diverse strategies such as semantic extension, derivation or compounding, and borrowing from other sign languages. Nevertheless, incorporating these signs into the dictionary also presents challenges since the signs must be standardized and widely adopted by the Deaf community.

The final paper in the book titled 'Using Wiktionary revision history to uncover lexical innovations related to topical events: Application to Covid-19 neologisms' presents a novel approach to identify and monitor new lexical innovations that are linked to topical events. The method is based on analyzing the Wiktionary revision histories of extensively modified articles from the active online communities of the English and French versions of Wiktionary. The results of the study suggest that this method is effective in tracking lexical innovations and can provide useful insights into the linguistic responses to current events.

The Lexicography of Coronavirus-related Neologisms is an excellent illustration of advanced research in European countries regarding coronavirus-related neologisms, featuring relevant and well-contextualized research approaches. The book presents research papers authored by experts with diverse linguistic backgrounds, exhibiting exceptional writing skills and professionalism in the field of lexicography. To enhance comprehension, the papers are furnished

with ample examples, pictures and figures. If the book were to be reprinted in the future, I would suggest adding chapter numbers to the table of contents. This would make it easier for readers to navigate through the book and find specific chapters they are interested in. Nevertheless, *The Lexicography of Coronavirus-related Neologisms* is a valuable resource for professional researchers, teachers and students of linguistics and lexicography worldwide. The findings of these papers may provide inspiration and ideas for the study of neologism, especially in the field of lexicographic compilation.

Acknowledgements

This study was funded by the Hebei Provincial Fundamental Research Projects for Higher Education Institutions (Project No. JSQ2021003).

Jian Zhou
North China University of Science and Technology
Tangshan
Heilongjiang University
Harbin
China
(zhou.jian@mail.ru)

and

Shuliang Dong
North China University of Science and Technology
Tangshan
China
(Corresponding Author, shuliangly@163.com)

The Future of Metalexicography: Reaching for the Mesosphere

Gilles-Maurice de Schryver, *BantUGent & LT³, Ghent University, Ghent, Belgium; and Department of African Languages, University of Pretoria, Pretoria, South Africa (gillesmaurice.deschryver@UGent.be; <https://orcid.org/0000-0001-7272-9878>)*

Abstract: In this research article, a quantified look is taken at the metalexicographic endeavours of the past half century, starting in 1971. It is argued that the year 2021 represents a tipping point, and for that reason, when illustrations are needed in the article, those are preferably taken from the work of Sue Atkins, who passed away in 2021. It analyses the formation of the various continental lexicography associations, as well as their conferences and linked proceedings, and ends with the current and future role of the global lexicographic alliance. In addition, a comparative bibliometric study is undertaken of the four main journals of our field. It is shown that the number of metalexicographic studies and the impact these have continued to grow, to the point where dedicated tools and databases are currently needed for the efficient examination and use of the now many thousands of conference papers, journal articles, and other publications. A brief section also deals with modern dictionary user research, which is, according to current thinking, subsumed under metalexicography. With and from this vast amount of data, the future direction of metalexicography is extrapolated. The undertaken research is thus very much data-driven, and refrains from thought experiments to arrive at that future.

Keywords: LEXICOGRAPHY, METALEXICOGRAPHY, DSNAL, EURALEX, AUSTRALEX, AFRILEX, ASIALEX, AMERICALEX-S, ELEX, GLOBALEX, CONFERENCE PROCEEDINGS, JOURNAL ARTICLES, DICTIONARIES: JOURNAL OF THE DSNAL, LEXICOGRAPHICA, INTERNATIONAL JOURNAL OF LEXICOGRAPHY, LEXIKOS, LEXICOGRAPHY: JOURNAL OF ASIALEX, BIBLIOMETRICS, GOOGLE SCHOLAR, DICTIONARY USE, ELEXIFINDER

Samenvatting: De toekomst van de metalexicografie: Op weg naar de mesosfeer. In dit onderzoeksartikel wordt een gekwantificeerde blik geworpen op de metalexicografische inspanningen van de afgelopen halve eeuw, beginnend in 1971. Er wordt beweerd dat het jaar 2021 een kantelpunt vertegenwoordigt, en om die reden maken de meeste illustraties in het artikel gebruik van het werk van Sue Atkins, die in 2021 overleed. Dit artikel analyseert de ontwikkeling van de verschillende continentale lexicografieverenigingen, evenals hun conferenties en gekoppelde conferentieverslagen, en eindigt met de huidige en toekomstige rol van de wereldwijde lexicografische alliantie. Daarnaast wordt een vergelijkende bibliometrische studie uitgevoerd van de vier belangrijkste tijdschriften uit ons vakgebied. Er wordt aangetoond dat de hoeveelheid metalexicografische studies en de impact die deze hebben blijven groeien, tot het punt waarop vandaag speciale tools en databases nodig zijn voor het efficiënt doorzoeken en gebruiken van de nu vele duizenden conferentiepapers, tijdschriftartikelen en andere publicaties. Een korte rubriek gaat ook over modern 'onderzoek naar het gebruik van woordenboeken', dat volgens de huidige opvat-

tingen onder metalexicografie valt. Met en uit deze enorme hoeveelheid gegevens wordt de toekomstige richting van de metalexicografie geëxtrapoleerd. Het uitgevoerde onderzoek is dus sterk data-gedreven en onthoudt zich van gedachte-experimenten om tot die toekomst te komen.

Sleuteltermen: LEXICOGRAFIE, METALEXICOGRAFIE, DSNA, EURALEX, AUSTRALEX, AFRILEX, ASIALEX, AMERICALEX-S, ELEX, GLOBALEX, CONFERENTIEVERSLAGEN, TIJD-SCHRIFTARTIKELEN, *DICTIONARIES: JOURNAL OF THE DSNA*, *LEXICOGRAPHICA*, *INTERNATIONAL JOURNAL OF LEXICOGRAPHY*, *LEXIKOS*, *LEXICOGRAPHY: JOURNAL OF ASIALEX*, BIBLIOMETRIE, GOOGLE SCHOLAR, WOORDENBOEKGEBRUIK, *ELEXIFINDER*

1. Metalexicography: From zero to the stratosphere in 50 years (1971–2021)

The discipline of metalexicography is generally considered to have started in earnest with the publication of the *Manual of Lexicography* by Ladislav Zgusta (1971). From humble beginnings only half a century ago, this discipline reached stratospheric heights in 2021. The year 2021 also literally represents a milestone: a first rocket was discarded and a second one took over. As is the case for so many other (if not most) aspects of human life, COVID-19 undeniably supercharged the change in gears. In the year 2021, we were treated to more (virtual) conferences, more research articles, and more innovative dictionaries (being the basis for more metalexigraphic research) than ever before. Even though some of it may be seen as the backlog from 2020 being released, metalexigraphers have now tasted the future, and the future is here to stay.

Given that metalexicography has now entered the future, we may contrast that 'present + future' with the 'past', as done in Table 1: the first line being a summary of the preceding paragraph; the next four lines summarising Sections 2 to 5 to follow.

Table 1: The discipline of metalexicography: From humble to stratospheric

§	Past	Present & Future
1	° with Zgusta in 1971, in a face-to-face world	entered the virtual world as of 2021; will become hybrid
2	from none to a few continental associations	associations for all continents, and an active world body
3	a few publications	massive amounts of research output
4	questionnaires and surveys to research dictionary use	unobtrusive logging and analysis of dictionary use
5	slow science via personal and public libraries	fast science accessible via dedicated databases and tools

Regardless of the fact that Grefenstette (1998) famously asked: "Will there be lexicographers in the year 3000?", which is of course a long shot, De Schryver (2023) established that there is still a chance for our dictionaries and our lexicographers to survive into the (near) future. If that is indeed the case, it also makes sense to look at the future of the scientific research devoted to lexicography, that is, the future of the discipline of metalexicography.

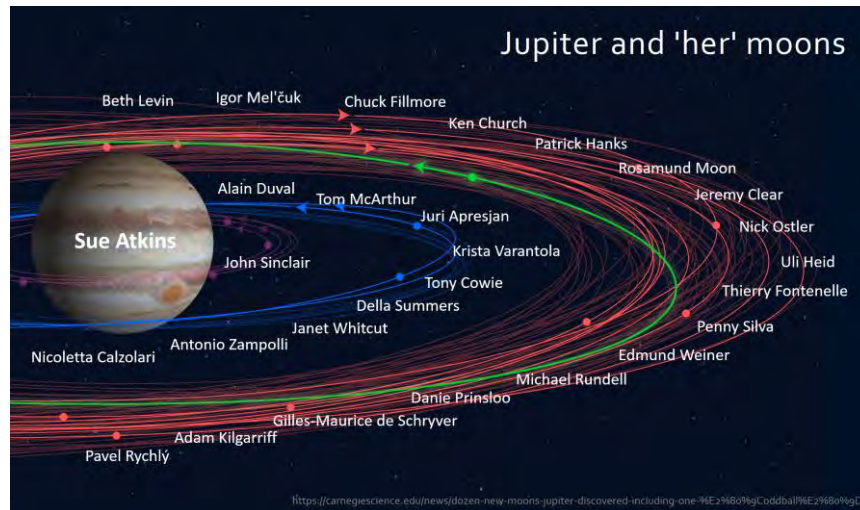


Figure 1: Visualisation of the past half century of (meta)lexicography, as seen through the eyes of Sue Atkins (1931–2021)

In what follows, various aspects of the metalexicographic endeavour will be looked at, and in places the metalexicographic repercussions of the *Grande Dame* of practical lexicography, Sue Atkins, who sadly passed away in the milestone-year 2021, will be studied. During the obituary read at the EURALEX congress the week following her passing, the then president of EURALEX, G.-M. de Schryver, urged every member of the lexicographic community to (re)watch her last interview (in conversation with Michael Rundell)¹ in which she masterfully surveyed the past half century of (meta)lexicography. The visualisation shown then, and reproduced in Figure 1, was an attempt to illustrate her interactions with all the (meta)lexicographers she mentioned during the interview. This will be addressed later.

2. Conferences of the continental associations for lexicography

Soon after the publication of Zgusta's manual, the first continental association for lexicography was formed, namely the DSNAL in 1975, for the North American continent. Over the next two decades, this effort to form continental associations (over and above local, national and regional ones) was copied on several other continents (see Table 2). EURALEX saw the light eight years later, AUSTRALLEX another seven years later, AFRILEX following five more years, and ASIALEX after a further two years. Key figures in setting up these associations were J. Edward Gates for the DSNAL, Reinhart Hartmann and Sue Atkins for EURALEX, Bill Ramson for AUSTRALLEX, Danie Prinsloo and Rufus Gouws for AFRILEX.

LEX, and Amy Chi and Gregory James for ASIALEX. For a very long time there was no association covering Central and South America, however, a problem pointed out over a decade ago in the metalexicographic literature (De Schryver 2012b: 482). An opportunity to do something about this situation arose at the DSN 2019 congress, where G.-M. de Schryver met with both Spanish and Portuguese-speaking lexicographers to discuss the ins and outs of establishing such an association. Three years later, in 2021, a working group was eventually formed with representatives from all over Latin America and the Caribbean, which baptised itself AMERICALEX-S. This working group will hold its first conference in October 2023, in São Paulo (Brazil), at which occasion the last continental association will be officially inaugurated.

Table 2: Continental associations for lexicography

Continent	Association	Acronym	Founded
North America	Dictionary Society of North America https://dictionarysofamerica.com/	DSNA	1975
Europe	European Association for Lexicography https://euralex.org/	EURALEX	1983
Oceania	Australasian Association for Lexicography https://www.australlex.org/	AUSTRALEX	1990
Africa	African Association for Lexicography https://www.afrilex.co.za/	AFRILEX	1995
Asia	Asian Association for Lexicography https://www.asialex.org/	ASIALEX	1997
Central & South America	Association of Lexicography for the Americas: South, Center, Caribbean and Mexico https://www.americlex.com/	AMERICALEX-S	(2023)
<i>n.a.</i>	<i>Electronic lexicography in the 21st century</i> https://elex.link/	<i>eLex</i>	2009
World	Global Alliance for Lexicography https://globalex.link/	GLOBALEX	2018

It is worth mentioning here the eLex conference series as well, first organised in Belgium by Sylviane Granger in 2009, which focuses exclusively on 'e-lexicography in the 21st century'. While not a continental association, given its focus, it certainly has the potential to go around the world (De Schryver et al. 2019a: 672-673). Cross-cutting these efforts, and merging Ilan Kernerman's wish to bring all of us together as 'Peoplex' (Kernerman 1997: 7) as well as the attempt to view all continents as one, hence as 'PangaeaLex' (De Schryver 2008: 113-114), a global alliance for lexicography was eventually set up in 2018, called GLOBALEX. For the first time in human history, then, *all* continents (save Antarctica (Adams 2014: 4)) will have their lexicographic association by the end of 2023, with a global alliance set to become ever more active.

Learned societies exist not just in name; their main activity is for their members to meet. As may be seen from Table 3, all continental associations have met regularly since their foundation, typically biennially, with only AFRILEX opting for an annual format, and ASIALEX having switched to that format recently. EURALEX purposely chose to organise its conferences in the off-years of the

DSNA, and then later eLex chose to organise theirs in the off-years of EURALEX. In half a century, over a hundred such conferences were organised, or thus two continental meetings per year on average. This average hides an important fact: Over time, there are simply ever more continental meetings, from just one every two years, to five annually, as shown in Figure 2.

Table 3: Conferences organised by the continental lexicography associations (up to 2022)

Association	Founded	Conferences	Proceedings?
DSNA	1975	biennial, in odd years (1977–2021): 23	No
EURALEX	1983	biennial, in even years (1983, 1986–2018, 2020>2021, 2022): 20	Yes
AUSTRALEX	1990	mostly biennial (1990–1998, 2000–2004, 2009–2021), some annual (1999, 2005–2008): 20	some online
AFRILEX	1995	annual (1996–2019, 2020>2021, 2022): 26	No
ASIALEX	1997	initially biennial (1997–2015), now annual (2016–2019, 2020>2021, 2022): 16	Yes
AMERICALEX-S	(2023)	t.b.a.	t.b.a.
eLex	2009	biennial, in odd years (2009–2021): 7	Yes
GLOBALEX	2018	workshops that co-locate with conferences: 4 with LREC & 4 with -lex	Yes for LREC

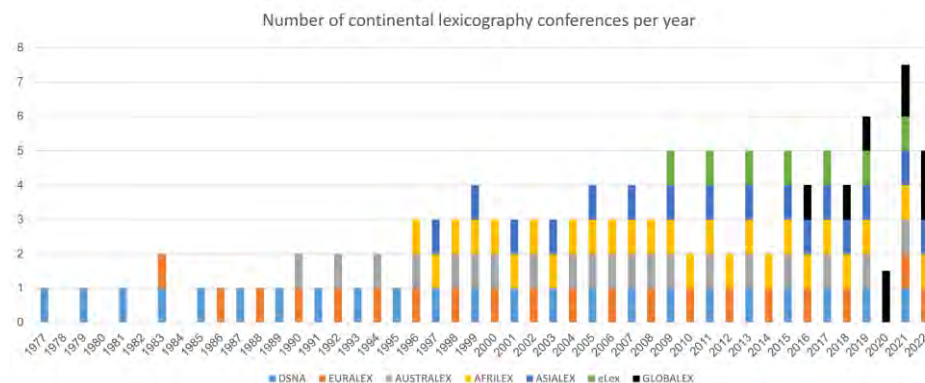


Figure 2: Number of continental lexicography conferences per year

Due to the COVID-19 pandemic, the year 2021 even saw the conferences of the DSNA, ASIALEX, AFRILEX, eLex, AUSTRALEX and EURALEX, as well as two GLOBALEX workshops (co-located with the last two), organised in a timespan of just four months. While there were very few colleagues who attended all eight meetings,² many colleagues attended several up to a handful. This was made possible by turning a necessity (the worldwide travel restrictions) into a virtue (move to a virtual format). Here is not the place to sing the praises of the virtual format — yes, it is comparable to the move from paper to digital dictionaries — but it may convincingly be argued that the future of our meetings

will be hybrid: Some of us will be physically present; others will attend via the streamed version; either in full or in part, including via the format of single-day registrations. As a matter of fact, EURALEX 2022 included a (free) hybrid day. All of this, then, means that lexicographers will have far more opportunities than ever to continue to meet in the future.

Another added advantage also became apparent: While previously only eLex had made a serious effort to record its presentations (at its physical meetings) to be streamed live as well as to be offered online after the event — at no charge, available to everyone —, in 2021 the presentations at each and every one of the continental lexicography conferences were recorded, and most of these also remain available to the world at large without any restrictions, as may be seen from Table 4. Going forward, Table 4 suggests that this will now become standard practice (cf. the entries for '2022').

Table 4: Recordings of the (2021) continental lexicography conferences (up to 2022)

Association	Rec < 2021?	Rec in 2021?	Online location, if any
DSNA	—	+	• 2021: all talks pre-recorded and shared with all participants after the congress
EURALEX	2018	+	• 2018: https://videolectures.net/euralex2018_ljubljana/ • 2020-2021: ³ https://videolectures.net/euralex2021/ • 2022: https://videolectures.net/euralex2022_mannheim/
AUSTRALEX	—	+	• 2021: recordings made during the congress but not yet shared
AFRILEX	—	+	• 2021: https://www.youtube.com/playlist?list=PLQbpi7xMa_B8Vfxtjg20d509L99LZa7g3 • 2022: https://www.youtube.com/playlist?list=PLQbpi7xMa_B40KX-4Yma0Z5VQqW4Skmw
ASIALEX	—	+	• 2021: https://drive.google.com/drive/folders/120H6oWP1Y29YFojiiy3-cp_A7ehDE36G
eLex	2011–2019	+	• 2011: https://videolectures.net/elex2011_bled/ • 2013: https://www.youtube.com/channel/UCDSa1Xot2bnZ03Y0mELxQJA/videos • 2015: https://www.youtube.com/channel/UCDSa1Xot2bnZ03Y0mELxQJA/videos • 2017: https://www.youtube.com/channel/UCDSa1Xot2bnZ03Y0mELxQJA/videos • 2019: https://videolectures.net/elexconference2019_sintra/ • 2021: https://www.youtube.com/channel/UCDSa1Xot2bnZ03Y0mELxQJA/videos
GLOBALEX	2020	+	• 2020: https://www.youtube.com/channel/UCxy9euhLou92uhf6LQqnMxA/videos • 2021: https://www.youtube.com/channel/UCxy9euhLou92uhf6LQqnMxA/videos

While the current picture may look a bit messy (YouTube channels, YouTube Playlists, a Google Drive, as well as VideoLectures) all this variation will surely converge, so that it will become ever easier in future to hear and see one another's talks and the discussions that ensued — both live and stored forever in the cloud. That, of course, does not mean that the video-format will replace the written-out texts that appear in conference proceedings.

3. Publications of the continental associations for lexicography

From Table 3 it is evident that EURALEX, ASIALEX and eLex religiously publish conference proceedings, AUSTRALEX and GLOBALEX only for some, with the DSNA and AFRILEX not at all. Even so, at the current rate, it still means that an average of about a thousand extra pages with conference papers on dictionaries,

dictionary makers and dictionary research appear each year — an amount that has been growing over the years as well. An analysis of the first fifteen EURALEX conference proceedings has shown (cf. De Schryver 2012a) that the impact of those publications also grows from conference to conference. Looking into the future, it seems there are no signs that this continuous upward trend will curb down.

As apparent in Table 5, all but one of the continental lexicography associations also publish their own journal: the DSNAL has *Dictionaries*, EURALEX has the *International Journal of Lexicography* (IJL), AFRILEX has *Lexikos*, and ASIALEX has *Lexicography*. In an early and beautiful example of cross-continent cooperation, the DSNAL and EURALEX also joined forces to produce the journal *Lexicographica* (in addition to the book series *Lexicographica. Series Maior*). All these journals started appearing some time after their respective associations were founded, except for *Lexikos*, a journal initially (and still) published by the *Bureau of the Woordeboek van die Afrikaanse Taal* (WAT) which became the mouthpiece of AFRILEX. GLOBALEX is the champion of cooperation, however. Four of its workshops took place at LREC (the biennial *International Conference on Language Resources and Evaluation*), in 2016, 2018, 2020 and 2022, so those workshop papers appeared in the LREC conference proceedings. While the GLOBALEX workshops that co-locate with LREC tend to focus on NLP lexicography, a second track of GLOBALEX workshops solely focuses on neologisms and runs in parallel with the various continental lexicography conferences. The first *Globallex Workshop on Lexicography and Neology* (GWLN) took place at DSNAL 2019 and its papers were published in *Dictionaries* 41(1) in 2020, the second took place at EURALEX 2020>2021 and its papers were published in IJL 34(3) in 2021, the third took place at AUSTRALEX 2021 and its papers were published in a book volume of *Lexicographica. Series Maior*, the fourth took place at EURALEX 2022 and its papers were published as part of that conference's proceedings. (The fifth is to be held in conjunction with ASIALEX 2023 and its papers will be submitted for publication in *Lexicography* 10(2) in 2023.)

Table 5: Journals of the continental lexicography associations

Association	Founded	Journal	Publisher	Since
DSNAL	1975	<i>Dictionaries: Journal of the Dictionary Society of North America</i>	self	1979
DSNAL + EURALEX	1975 + 1983	<i>Lexicographica: International Annual for Lexicography</i>	Max Niemeyer	1985
EURALEX	1983	<i>International Journal of Lexicography</i> (IJL)	OUP	1988
AUSTRALEX	1990	— (in local/regional linguistics journals)	—	—
AFRILEX	1995	<i>Lexikos</i>	WAT	1991
ASIALEX	1997	<i>Lexicography: Journal of ASIALEX</i>	Springer; Equinox (as of 2021)	2014
AMERICALEX-S	(2023)	t.b.a.	t.b.a.	t.b.a.
eLex	2009	n.a.	n.a.	n.a.
GLOBALEX	2018	— (in LREC proceedings; in other lexicography journals; in lexicography book series)	—	—

A detailed analysis of some of these journals — as was undertaken for *Lexikos* (De Schryver 2009b) and *IJL* (De Schryver 2012b) — indicates that ever more material appears in them, even where the publisher sets page limits. The latter is for instance the case for *IJL*, where, starting with the June 2018 issue, they simply introduced a different typeface, a much smaller font size and a new layout, so they could cram in more text on the same number of pages. But is 'more' — ever more publication outlets, ever more pages, and/or ever smaller font sizes to include even more articles — also 'better'? What is the true impact of all this increased publishing? Are we now ending up with hundreds, perhaps thousands, of research articles that are simply never being cited? To try answering this question in a concise way, Google Scholar profiles were manually created for each of the four main journals that are linked to the continental associations: *Dictionaries*, *IJL*, *Lexikos* and *Lexicography*.⁴ Given that the main topic of this contribution is not bibliometrics in lexicography (see De Schryver 2009a), the present study is limited to the basic outcomes only. All screenshots and counts which follow here and in the subsequent sections refer to the situation on 13 February 2023.

Figures 3 to 6 show the number of citations that the various articles in the four journals attract per year. With ever more material, the number of times that material is cited year after year indeed goes up for *Dictionaries*, *Lexikos* and *Lexicography*, although this seems to have plateaued for *IJL* over the past decade, with an average of about 1,300 citations per year. So the 'more' for *IJL* seems not to have an increased impact. On the other hand, that plateau for *IJL* is way above the current number of annual hits for *Dictionaries* (about 230), *Lexikos* (about 600) and *Lexicography* (about 100).

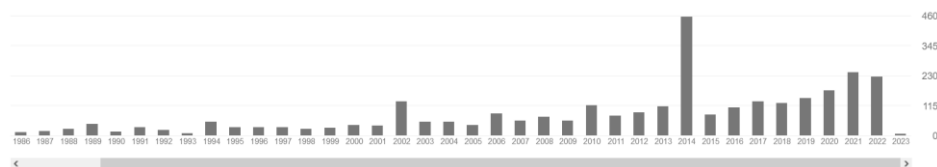


Figure 3: Number of cites per year for *Dictionaries*

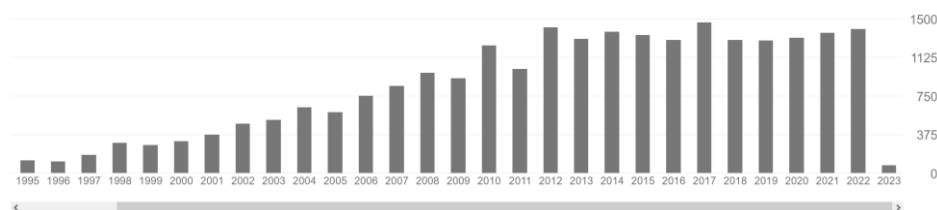


Figure 4: Number of cites per year for *IJL*

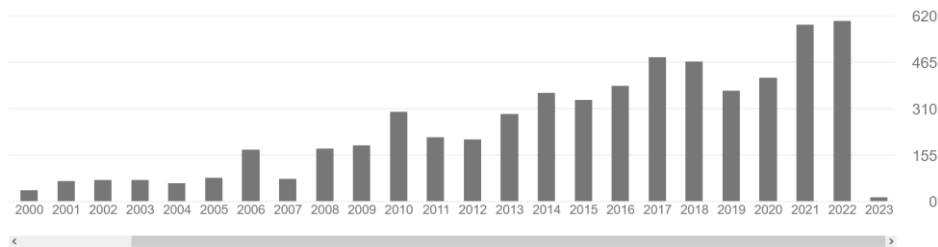


Figure 5: Number of cites per year for *Lexikos*

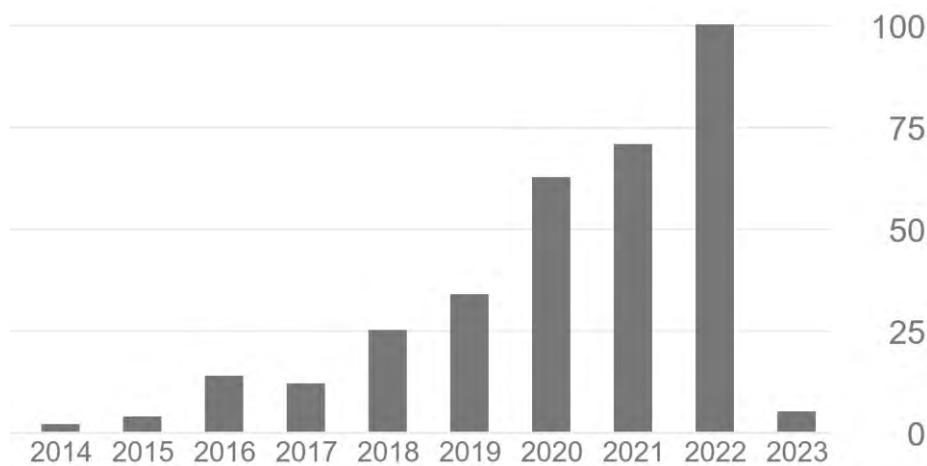


Figure 6: Number of cites per year for *Lexicography*

With regard to the number of citations depicted in Figure 6 for the journal *Lexicography*, it is important to point out that Kilgarrieff et al.'s (2014) contribution 'The Sketch Engine: Ten Years on' (published in this journal) is not included, as Google Scholar lumps its citations with Kilgarrieff et al.'s earlier (2004) 'The Sketch Engine' (published in the EURALEX 2004 proceedings). This single article actually carries the entire journal, as it must have attracted about one thousand six hundred citations since 2014, compared to 345 for all other articles together over the same time period.

A single article can also produce a huge spike for just one year in the data, dwarfing everything else, as is the case for the number of hits in 2014 for *Dictionaries*. The reason for the spike is a jubilee article by David Jost (2014) which quotes every single item that had been published during the 34 preceding years in the journal. In order to smooth out spikes, working with windows that cover a number of years is a solution. Two metrics are presented to that effect in

Table 6: 'overall' (= 'all') and 'across the last five years' (= 'since 2018'). Also shown are the *h*-index and the *i10*-index: the "*h*-index is the largest number *h* such that *h* publications have at least *h* citations", while the "*i10*-index is the number of publications with at least 10 citations".

Table 6: Basic citation metrics for the four main lexicographic journals [as of 13 February 2023]

	<i>Dictionaries</i>		IJL		<i>Lexikos</i>		<i>Lexicography</i> *	
	All	Since 2018	All	Since 2018	All	Since 2018	All	Since 2018
Citations	3,116	911	25,376	6,768	6,222	2,454	345	308
<i>h</i> -index	22	11	67	36	33	20	11	10
<i>i10</i> -index	83	13	246	154	193	75	12	10

* Without the counts for 'The Sketch Engine: Ten Years on' (Kilgariff et al. 2014), which number about 1,600 overall.

From the values in Table 6, one may conclude that despite the fact that *Dictionaries* is the oldest journal, IJL clearly comes out top, followed by *Lexikos*, then *Dictionaries*, and a distant last (although it is of course still a young journal) *Lexicography*.

As to the question of how many articles end up never being cited, the four journals surprisingly behave very similarly. In Table 7, the number of published items per journal up to the end of 2012 were tabulated, together with calculations of the number of items not cited so far, as well as those cited only once. After 'correcting' the values for *Dictionaries* (to subtract all the hits from the D. Jost overview), it turns out that the percentage of articles never cited in the four main journals of our field is exactly the same for IJL, *Lexikos* and *Lexicography*, namely 41%, and 45% for the journal *Dictionaries*. Given that hapaxes often mean very little in corpus studies, and to truly compare apples with apples and not oranges, the last column in Table 7 lists the number of items cited more than once, which reveals that only half of everything we write and publish in IJL, *Lexikos* and *Lexicography* is cited more than once, which drops to less than a third for *Dictionaries*.⁵

Table 7: Items 'not cited' and 'cited more than once' for the four main lexicographic journals [as of 13 February 2023]

Journal	Total items	Not cited	Not cited, in %	Cited once	Cited once, in %	Cited more than once, in %
<i>Dictionaries</i>	725	249	34.3%	104	14.3%	51.3%
<i>Dictionaries</i> *	725	329	45.4%	170	23.4%	31.2%
IJL	658	269	40.9%	47	7.1%	52.0%
<i>Lexikos</i>	954	391	41.0%	70	7.3%	51.7%
<i>Lexicography</i>	85	35	41.2%	10	11.8%	47.1%

* In order to 'correct' for the fact that the Jost (2014) overview cites every single item that had appeared in *Dictionaries* during the preceding 34 years, all items cited only once from before 2014 actually need to be added to the 'not cited'. There are 80 such items, which means that the corrected number of items not cited is 329, which corresponds to 45.4%. Likewise for the material cited twice: those from before 2014 need to be added to 'cited once', of which there are 66 cases, so the corrected number of cited once is 170, which corresponds to 23.4%, and the 'cited more than once' then becomes 31.2%.

On average, for every five journal articles lexicographers write, two are never cited. Given that this pattern is seen throughout, no matter whether a journal is over four decades old or just a decade old, this does imply that producing more does get you noticed. The message, however, is also that we should all focus on researching and writing the three articles out of every five that will get (you) noticed, and stop spending time on the other two — if only we knew which two. Moreover, if we wish our output to be more than just once-offs, single citations, hapaxes, we should simply cut out half up to two-thirds of everything we do metalexicographically, as that part is clearly not more than a fluke in the data, a statistical error.

Probably more interesting than knowing that over two-fifths of the material published in the four main lexicographic journals is never cited, is to know *what* is cited, and especially what attracts *most* cites. This question is answered in Figures 7 through 10, which show the Top 15 journal articles (together with their citation numbers to date) for each of these four journals. The four journals clearly have their own identity. Firstly, popular topics in *Dictionaries* include various theoretical discussions (with a link to linguistic theories as well as dictionary practice), dictionary use, translation equivalence, combinations, examples, and neologisms; popular in *IJL* are the various 'nets' (WordNet, FrameNet, ...), digital dictionaries, corpora and computational aspects, as well as dictionary use, and combinations; popular in *Lexikos* are various theoretical discussions (with a focus on specialised and pedagogical lexicography), digital dictionaries, corpora and computational aspects, and dictionary use; and popular in *Lexicography* is dictionary software, frame-based terminology, dictionary culture and skills, dictionary use, dictionary apps, and phraseology.

Secondly, in terms of languages dealt with in the Top 15, *Dictionaries* is heavily biased towards the various historical stages of American (and British) English; *IJL* is solely interested in present-day (British) English; *Lexikos* is so theoretical in the top material that it nearly appears to be 'against language', with only De Schryver et al. (2006) and Gouws and Prinsloo (1998) doing a serious effort to involve African languages, while Lew et al. (2013) produced an eye-tracking study involving Polish; and *Lexicography* is interested in varieties of Asian English on the one hand, and English as a Foreign Language (EFL) on the other.

Thirdly, in terms of authors, each journal furthermore tends to be a microcosm. *Lexikos* is especially affected, with the same Danish scholars restating their positions. The other journals do indeed list the important figures of our field, including S. Atkins, M. Benson, H. Nesi, L. Zgusta, D. and R. Barnhart, A. Wierzbicka, G. Stein and J. Apresjan in *Dictionaries*; G. Miller (and his team), C. Fillmore, G.-M. de Schryver, W. Teubert, A. Kilgariff, S. Atkins, K. Varantola, N. Calzolari, M. Benson and H. Nesi in *IJL*; and A. Kilgariff (and his colleagues from the Czech Republic), M.-C. L'Homme, P. Faber, M. Rundell and R. Gouws in *Lexicography*. Judging from the Top 15s, *Lexicography* is probably also the most 'local' in its continentalness, with contributors like D. Salazar (Philip-

piners), C. Winestock (South Korea), Y. Jeong (South Korea), J. Ding (China), S. Bae (South Korea) and T. Koyama (Japan). Actually, only a few names make an impact across the journals: S. Atkins, M. Benson, G.-M. de Schryver, R. Gouws, A. Kilgarriff and H. Nesi. More intriguing is that Sue Atkins's recounting of the first half century of metalexicography, as depicted in Figure 1 (starting with her brother J. Sinclair close to her, then on a first plane colleagues from A. Duval to N. Calzolari, and finally on a second plane colleagues from B. Levin to P. Rychlý), included no less than eleven of the most cited authors in the four journals under consideration: J. Apresjan, S. Atkins, N. Calzolari, G.-M. de Schryver, C. Fillmore, T. Fontenelle, A. Kilgarriff, D.J. Prinsloo, M. Rundell, P. Rychlý and K. Varantola — which thus includes, chronologically speaking, the last five from her summary. Even as a nonagenarian, then, Sue Atkins was still fully in the present and with that, predicting the future.

4. Research into dictionary use as metalexicography

As stressed by Gouws (2020: 3), while at some point in the past 'metalexicography' was considered separate from 'dictionary research' — itself subdivided into four categories (cf. Wiegand 1998: 114), one of which being 'research into dictionary use' — this distinction is no longer upheld. As result, in addition to the metalexicographic endeavours discussed so far, the actual study of how dictionaries are used is now part and parcel of the field of metalexicography.

Ironically, one aspect where Sue Atkins did lose touch with present praxis, and thus failed to foresee the future, was the direction in which research into dictionary use, to which she herself contributed a major study (Atkins 1998), was heading. Just as dictionaries have moved from paper to digital, research on dictionary use has now fully moved from (paper-based) questionnaires and surveys, as undertaken by Atkins in the 1990s, to the unobtrusive logging and analysis of real (digital, online) dictionary use.

Overviews of the current and future potential in this domain have already been written, and the interested reader is referred to those writings, the main ones being Dziemianko (2018) and Müller-Spitzer et al. (2018). Suffice it to say here that a particularly promising venue for the study of dictionary use with the aim to improve future dictionaries is the utilisation of eye-tracking during dictionary consultation (Lew et al. 2018). On a more basic level, recent studies have also shown the potential of carefully analysing up to a decades-worth of online dictionary logs, revealing not only real, unobtrusive dictionary usage behaviour, but also allowing for new insights into graphical-user-interface design in the context of lexicography (De Schryver et al. 2019b).



Dictionaries: Journal of The Dictionary Society of North America



publisher: Dictionary Society of North America
No verified email - [Homepage](#)
[Lexicography](#)

TITLE	CITED BY	YEAR
Processing strategies and problems encountered in the use of dictionaries A Neubach, AD Cohen Dictionaries: Journal of the Dictionary Society of North America 10 (1), 1-20	134	1988
Theoretical Lexicography and its Relation to Dictionary-making BTS Atkins Dictionaries: Journal of The Dictionary Society of North America 14 (1), 4-43	90	1992
The Dictionary of Old English A Cameron, A diPaolo Healey Dictionaries: Journal of the Dictionary Society of North America 1 (1), 87-96	76	1979
A combinatory dictionary of English M Benson Dictionaries: Journal of the Dictionary Society of North America 7 (1), 189-200	54	1985
The Middle English Compendium: past, present, future F McSparran Dictionaries: Journal of the Dictionary Society of North America 23 (1), 126-141	48	2002
The role of illustrative examples in productive dictionary use H Nesi Dictionaries: Journal of the Dictionary Society of North America 17 (1), 198-206	48	1996
Guidelines for reviewers of bilingual dictionaries RJ Steiner Dictionaries: Journal of the Dictionary Society of North America 6 (1), 166-181	45	1984
Translational equivalence in a bilingual dictionary: Bāhukośyam L Zgusta Dictionaries: Journal of the Dictionary Society of North America 9 (1), 1-47	44	1987
The image of the dictionary for American college students S Greenbaum, CF Meyer, J Taylor Dictionaries: Journal of the Dictionary Society of North America 6 (1), 31-52	42	1984
Prizes and pitfalls of computerized searching for new words for dictionaries DK Barnhart Dictionaries: Journal of the Dictionary Society of North America 7 (1), 253-260	40	1985
What are the uses of theoretical lexicography? A Wierzbicka Dictionaries: Journal of the Dictionary Society of North America 14 (1), 44-78	34	1992
The best of British and American lexicography G Stein Dictionaries: Journal of the Dictionary Society of North America 1 (1), 1-23	32	1979
Some thoughts about neologisms before starting BDNE IV RK Barnhart Dictionaries: Journal of the Dictionary Society of North America 16 (1), 51-64	28	1995
Systemic lexicography as a basis of dictionary-making JD Apresjan Dictionaries: Journal of the Dictionary Society of North America 14 (1), 79-87	28	1992
Retrieving ergative verbs from a lexical data base T Fontenelle, J Vanandroye Dictionaries: Journal of the Dictionary Society of North America 11 (1), 11-39	28	1989

Figure 7: Top 15 most cited articles from *Dictionaries*

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1803> (Article)



International Journal of Lexicography

publisher: Oxford University Press

No verified email - [Homepage](#)

[Lexicography](#)



TITLE	CITED BY	YEAR
Introduction to WordNet: An on-line lexical database GA Miller, R Beckwith, C Fellbaum, D Gross, KJ Miller International journal of lexicography 3 (4), 235-244	7352	1990
Background to framenet CJ Fillmore, CR Johnson, MRL Petruck International journal of lexicography 16 (3), 235-250	1173	2003
Word families L Bauer, P Nation International journal of Lexicography 6 (4), 253-279	1022	1993
Nouns in WordNet: a lexical inheritance system GA Miller International journal of Lexicography 3 (4), 245-264	495	1990
English verbs as a semantic net C Fellbaum International journal of Lexicography 3 (4), 278-301	486	1990
Lexicographers' dreams in the electronic-dictionary age GM de Schryver International Journal of Lexicography 16 (2), 143-199	432	2003
The structure of the FrameNet database CF Baker, CJ Fillmore, B Cronin International Journal of Lexicography 16 (3), 281-296	301	2003
Comparable or parallel corpora? W Teubert International journal of lexicography 9 (3), 238-264	301	1996
Putting frequencies in the dictionary A Kilgariff International Journal of Lexicography 10 (2), 135-155	280	1997
Monitoring dictionary use BTS Atkins, K Varantola International Journal of Lexicography 10 (1), 1-45	276	1997
SIMPLE: A general framework for the development of multilingual lexicons A Lenci, N Bel, F Busa, N Calzolari, E Gola, M Monachini, A Ogonowski, ... International Journal of Lexicography 13 (4), 249-263	259	2000
The structure of the collocational dictionary M Benson International Journal of Lexicography 2 (1), 1-14	251	1989
Collocations and general-purpose dictionaries M Benson International Journal of Lexicography 3 (1), 23-34	237	1990
A study of dictionary use by international students at a British university H Nesi, R Hail International journal of Lexicography 15 (4), 277-305	236	2002
Co-occurrence and antonymy C Fellbaum International journal of lexicography 8 (4), 281-303	231	1995

Figure 8: Top 15 most cited articles from *IJL*

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1803> (Article)

150 Gilles-Maurice de Schryver



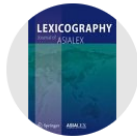
Lexikos

publisher: Bureau of the WAT
No verified email - [Homepage](#)
[Lexicography](#)



TITLE	CITED BY	YEAR
Theoretical challenges to practical specialised lexicography S Tarp Lexikos 10	127	2000
Reflections on lexicographical user research S Tarp Lexikos 19	125	2009
The effect of lexicographical information costs on dictionary making S Nielsen Lexikos 18	95	2008
Multimodal lexicography: The representation of meaning in electronic dictionaries R Lew Lexikos 20, 290-306	79	2010
Do dictionary users really look up frequent words? — On the overestimation of the value of corpus-based lexicography GM de Schryver, D Joffe, P Joffe, S Hillewaert Lexikos 16, 67-83	79	2006
Bilingual dictionaries and communicative equivalence for a multilingual society RH Gouws Lexikos 6, 14-31	79	1996
What is lexicography? H Bergenholtz, RH Gouws Lexikos 22, 31-42	72	2012
Lexicography in the information age S Tarp Lexikos 17	71	2007
User-oriented understanding of descriptive, proscriptive and prescriptive lexicography H Bergenholtz Lexikos 13	65	2003
Basic problems of learner's lexicography S Tarp Lexikos 14	61	2004
Outer texts in bilingual dictionaries R Gouws Lexikos 14	60	2004
Cross-referencing as a Lexicographic Device RH Gouws, DJ Prinsloo Lexikos 8, 17-36	60	1998
How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study R Lew, M Grzelak, M Leszkowicz Lexikos 23, 228-254	59	2013
Needs-adapted data presentation in e-information tools H Bergenholtz, TJD Bothma Lexikos 21	54	2011
Pedagogical lexicography: Towards a new and strict typology corresponding to the present state-of-the-art S Tarp Lexikos 21	52	2011

Figure 9: Top 15 most cited articles from *Lexikos*



Lexicography: Journal of ASIALEX

publishers: Springer (2014-2020) & Equinox Publishing (2021-)

Verified email at UGent.be - [Homepage](#)

[Lexicography](#)



TITLE	CITED BY	YEAR
The Sketch Engine: ten years on A Kilgariff, V Baisa, J Bušta, M Jakubiček, V Kovář, J Michelfeit, P Rychlý, ... Lexicography 1 (1), 7-36	3021	2014
Maintaining the balance between knowledge and the lexicon in terminology: a methodology based on frame semantics MC L'Homme Lexicography 4 (1), 3-21	31	2018
Lacunarity, lexicography and beyond: integration of the introduction of a linguo-cultural concept and the development of L2 learners' dictionary skills J Szerszunowicz Lexicography 2 (1), 101-118	28	2015
Towards improved coverage of Southeast Asian Englishes in the <i>Oxford English Dictionary</i> D Salazar Lexicography 1 (1), 95-108	21	2014
Phraseology in specialized resources: An approach to complex nominals M Cabezas-García, P Faber Lexicography 5 (1), 55-83	19	2018
Assessing dictionary skills MC Campoy-Cubillo Lexicography 2 (1), 119-141	19	2015
An analysis of the smartphone dictionary app market C Winestock, Y Jeong Lexicography 1 (1), 109-119	17	2014
Improving the consistency of usage labelling in dictionaries with <i>tei lex-0</i> A Salgado, R Costa, T Tasovac Lexicography 6 (2), 133-156	14	2019
Searching for extended units of meaning—and what to do when you find them M Rundell Lexicography 5 (1), 5-21	14	2018
A study of English majors in a Chinese university as dictionary users J Ding Lexicography 2 (1), 5-34	14	2015
Developing a dictionary culture through integrated dictionary pedagogy in the outer texts of South African school dictionaries: the case of Oxford Bilingual School Dictionary ... D Nkomo Lexicography 2 (1), 71-99	12	2015
A course in dictionary use for Korean EFL teachers S Bae Lexicography 2 (1), 45-69	12	2015
The impact of E-dictionary strategy training on EFL class T Koyama Lexicography 2 (1), 35-44	10	2015
Expanding the notion of addressing relations RH Gouws Lexicography 1 (2), 159-184	9	2014
Frame-based terminology applied to military science: transforming a glossary into a knowledge resource P Faber, P León-Araúz Lexicography 6 (2), 105-131	8	2019

Figure 10: Top 15 most cited articles from *Lexicography*⁶

5. *Elexifinder*: A tool with the potential to browse all of the world's meta-lexicographic publications

The worldwide annual metalexicographic output includes much more than the published congress papers and journal articles mentioned in Sections 3 and 4, as the scientific output of our discipline is also found in several other journals, monographs and book series, edited collections, festschrifts, as well as manuals and handbooks — all of these with either a narrow focus on lexicography, or linguistics in general; written in English, as well as many other languages. How will future students of lexicography, dictionary makers and metalexicographers navigate this wealth of data? A good candidate is *Elexifinder*, a prototype discovery portal for lexicographic literature produced by the Elexis project,⁷ designed and described by Kosem and Lindemann (2021). With currently just over 6,500 articles and books in the second iteration of *Elexifinder*, this is a good start.⁸

In order to illustrate what *Elexifinder* can currently do, let us return to (some of) the metalexicographic output of Sue Atkins. *Elexifinder* 2 contains meta-data and pointers to 18 of the articles she (co-)wrote — clearly an incomplete set. By way of example, Figure 11 displays the co-authors for those 18 articles. Figure 12 shows the automatically generated 'tag cloud' from those 18 articles, where the most salient keywords are (correctly) 'frame' and 'FrameNet', 'transitive' and 'intransitive' uses of the 'verbs' 'cook' ('cuire' in 'French') and 'bake' (used in Atkins et al. 2003), the 'database' 'DANTE' (Atkins et al. 2010, Rundell and Atkins 2011) which was used for an 'Irish' 'bilingual' (Convery et al. 2010), 'lexicographic' 'training' and her colleague 'Kilgariff'. This is clearly an impressive 'fit', all the more so that each of those tags are clickable, and leads to the relevant article abstracts and when in open access the full texts.

Figure 13 shows the concept graph for the same 18 articles, where concepts that frequently co-occur are connected, but this view is less impressive; for one would have expected to see the concept 'virtual dictionary' there (Atkins 1996). Searching for 'virtual dictionary' in the entire database, however, does reveal a number of relevant studies that have referred to the original concept by Sue Atkins, as seen in Figure 14, but surprisingly not to the initial publication itself (Atkins 1996), which is in the database.

While all of the previous wizardry of *Elexifinder* could still be achieved by mostly automated ways, Figure 15 illustrates the result of the painstaking effort undertaken by David Lindemann to assign categories and subcategories to each and every publication in the database.⁹ Here, 5.84% of the articles by Sue Atkins (covered in the database) deal with 'dictionary use', 3.55% with the sub-category 'user profile', and finally 2.03% of all terms assigned to articles by Sue Atkins deal with the sub-sub-category 'native speaker'. Needless to say, all these labels and pie sections are again clickable. As an illustration, clicking on the 'native speaker' slice in Figure 15 reveals what is displayed in Figure 16. Therefore, despite the fact that metalexicographers will have to master more

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1803> (Article)

material than ever in future, tools such as *Ellexifinder* will indeed help them to find and discover earlier research in a user-friendly way.

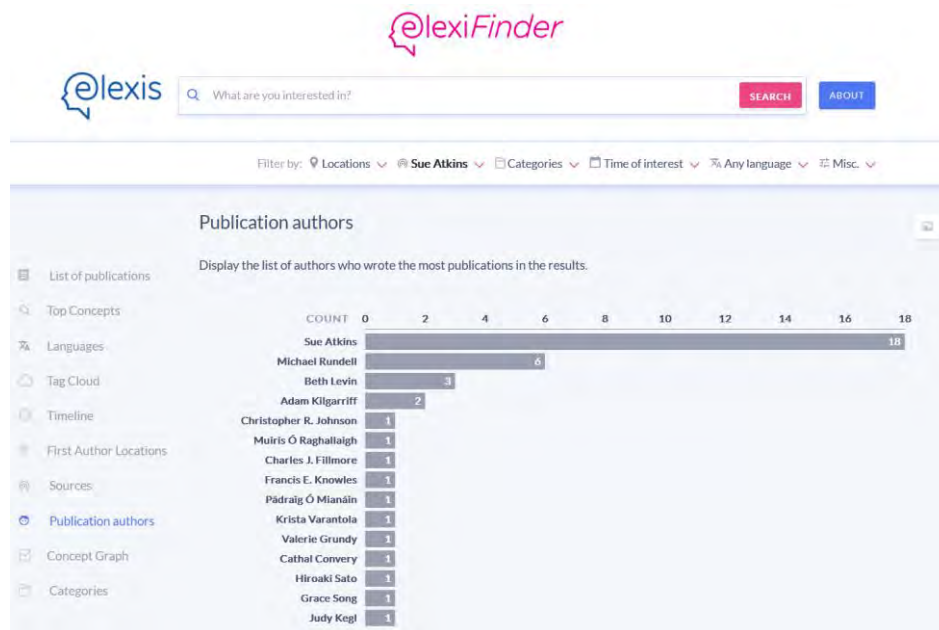


Figure 11: List of authors for articles (co-)written by Sue Atkins in *Elexifinder 2*



Figure 12: Tag cloud of top keywords extracted from articles by Sue Atkins in *Elexifinder 2*

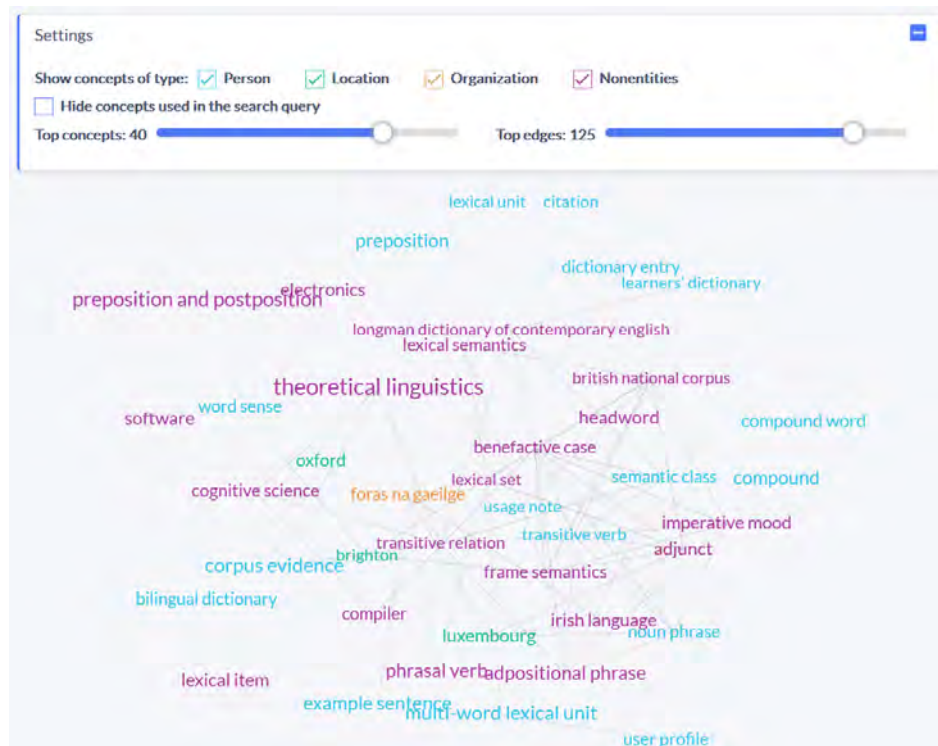


Figure 13: Concept graph extracted from articles by Sue Atkins in *Elexifinder 2*

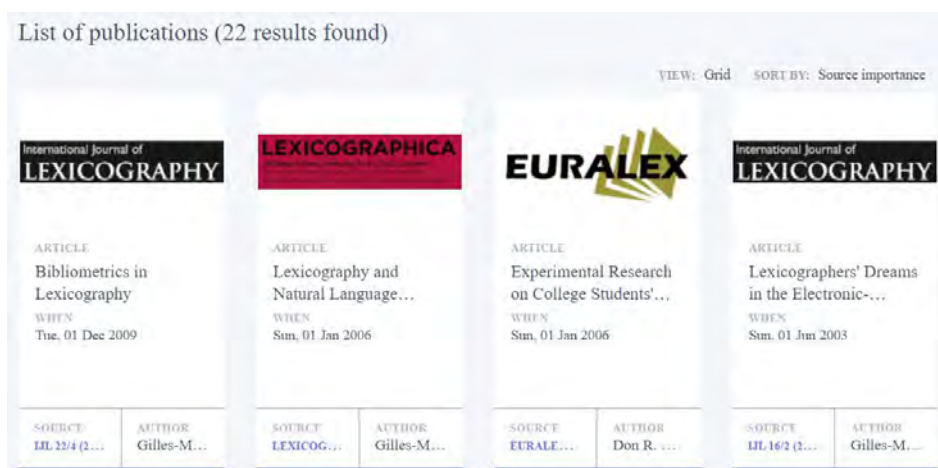


Figure 14 [first row]: List of publications with the concept 'virtual dictionary' in *Elexifinder 2*

A Practical Guide to Lexicography <small>EDITED BY Piet van Sterkenburg</small> <small>ARTICLE</small> Linguistic corpora (databases) and the... <small>WHEN</small> Wed, 01 Jan 2003 <small>SOURCE</small> STERKEN... <small>AUTHOR</small> Krista V...	<small>Herbert Ernst Wiegand (Ed.)</small> WÖRTERBÜCHER IN DER DISKUSSION IV <small>ARTICLE</small> Zur Hypertextualisierung... <small>WHEN</small> Sat, 01 Jan 2000 <small>SOURCE</small> WIEGAN... <small>AUTHOR</small> Krzyszto...	EURALEX <small>ARTICLE</small> Computer Words in Our Everyday Lives: How... <small>WHEN</small> Sat, 01 Jan 2000 <small>SOURCE</small> EURALE... <small>AUTHOR</small> Ingrid M...	LEXICOGRAPHICA <small>ARTICLE</small> Problems of Exemplification in... <small>WHEN</small> Fri, 31 Dec 1999 <small>SOURCE</small> LEXICOG... <small>AUTHOR</small> Thomas ...
DICTIONARIES <small>ARTICLE</small> An Annotated Survey of English Etymological... <small>WHEN</small> Thu, 01 Jan 1998 <small>SOURCE</small> DSNA JO... <small>AUTHOR</small> Anatoly ...	EURALEX <small>ARTICLE</small> From Virtual Sex to Virtual Dictionaries: ... <small>WHEN</small> Thu, 01 Jan 1998 <small>SOURCE</small> EURALE... <small>AUTHOR</small> Krista V...	International Journal of LEXICOGRAPHY <small>ARTICLE</small> BudaLEX Presidential Debate 1988: Part 2 T... <small>WHEN</small> Tue, 01 Dec 1992 <small>SOURCE</small> IJL 5/4 (1992) <small>AUTHOR</small> Adam M...	eLex <small>ARTICLE</small> A Corpus-assisted Approach to Paronym... <small>WHEN</small> Sun, 01 Jan 2017 <small>SOURCE</small> ELEX (2017) <small>AUTHOR</small> Petra Sto...

[truncated]

eLex <small>ARTICLE</small> Collocational Networks and Their Application... <small>WHEN</small> Sat, 01 Jan 2011 <small>SOURCE</small> ELEX (2011) <small>AUTHOR</small> Geoffrey...	LEXIKOS <small>ARTICLE</small> National Lexicography Units: Past, Present,... <small>WHEN</small> Sat, 01 Jan 2011 <small>SOURCE</small> LEXIKOS... <small>AUTHOR</small> Mariëtta ...
--	--

Figure 14 [continued]: List of publications with the concept 'virtual dictionary' in *Elexifinder 2*

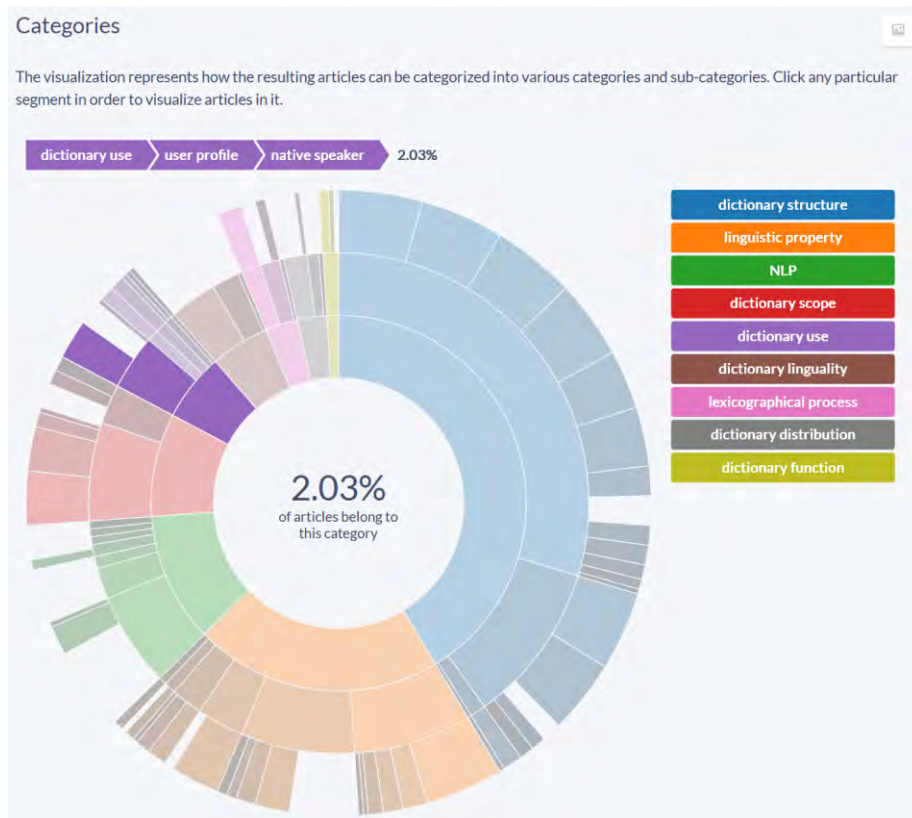


Figure 15: Categorization for articles by Sue Atkins (zooming in on user profiling for native speakers during dictionary use) in *Elexifinder 2*

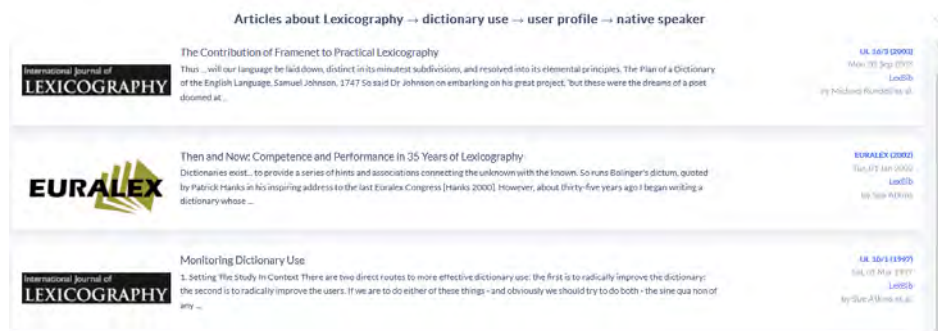


Figure 16: Articles by Sue Atkins on 'Lexicography > dictionary use > user profile > native speaker' in *Elexifinder 2* [truncated here]

6. Discussion and conclusion

Colleagues have thought about, and then spoken and written about, the future of the dictionary and/or the future of (meta)lexicography before, but in very uneven ways. This may be illustrated with a search in our database of lexicographic publications.¹⁰ A representative sample of what was found, organised into six categories (I to VI), is listed on pages 158-159. The list contains 34 contributions, and to assign each to one of the six categories, they were first judged on whether their predictions about the future were on the whole either practice-inspired, theory-oriented or data-driven, and second whether they were mostly about lexicography or metalexicography. Each contribution was thus assigned to just one of the cells of the three-by-two matrix shown in the centre of Table 8, even though some could arguably have been put in several cells. The raw counts are also expressed as percentages.

Table 8: Distribution of articles and talks on the future of our discipline

	Practice-inspired	Theory-oriented	Data-driven	
Lexicography	Ia. (specific) 11 (32%) Ib. (general) 9 (26%)	III. 3 (9%)	V. 3 (9%)	76%
Metalexicography	II. 3 (9%)	IV. 4 (12%)	VI. 1 (3%)	24%
	68%	21%	12%	

As one may deduce from the data, when lexicographers talk about the future of our discipline they mostly do so based on their own experience in practical lexicography, and their suggestions concern new types of dictionaries or dictionary features, either focused on certain languages and regions, or limited to certain fields and topics (cat. Ia: 11 contributions, 32%). The next category is when they use their own experience to make more general claims, supposedly valid for all dictionaries (cat. Ib: 9 contributions, 26%). A small number of colleagues also departs from their own practice to make wider, metalexicographic claims (cat. II: 3 contributions, 9%). In all, two-thirds (68%) of all work on the future is thus rooted in the lexicographic practice.

The next group, good for two-fifths of the total (21%), is when colleagues depart from theoretical concepts to make predictions about the future. These concern dictionaries proper (cat. III: 3 contributions, 9%), or they concern meta-levels such as the interaction between language policy and dictionary-making (cat. IV: 4 contributions, 12%).

The last group is the smallest, with just a tenth (12%) of the contributions, and concerns those studies where large quantities of data are processed, with which various trend lines are drawn, and from which extrapolations are made to predict the future. These extrapolations either concern lexicography proper (cat. V: 3 contributions, 9%) or metalexicography (cat. VI: 1 contribution, 3%).

I. Practice-inspired predictions about the future of lexicography

Ia. Specific (language or region; field or topic)

Language or region

- Gove (1972) *English Dictionaries of the Future*
- Schäfer (1984) *The History of Ideas and Cross-referencing in the Future* *Early Modern English Dictionary*
- Hoftijzer (1995) *The Present and Future of the Dictionary of the Northwest-Semitic Epigraphy*
- McSparran (2002) *The Middle English Compendium: Past, Present, Future*
- Nkomo and Wababa (2013) *IsiXhosa Lexicography: Past, Present and Future*
- Prinsloo (2019) *A Perspective on the Past, Present and Future of Lexicography with Specific Reference to Africa*

Field or topic

- Sobkowiak (1994) *Beyond the Year 2000: Phonetic-access Dictionaries*
- Nesi (1996) *Review Article: For Future Reference? Current English Learners' Dictionaries in Electronic Form*
- Leech and Nesi (1999) *Moving Towards Perfection: The Learners' (Electronic) Dictionary of the Future*
- Kwary (2010) *Access Routes of Internet Finance Dictionaries: Present Solutions and Future Opportunities*
- Liberman (2015) *The Future of Etymological Dictionaries*

Ib. General

- Kay (1983) *The Dictionary of the Future and the Future of the Dictionary*
- Abate (1985) *Dictionaries Past & Future: Issues and Prospects*
- Bailey (1986) *Dictionaries of the Next Century*
- Crystal (1986) *The Ideal Dictionary, Lexicographer and User*
- Dodd (1989) *Lexicomputing and the Dictionary of the Future*
- Zaenen (2002) *Musings about the Impossible Electronic Dictionary*
- Rundell (2007) *The Dictionary of the Future*
- De Schryver et al. (2019a) *An Overview of Digital Lexicography and Directions for Its Future*
- Ogilvie (2021) *The Future of Dictionaries and Lexicography*

II. Practice-inspired predictions about the future of metalexicography

- Atkins (1996) *Bilingual Dictionaries: Past, Present and Future*
- De Schryver (2003) *Lexicographers' Dreams in the Electronic-Dictionary Age*
- Hanks (2018) *Phraseology, Meaning, and the Future of Lexicography*

III. Theory-oriented predictions about the future of lexicography

- Zgusta (1991) *Probable Future Developments in Lexicography*
- Steurs et al. (2020) *The Future of Academic Lexicography*
- Gouws and Prinsloo (2021) *Lexicographic Data Boxes, Part 3: Aspects of Data Boxes in Bilingual Dictionaries and a Perspective on Current and Future Data Boxes*

IV. Theory-oriented predictions about the future of metalexicography

- Tarp and Gouws (2008) A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries
- Alberts (2011) National Lexicography Units: Past, Present, Future
- Tarp (2012) Online Dictionaries: Today and Tomorrow
- Nielsen (2013) The Future of Dictionaries, Dictionaries of the Future

V. Data-driven predictions about the future of lexicography

- Grefenstette (1998) The Future of Linguistics and Lexicographers
- Lew and De Schryver (2014) Dictionary Users in the Digital Revolution
- De Schryver (2023) The Future of the Dictionary

VI. Data-driven predictions about the future of metalexicography

- De Schryver (2019b) Past, Present and Future in Asian Lexicography

Viewed in another direction, three-quarters (76%) of all contributions on the future concern lexicography, while just one-quarter (24%) concern metalexicography. Within the latter group, the data-driven approach is not only the smallest (with a single contribution),¹¹ but also the newest, as it was introduced at ASIALEX 2019, where making 'data-driven predictions about the future of metalexicography' was termed 'meta-metalexicography':

In meta-metalexicography (i) a bird's eye view is taken of (ii) the scientific research devoted to (iii) the compilation of dictionaries. Level (iii) is known as lexicography, levels (ii-iii) as metalexicography, and levels (i-ii-iii) is therefore 'meta-metalexicography' (De Schryver 2019a).

One may thus conclude that all colleagues, when making predictions about the future of our field, somehow base their claims on facts, mainly observed and internalised (practice), or derive them during thought experiments (theory). The base for the resulting claims is not always made explicit, however. When the predictions are data-driven, that base is the starting point, so a given. This, then, also differentiates the present contribution from all but one of the earlier studies, as all provided predictions about the future of metalexicography are based on *extrapolations* of quantifiable data (see Tables 2 through 8, and Figures 2 through 16). Making predictions based on assumed continuations of trends seen in the material of the past half century has the advantage that one is able to say *why* (as *how*) the predictions were made. It has, admittedly, the disadvantage of being 'less sexy' as one is not freewheeling, thinking out of the box, dreaming so to say, as was done in a number of earlier studies on the future of lexicography (Atkins 1996, Zaenen 2002, De Schryver et al. 2019a).

The metalexicographic output has indeed been impressive: from a handful of conferences before 1971 (most noteworthy the *Indiana University Conference on Lexicography* in 1960) to well over a hundred continental ones during the past

half century; from conference proceedings in the single digits (the proceedings of the 1960 conference first appeared as Householder and Saporta (1962)) to over 3,000 published conference papers since then (De Schryver 2019b); and from a few dozen (meta)lexicographic articles to over 10,000 today. Starting in 2011, eLex conferences began to be streamed online and the presentations stored in the cloud, a practice now copied (initially out of pandemic-necessity in 2021) by all other continental associations a decade later. All this material is generating ever more citations (so one has to conclude that it *is* also consulted, and not (all) written in vain). Overall, the papers published in the ASIALEX proceedings, for instance, generated about 1,000 citations so far,¹² while the four main lexicographic journals aggregated well over 36,500 citations to date (see Table 6).

Metalexicography is therefore not really going through an existential crisis, as provocatively opined by De Schryver (2022), although the slow pace of some publication channels as well as the realisation that a considerable amount of what we write does not even attract a single citation is certainly food for thought. Lastly, GLOBALEX should take up an even bigger role than is currently the case, in order to achieve a better mix of novel ideas and a healthy exchange of the practices used on other continents, for languages other than Indo-European ones. A current English bias is undeniable, and to the speakers and lexicographers working on this language, it would be advisable to read up on the challenges and solutions devised elsewhere, say in North America (e.g., Spence 2021), in South America (e.g., Dietrich 2014), in Africa (e.g., De Schryver 2010), in Europe and North Asia for its endangered languages (e.g., Salminen 2022), in Central Asia (e.g., Hill and Garrett 2017), in Southeast Asia (e.g., Bradley 2015), further east in Asia (e.g., Tono 2022), or finally in Australia and the Pacific (e.g., Thieberger 2015). At that point, and with all this cross-fertilisation — which is bound to result in more original dictionaries, smarter conference papers and journal articles, and even better data for metalexicographic studies — there is a good chance that we will be able to reach the next level: the metalexicographic mesosphere.

Endnotes

1. Available at <https://www.youtube.com/watch?v=dZck6jUP-Hw>.
2. There was actually just one person to do so, the author of the present contribution.
3. The recordings were initially only available to the EURALEX 2020>2021 congress participants at <https://portalapp.artion.eventsair.com/VirtualAttendeePortal/euralex-2021/onair/>, but EURALEX board members Iztok Kosem and G.-M. de Schryver took charge, and transferred the data to the more permanent VideoLectures.net.
4. These Google Scholar profiles are available at the following locations:
 - *Dictionaries*: <https://scholar.google.com/citations?user=8sqTJ6gAAAAJ>
 - *IJL*: <https://scholar.google.com/citations?user=fXZcyNQAAAAJ>
 - *Lexikos*: <https://scholar.google.com/citations?user=G14hyzcAAAAJ>
 - *Lexicography*: <https://scholar.google.com/citations?user=cLVrkSEAAAAJ>

5. See De Schryver (2022: 198, note 3) for a wider perspective on this issue.
6. Recall that Google Scholar over-counts the top article published in the journal *Lexicography*, as it fails to differentiate between Kilgariff et al. (2014), which it should include, and Kilgariff et al. (2004), which it should not include here. The true count here should be about 1,600.
7. Available at <https://finder.elex.is/intelligence>.
8. For comparison, the present author's private electronic corpus of lexicographic publications contains about 10,000 articles and books. An earlier version was used and described in Lew and De Schryver (2014).
9. For more on the *Elexifinder* categories, see <https://lexbib.elex.is/wiki/LexVoc>.
10. See Endnote 8.
11. It is tempting to add other metalexicographical data-driven studies here, but on closer inspection they stopped short of also predicting the future. These include the comparative analyses of the journals *IJL*, *Dictionaries* and *Lexikos* in De Schryver (2009a), the dedicated studies of two of these journals: *Lexikos* in De Schryver (2009b) and *IJL* in De Schryver (2012b), as well as an analysis of the first thirty years of EURALEX congress proceedings in De Schryver (2012a).
12. See the Google Scholar profile for the ASIALEX conference proceedings (part of the supporting data of De Schryver (2019b)): https://scholar.google.com/citations?user=tJy_3GQAAAAJ.

References

- Abate, Fank R. 1985. Dictionaries Past & Future: Issues and Prospects. *Dictionaries: Journal of the Dictionary Society of North America* 7: 270-283.
- Adams, Michael. 2014. The Dictionary Society of North America: A History of the Early Years (Part I). *Dictionaries: Journal of the Dictionary Society of North America* 35: 1-35.
- Alberts, Mariëtta. 2011. National Lexicography Units: Past, Present, Future. *Lexikos* 21: 23-52.
- Atkins, B.T. Sue. 1996. Bilingual Dictionaries: Past, Present and Future. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catalina Røjder Pappmehl (Eds.). *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.
- Atkins, B.T. Sue (Ed.). 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators* (Lexicographica. Series Maior 88). Tübingen: Max Niemeyer.
- Atkins, B.T. Sue, Adam Kilgariff and Michael Rundell. 2010. Database of ANalysed Texts of English (DANTE): The NEID Database Project. Dykstra, Anne and Tanneke Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*: 549-556. Leeuwarden/Ljouwert: Fryske Akademy.
- Atkins, B.T. Sue, Michael Rundell and Hiroaki Sato. 2003. The Contribution of FrameNet to Practical Lexicography. *International Journal of Lexicography* 16(3): 333-357.
- Bailey, Richard W. 1986. Dictionaries of the Next Century. Ilson, Robert F. (Ed.). 1986. *Lexicography: An Emerging International Profession* (Fulbright Papers 1): 123-137. Manchester: Manchester University Press.
- Bradley, David. 2015. The Lexicography of Minority Languages in Southeast Asia. Hanks, Patrick and Gilles-Maurice de Schryver (Eds.). 2015. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.

-
- Convery, Cathal, Pádraig Ó Mianáin, Muiris Ó Raghallaigh, B.T. Sue Atkins, Adam Kilgarrieff and Michael Rundell. 2010. The DANTE Database (Database of ANalysed Texts of English). Dykstra, Anne and Tanneke Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*: 293–295. Leeuwarden/Ljouwert: Fryske Akademy.
- Crystal, David. 1986. The Ideal Dictionary, Lexicographer and User. Ilson, Robert F. (Ed.). 1986. *Lexicography: An Emerging International Profession* (Fulbright Papers 1): 72–81. Manchester: Manchester University Press.
- De Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143–199.
- De Schryver, Gilles-Maurice. 2008. EURALEX Newsletter — PangaeaLex: New Lexicography Portal. *International Journal of Lexicography* 21(1): 113–114.
- De Schryver, Gilles-Maurice. 2009a. Bibliometrics in Lexicography. *International Journal of Lexicography* 22(4): 423–465.
- De Schryver, Gilles-Maurice. 2009b. Lexikos at Eighteen: An Analysis. *Lexikos* 19: 372–403.
- De Schryver, Gilles-Maurice. 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161–201.
- De Schryver, Gilles-Maurice. 2012a. Lexicography in the Crystal Ball: Facts, Trends and Outlook. Fjeld, Ruth V. and Julie M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo*: 93–163. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- De Schryver, Gilles-Maurice. 2012b. Trends in Twenty-five Years of Academic Lexicography. *International Journal of Lexicography* 25(4): 464–506 + 42 pages of supplementary material online.
- De Schryver, Gilles-Maurice. 2019a. Introducing 'meta-metalexigraphy' (Post on the Euralex Facebook Site, 25 June 2019). Available online at https://www.facebook.com/permalink.php?story_fbid=pfbid02xpR87cxp3xEqaUJMgmaNpH3Gaa7EYRYURiDHZW1RRuWeB6zGiz2QHBpVEWt1vJhZl&id=425976994160909.
- De Schryver, Gilles-Maurice. 2019b. *Past, Present and Future in Asian Lexicography*. Paper presented at the 13th International Conference of the Asian Association for Lexicography, ASIALEX 2019, Istanbul, Turkey, 19–21 June 2019.
- De Schryver, Gilles-Maurice. 2022. Metalexigraphy: An Existential Crisis. Klosa-Kückelhaus, Annette, Stefan Engelberg, Christine Möhrs and Petra Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress, 12–16 July 2022, Mannheim, Germany*: 196–206. Mannheim: IDS-Verlag.
- De Schryver, Gilles-Maurice. 2023. The Future of the Dictionary. Finegan, Edward and Michael Adams (Eds.). 2023. *The Cambridge Handbook of the Dictionary*. Cambridge: Cambridge University Press.
- De Schryver, Gilles-Maurice, David Joffe, Pitta Joffe and Sarah Hillewaert. 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67–83.
- De Schryver, Gilles-Maurice, Rove Chishman and Bruna da Silva. 2019a. An Overview of Digital Lexicography and Directions for Its Future: An Interview with Gilles-Maurice de Schryver. *Calidoscópio* 17(3): 659–683.
- De Schryver, Gilles-Maurice, Sascha Wolfer and Robert Lew. 2019b. 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 + supplementary material online.
- Dietrich, Wolf.** 2014. The Lexicography of Indigenous Languages in South America. Hanks, Patrick and Gilles-Maurice de Schryver (Eds.). 2014. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.
- Dodd, W. Steven.** 1989. Lexicocomputing and the Dictionary of the Future. James, Gregory (Ed.). 1989. *Lexicographers and Their Works* (Exeter Linguistic Studies 14): 83-93. Exeter: Exeter University Press.
- Dziemianko, Anna.** 2018. Electronic Dictionaries. Fuertes-Olivera, Pedro A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 663-683. Milton Park, Abingdon: Routledge.
- Gouws, Rufus H.** 2020. Metalexicography, Dictionaries and Culture. *Lexicographica: International Annual for Lexicography* 36: 3-9.
- Gouws, Rufus H. and D.J. Prinsloo.** 1998. Cross-referencing as a Lexicographic Device. *Lexikos* 8: 17-36.
- Gouws, Rufus H. and D.J. Prinsloo.** 2021. Lexicographic Data Boxes, Part 3: Aspects of Data Boxes in Bilingual Dictionaries and a Perspective on Current and Future Data Boxes. *Lexikos* 31: 402-433.
- Gove, Philip B.** 1972. English Dictionaries of the Future. Weinbrot, Howard D. (Ed.). 1972. *New Aspects of Lexicography: Literary Criticism, Intellectual History, and Social Change*: 152-168. Carbondale, IL: Southern Illinois University Press.
- Grefenstette, Gregory.** 1998. The Future of Linguistics and Lexicographers: Will There be Lexicographers in the Year 3000? Fontenelle, Thierry, Philippe Hilgsmann, Archibald Michiels, André Moulin and Siegfried Theissen (Eds.). 1998. *ACTES EURALEX'98 PROCEEDINGS, Communications soumises à EURALEX'98 (Huitième Congrès International de Lexicographie) à Liège, Belgique / Papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*: 25-41. Liège: English and Dutch Departments, University of Liège.
- Hanks, Patrick.** 2018. *Phraseology, Meaning, and the Future of Lexicography*. Paper presented at Euralex 2018. http://videlectures.net/euralex2018_hanks_phraseology/.
- Hill, Nathan W. and Edward Garrett.** 2017. The Lexicography of Tibetan. Hanks, Patrick and Gilles-Maurice de Schryver (Eds.). 2017. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.
- Hoftijzer, Jacob.** 1995. The Present and Future of the Dictionary of the Northwest-Semitic Epigraphy. *Studi Epigrafici e Linguistici sul Vicino Oriente Antico* 12: 85-97.
- Householder, Fred W. and Sol Saporta (Eds.).** 1962. *Problems in Lexicography* (International Journal of American Linguistics 28(2), Supplement 4 / Indiana University Research Centre in Anthropology, Folklore, and Linguistics 21). Bloomington, IN: Indiana University.
- Jost, David.** 2014. Thirty-four Years of Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 35: 36-92.
- Kay, Martin.** 1983. The Dictionary of the Future and the Future of the Dictionary. Zampolli, Antonio and Amadeo Cappelli (Eds.). 1983. *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries: Proceedings of the European Science Foundation Workshop, Pisa '81* (Linguistica Computazionale 3): 161-174. Pisa: Giardini Editori.
- Kernerman, Ilan.** 1997. Towards PEOPLEX. *Kernerman Dictionary News* 5: 7.
- Kilgariff, Adam, Pavel Rychlý, Pavel Smrz and David Tugwell.** 2004. The Sketch Engine. Williams, Geoffrey and Sandra Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 105-115. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel.** 2014. The Sketch Engine: Ten Years on. *Lexicography: Journal of ASIALEX* 1(1): 7-36.
- Kosem, Iztok and David Lindemann.** 2021. New Developments in Elexifinder, a Discovery Portal for Lexicographic Literature. Gavrilidou, Zoe, Lydia Mitits and Spyros Kiosses (Eds.). 2021. *EURALEX XIX, Congress of the European Association for Lexicography, Lexicography for Inclusion, 7-9 September 2021, Virtual, Proceedings Book Volume 2*: 759-766. Komotini: SynMorPhoSe Lab, Democritus University of Thrace.
- Kwary, Deny Arnos.** 2010. Access Routes of Internet Finance Dictionaries: Present Solutions and Future Opportunities. *Lexikos* 20: 272-289.
- Leech, Geoffrey and Hilary Nesi.** 1999. Moving Towards Perfection: The Learners' (Electronic) Dictionary of the Future. Herbst, Thomas and Kerstin Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?)* (Lexicographica. Series Maior 95): 295-306. Tübingen: Max Niemeyer.
- Lew, Robert and Gilles-Maurice de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
- Lew, Robert, Marcin Grzelak and Mateusz Leszkowicz.** 2013. How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-tracking Study. *Lexikos* 23: 228-254.
- Lew, Robert, Rafał Kaźmierczak, Ewa Tomczak and Mateusz Leszkowicz.** 2018. Competition of Definition and Pictorial Illustration for Dictionary Users' Attention: An Eye-tracking Study. *International Journal of Lexicography* 31(1): 53-77.
- Liberman, Anatoly.** 2015. The Future of Etymological Dictionaries. *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 132(4): 229-237.
- McSparran, Frances.** 2002. The Middle English Compendium: Past, Present, Future. *Dictionaries: Journal of the Dictionary Society of North America* 23: 126-141.
- Müller-Spitzer, Carolin, Alexander Koplenig and Sascha Wolfer.** 2018. Dictionary Usage Research in the Internet Era. Fuertes-Olivera, Pedro A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 715-734. Milton Park, Abingdon: Routledge.
- Nesi, Hilary.** 1996. Review Article: For Future Reference? Current English Learners' Dictionaries in Electronic Form. *System* 24(4): 537-546.
- Nielsen, Sandro.** 2013. The Future of Dictionaries, Dictionaries of the Future. Jackson, Howard (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 355-372. London: Bloomsbury.
- Nkomo, Dion and Zola Wababa.** 2013. IsiXhosa Lexicography: Past, Present and Future. *Lexikos* 23: 348-370.
- Ogilvie, Sarah.** 2021. The Future of Dictionaries and Lexicography. *Dictionaries: Journal of the Dictionary Society of North America* 42(2): 91-94.
- Prinsloo, D.J.** 2019. A Perspective on the Past, Present and Future of Lexicography with Specific Reference to Africa. Gürlek, Mehmet, Ahmet Naim Çiçekler and Yasin Taşdemir. 2019. *Proceedings of the 13th International Conference of the Asian Association for Lexicography, ASIALEX 2019, 19-21 June 2019, Istanbul, Turkey*: 148-160. Istanbul: Asos
- Rundell, Michael.** 2007. The Dictionary of the Future. Granger, Sylviane (Ed.). 2007. *Optimizing the Role of Language in Technology-enhanced Learning* (Proceedings of the expert workshop of the integrated digital language learning seed grant project): 49-51. Louvain-la-Neuve: Université catholique de Louvain.

- Rundell, Michael and B.T. Sue Atkins.** 2011. The DANTE Database: A User Guide. Kosem, Iztok and Karmen Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011, Bled, 10–12 November 2011*: 233-246. Ljubljana: Trojina, Institute for Applied Slovene Studies.
- Salminen, Tapani.** 2022. The Lexicography of Endangered Languages in Europe and North Asia. Hanks, Patrick and Gilles-Maurice de Schryver (Eds.). 2022. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.
- Schäfer, Jürgen.** 1984. The History of Ideas and Cross-referencing in the Future EMED. *Dictionaries: Journal of the Dictionary Society of North America* 6: 182-198.
- Sobkowiak, Włodzimierz.** 1994. Beyond the Year 2000: Phonetic-access Dictionaries (with Word-frequency Information) in EFL. *System* 22(4): 509-523.
- Spence, Justin.** 2021. A Corpus too Small: Uses of Text Data in a Hupa–English Bilingual Dictionary. *International Journal of Lexicography* 34(4): 413-436.
- Steurs, Frieda, Tanneke Schoonheim, Kris Heylen and Vincent Vandeghinste (Eds.).** 2020. *The Future of Academic Lexicography — A White Paper*. Leiden: Instituut voor de Nederlandse Taal. Available online at <https://ivdnt.org/wp-content/uploads/2021/02/The-Future-of-Academic-Lexicography-A-White-Paper.pdf>.
- Tarp, Sven.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica: International Annual for Lexicography / Revue Internationale de Lexicographie / Internationales Jahrbuch für Lexikographie* 28(1): 253-267.
- Tarp, Sven and Rufus H. Gouws.** 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- Thieberger, Nick.** 2015. The Lexicography of Indigenous Languages in Australia and the Pacific. Hanks, Patrick and Gilles-Maurice de Schryver (Eds.). 2015. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.
- Tono, Yukio.** 2022. The Lexicography of Japanese. Hanks, Patrick and Gilles-Maurice de Schryver (Eds.). 2022. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.
- Wiegand, Herbert E.** 1998. *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin: Walter de Gruyter.
- Zaenen, Annie.** 2002. Musings about the Impossible Electronic Dictionary. Corréard, Marie-Hélène (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 230-244. s.l.: Euralex.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography* (Janua Linguarum Series Maior 39). Prague / The Hague: Academia / Mouton.
- Zgusta, Ladislav.** 1991. Probable Future Developments in Lexicography. Hausmann, Franz J., Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (Eds.). 1991. *Wörterbücher / Dictionaries / Dictionnaires, Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie* (Handbücher zur Sprach- und Kommunikationswissenschaft, HSK 5.3): 3158-3168. Berlin: Walter de Gruyter.

Phonostilistische und regionale Variation der deutschen Aussprache in den orthoepischen Nachschlagewerken der dritten Generation

Krzysztof Nycz, *Institut für Germanistik, Universität Rzeszów,
Rzeszów, Polen (knycz@ur.edu.pl)*

und

Zygmunt Tęcza, *Institut für Germanistik, Universität Rzeszów,
Rzeszów, Polen (ztecza@ur.edu.pl)*

Abstract: Mit der Veröffentlichung des *Deutschen Aussprachewörterbuchs* (2009) und der komplett überarbeiteten siebten Ausgabe des *DUDEN-Aussprachewörterbuchs* (2015) sind im 21. Jahrhundert zwei orthoepische Nachschlagewerke erschienen, die der aktuellen Landschaft der deutschen Aussprachewörterbücher in vielerlei Hinsicht eine neue Dimension verliehen. Aus diesem Grund halten wir es für gerechtfertigt, diese als Aussprachewörterbücher einer neuen Generation aufzufassen.

Die beiden hier zu analysierenden Wörterbücher zeichnen sich durch neuartige medientechnische Lösungen und die erstmalige Einbeziehung multimedialer Komponenten aus. Vor allem aber präsentieren sie eine gewandelte, deutlich erweiterte und realistischere Auffassung des phonetischen Standards, die der phonostilistischen Vielfalt der deutschen Aussprache Rechnung trägt. Dabei wird auch zum ersten Mal in der deutschen Phonolexikographie die polyzentrische und polyareale Natur des Deutschen anerkannt, indem die Standardaussprachen Österreichs und der Schweiz sowie deren subnationale bzw. regionale Varietäten berücksichtigt und zum Gegenstand der Beschreibung gemacht werden. Diese beiden grundlegenden Neuerungen der zur Debatte stehenden Aussprachewörterbücher werden im vorliegenden Aufsatz einer kritisch-vergleichenden Betrachtung unterzogen.

Schlagwörter: PHONOLEXIKOGRAPHIE, AUSSPRACHEKODIFIZIERUNG, AUSSPRACHEWÖRTERBUCH, DEUTSCHE AUSSPRACHE, STANDARD AUSSPRACHE, PHONETISCHER STANDARD, PHONETISCHE VARIATION, PHONOSTILISTISCHE VARIATION, REGIONALE VARIATION, SPRACHLICHER PLURIZENTRISMUS

Abstract: Phonostylistic and Regional Variation of German Pronunciation in the Third-generation Orthoepic Dictionaries. By publishing the *Deutsches Aussprachewörterbuch* (2009) and the seventh, completely revised and updated edition of the *DUDEN Aussprachewörterbuch* (2015), two orthoepic reference works have thus far been released in the 21st century that in many respects add a new dimension to the current dictionary landscape in the field

of German pronunciation. For this reason, we consider it justified to regard them as pronunciation dictionaries of a new generation.

The two dictionaries to be discussed in this paper are remarkable for their novel editorial solutions and the unprecedented inclusion of a number of multimedia components. But more crucially, they present a changed, significantly expanded and more realistic understanding of the phonetic standard that takes into account the phonostylistic diversity in German pronunciation. Furthermore, and no less importantly, they do also acknowledge, for the first time in German phonolexicography, the polycentric and polyareal nature of German by including the standard pronunciations of Austria and Switzerland, as well as their subnational or regional varieties, and making them subjects of description. These two fundamental innovations of both dictionaries receive a critical comparative examination in the present article.

Keywords: PHONOLEXICOGRAPHY, PRONUNCIATION CODIFICATION, PRONUNCIATION DICTIONARY, GERMAN PRONUNCIATION, STANDARD PRONUNCIATION, PHONETIC STANDARD, PHONETIC VARIATION, PHONOSTYLISTIC VARIATION, REGIONAL VARIATION, LINGUISTIC PLURICENTRISM

1. Einleitung. Das Ziel der Untersuchung

Die Landschaft der deutschen orthoepischen Nachschlagewerke hat in den vergangenen Jahren eine so wesentliche Bereicherung wie schon lange nicht mehr erfahren. Im Abstand von nur sechs Jahren sind nämlich zwei in jeder Hinsicht moderne Werke erschienen, die der gegenwärtigen deutschen Phonolexikographie eine komplett neue Dimension verleihen. Es handelt sich um folgende Werke:

- *Deutsches Aussprachewörterbuch* von 2009 (nachstehend DAWB) und
- *DUDEN. Das Aussprachewörterbuch* (Bd. 6 der DUDEN-Reihe) in seiner jüngsten, komplett überarbeiteten und aktualisierten siebten Auflage von 2015 (nachstehend DU-15).

Dabei hat jedes der beiden Wörterbücher auch seine Vorgeschichte, die bis in die 1960er Jahre zurückreicht. Das DAWB steht in der Nachfolge des *Großen Wörterbuchs der deutschen Aussprache* von 1982 (im Weiteren GWDA), welches wiederum eine erweiterte Form des zum ersten Mal 1964 in Leipzig erschienenen *Wörterbuchs der deutschen Aussprache* darstellte, und geht — wie im Einführungsteil explizit ausgeführt — bei vergleichbarer Zielsetzung von den gleichen konzeptionellen und methodologischen Grundpositionen wie das GWDA aus (vgl. DAWB: 15). Beim DU-15 wiederum ist eine Kontinuität im Verhältnis zu den früheren Ausgaben des DUDEN-Aussprachewörterbuchs von 1962, 1974 und 1990 sowie drei weiteren, um die Jahrtausendwende erschienenen (aber kaum mehr veränderten) Auflagen allein schon durch den redaktionellen Rahmen gewährleistet.

Doch sind die hier zur Debatte stehenden Auflagen von DAWB und DU-15 schon autonome Nachschlagewerke, die durch moderne editorische Lösungen

auffallen, neuartige, multimediale Komponenten — dem digitalen Zeitalter entsprechend — enthalten, vor allem aber eine gewandelte Auffassung des phonetischen Standards samt dessen innerer phonostilistischer und geografischer Differenzierung präsentieren und sich zum ersten Mal auch mit der Spezifik der Standardaussprache Österreichs und der Schweiz befassen. All das bewirkt, dass wir sie als Aussprachewörterbücher einer neuen Generation betrachten wollen, und zwar der dritten schon, wenn man die Väter der deutschen Phonolexikographie: Wilhelm Viëtor und vor allem Theodor Siebs mit ihren vom Ende des 19. und Anfang des 20. Jahrhunderts stammenden Werken als die erste und die oben aufgezählten Wörterbücher aus der zweiten Hälfte des 20. Jahrhunderts als die zweite Generation definiert (vgl. Tęcza 2018: 104ff.; zur Geschichte der dt. Phonolexikographie s. auch Tęcza und Nycz 2016: 361ff., Nycz und Tęcza 2020a: 74ff.).

Die zuvor genannten grundlegenden Neuerungen von DAWB und DU-15, nämlich ein differenzierteres Verständnis der Standardaussprache sowie die Berücksichtigung ihrer nationalen und regionalen Varietäten, sollen nun im Folgenden einer kritisch-vergleichenden Betrachtung unterzogen werden. Im Einzelnen ist zu prüfen, wie stark diese neue Deutung des phonetischen Standards von der in den älteren Wörterbüchern vertretenen divergiert und inwieweit der nationalen und der regionalen Dimension der deutschen Aussprache sowohl im Einführungsteil als auch im Wörterverzeichnis jedes der beiden untersuchten Nachschlagewerke Rechnung getragen wird.

2. Die Auffassung des phonetischen Standards

Das Hauptanliegen eines Aussprachewörterbuches ist bekanntlich die Normierung und Kodifizierung dessen, was jeweils als *Standardaussprache* bzw. *Standardlautung* verstanden wird. Die beiden Wörterbücher vertreten in dieser Hinsicht zum Teil divergierende Positionen. Im DAWB wird die Standardaussprache gleich zu Beginn des Einführungsteils als „die mündliche Form der Standardvarietät in der Bundesrepublik Deutschland“ definiert, die „in geografischer und sozialer Hinsicht über eine weite Geltung [verfügt] und [...] insbesondere durch die elektronischen Medien verbreitet [wird]“ (DAWB: 6). Die so verstandene Standardaussprache:

- ist **dialektneutral**, was sich darin zeigt, dass sie keine mundartlichen oder regional gefärbten umgangssprachlichen Aussprachevarianten aufweist; sie kann folglich überregional in allen sozialen Gruppen und von jeder Person mit Deutsch als Muttersprache verstanden werden;
- wird **bevorzugt in offiziellen öffentlichen Situationen** gebraucht und erwartet, in denen sie auch als „prestigefördernd“ empfunden wird;
- ist durch **unterschiedliche Grade der Artikulationspräzision** gekennzeichnet (= phonostilistisch differenziert), was ihre Anwendung auch im nicht öffentlichen Bereich ermöglicht;

- ist **kodifiziert** und erfüllt damit eine **normative bzw. regulative Funktion**; dabei weist sie einen unterschiedlichen Grad an Verbindlichkeit auf, wobei diese vor allem für die soziale Gruppe der Berufssprecher(innen) gilt (vgl. ebd.: 7)¹.

Die Autor(inn)en des DU-15 verwenden hierfür den Terminus (*bundesdeutsche überregionale Standardaussprache*) und beziehen ihn auf die überregionale Aussprachevarietät, wie sie „z. B. in bundesdeutschen Nachrichtensendungen praktisch umgesetzt wird“ (DU-15: 30). Sie fügen ergänzend hinzu, dass die fragliche Varietät „heute in großem Umfang auch von nicht in Sprachberufen Tätigen verwendet wird“, zugleich räumen sie aber ein, dass sie — obwohl im gesamten deutschen Sprachraum als vorbildhaft akzeptiert — „offensichtlich nur begrenzte Wirkung [hat], da sie von der sprachlichen Allgemeinheit nur teilweise und in von Region zu Region und individuell unterschiedlichem Umfang aktiv umgesetzt wird“ (ebd.). Dies resultiert darin, dass die Standardaussprache, wie sie tatsächlich von der Bevölkerung verwendet wird, „auch in formellen Situationen in mehr oder weniger großem Umfang, bei einzelnen Wörtern oder systematisch, von der überregionalen Standardaussprache abweicht“ (ebd.). Vor diesem Hintergrund postulieren die Autor(inn)en des DU-15, sämtliche Ausspracheformen, die von deutschen Muttersprachler(inne)n in formellen Sprechsituationen gebraucht und als situationsangemessen eingestuft werden, als standardsprachlich zu betrachten (vgl. ebd.) und damit die Kodifikation auf den tatsächlichen Sprachgebrauch zu stützen. Zugleich wird das im DAWB angenommene Konzept, allein die Aussprache von Berufssprecher(inne)n zur Grundlage der Kodifikation zu machen und damit jegliche Variation auszublenken, explizit kritisiert (vgl. ebd.: 31).

Demnach wird die Standardaussprache im DU-15 durch folgende Merkmale charakterisiert:

- **größtenteils überregionalen Charakter:** Die Standardaussprache umfasst *in der Mehrzahl* überregionale Ausspracheformen — aber auch solche, die im deutschsprachigen Raum nur national oder großregional verwendet werden; damit wird das Postulat der Dialektneutralität, wie es im DAWB vorgebracht wurde (s.o.), deutlich abgeschwächt;
- **Realitätsnähe:** Es ist eine Gebrauchsnorm, die der Sprechwirklichkeit nahekommt und neben der als überregional geltenden Aussprache geschulter Berufssprecher(innen) auch die in der Bevölkerung übliche Standardaussprache berücksichtigt; ihr wird ferner das Merkmal der Schriftnähe zugeschrieben — allerdings mit der Einschränkung, dass sie „im Zweifelsfall auch dem tatsächlichen Sprachgebrauch [folgt], wenn dieser Diskrepanzen zu regelhaften Schreibungs-Aussprache-Korrespondenzen aufweist“ (ebd.: 32);
- **innere Differenziertheit bzw. Variabilität:** Die von der Bevölkerung tatsächlich verwendete Standardaussprache weicht auch in formellen Situa-

tionen in mehr oder weniger großem Umfang von der überregionalen Standardaussprache ab (vgl. ebd.: 30).

Das zum Teil unterschiedliche Verständnis des phonetischen Standards im DAWB und im DU-15 schlägt sich des Öfteren sowohl im deskriptiven Teil der beiden Wörterbücher als auch in deren Wörterverzeichnissen in verschiedener Hinsicht nieder.

3. Das Problem phonostilistischer Differenzierung

Im DAWB wird der Standardaussprache zwar das Merkmal der Überregionalität attestiert, zugleich schließen aber die Autor(inn)en ihre Variabilität und stilistische Vielfalt nicht aus. Diese ergeben sich aus der Tatsache, dass die Standardaussprache in unterschiedlichen sozialen Gruppen, unter vielfältigen kommunikativen Bedingungen und in unterschiedlichen Anwendungsgebieten, d. h. in öffentlichen oder privaten, formellen oder weniger formellen Situationen, genutzt und dementsprechend nicht immer in gleicher Form realisiert wird (vgl. DAWB: 98). Andererseits wird ausdrücklich darauf hingewiesen, dass die im gegenständlichen Wörterbuch formulierten Empfehlungen ausschließlich auf Äußerungen in öffentlichen Situationen zutreffen. Da aber auch die Sprechweise in der Öffentlichkeit „unterschiedliche Grade in der Ausprägung von Sprechspannung und Artikulationspräzision“ (ebd.: 99) aufweist, werden im DAWB drei phonostilistische Ebenen der Standardaussprache (bzw. drei Grade der Artikulationspräzision oder Varianten der prosodischen Gestaltung) unterschieden:

- (1) *eine sehr hohe Artikulationspräzision* — bei einem feierlichen Vortrag und in der sprechkünstlerischen Kommunikation, etwa beim Sprechen der Schauspieler(innen) auf der Bühne oder Rezitation von Lyrik (vgl. ebd.: 102);
- (2) *eine hohe bis mittlere Artikulationspräzision* — beim reproduzierenden Sprechen wie Vorlesen von Nachrichtentexten in Funk und Fernsehen, beim öffentlichen Vorlesen, bei wissenschaftlichen und populärwissenschaftlichen Vorträgen sowie in der sprechkünstlerischen Kommunikation, etwa beim figurendarstellenden Sprechen der Schauspieler(innen) im Theater, Hörspiel, Film, bei Synchronisation und im Bereich der Kleinkunst (vgl. ebd.);
- (3) *eine verminderte Artikulationspräzision* — beim freien Sprechen, vor allem in öffentlich geführten Gesprächen in Funk und Fernsehen (Talkshows), in Film- und Hörspieldialogen, im Kabarett o.Ä., aber auch in nichtöffentlichen, inoffiziellen und informellen Gesprächen im Alltag (vgl. ebd.: 99, 104).

Für jede der drei Stufen werden typische Merkmale, ggf. Abweichungen von den anderen Registern, angeführt und anschließend an einzelnen Beispielen illustriert. Recht detailliert wird dabei auf die Aussprachevariante mit *verminderter Artikulationspräzision* eingegangen. Für diesen phonostilistischen Bereich

werden in zwei relativ umfangreichen Unterkapiteln separat Lautschwächungen im Vokalismus sowie Reduktionen und Assimilationen im Konsonantismus behandelt. Sie werden als „Tendenzen“ bezeichnet, die „sich in öffentlich geführten Gesprächen beim Gebrauch der Standardaussprache beobachten lassen“, „exemplarischen Charakter [haben]“ und „meist nicht alle gleichzeitig [vorkommen]“ (DAWB: 105).

Die Beschreibung unterschiedlicher phonostilistischer Varianten in einem Wörterbuch wird dabei wie folgt begründet:

Die Beschreibung phonostilistischer Differenzierungen im Rahmen einer Aussprachekodifikation ist notwendig, weil Verminderungen der Artikulationspräzision, d. h. auch eine Zunahme assimilatorisch bedingter Lautreduktionen, nicht in allen Sprachen nach gleichen Gesetzmäßigkeiten ablaufen. Sie sind daher für den Deutsch lernenden Ausländer ebenso wenig ohne Weiteres erschließbar wie für Sprecher regional gefärbter Umgangssprachen, und sie ergeben sich bei nachlassender Sprechspannung auch nicht von selbst. Umgekehrt bereitet es z. B. der letztgenannten Sprechergruppe häufig Probleme, eine (angemessen) hohe Artikulationspräzision zu realisieren, ohne zu unangebrachten Übertreibungen zu gelangen. (ebd.: 100)

Was die Spezifik der Standardaussprache in besonderen kommunikativen Situationen angeht, so ist im DAWB allein dem Gesang ein eigenes Unterkapitel gewidmet. Dabei werden vorerst Unterschiede zwischen Sprechen und Singen aufgeführt, die sich auf die Funktion der Sprechwerkzeuge, die Form der Melodieführung sowie den Einsatz der prosodischen Mittel beziehen. Anschließend werden Tendenzen im klassischen Solo- und Chorgesang diskutiert, wie sie etwa bei der Bildung der Vokale und Konsonanten zu beobachten sind. In einem separaten Abschnitt werden Gesangstils in der Populär- und Volksmusik sowie deren Spezifika behandelt.

Varianten mit stark reduzierter bzw. besonders geringer Artikulationspräzision, wie sie etwa „bei nicht öffentlichen, inoffiziellen und informellen Gesprächen zwischen miteinander vertrauten Kommunikationspartnern im Alltags- und Privatbereich gebräuchlich“ (ebd.: 104f.) sind, werden dagegen im DAWB nicht mehr berücksichtigt. Da sie „verstärkt fließende Übergänge zu regionalen Umgangssprachen sowie zu Sozio- und Idiolekten“ (ebd.: 105) zeigen, scheinen sie nach Meinung der Autor(inn)en in einem Randbereich der Standardaussprache oder sogar bereits außerhalb derselben zu liegen. Ebenfalls unberücksichtigt bleibt die sog. *Überlautung*, die „z. B. beim Sprechen unter Lärmbedingungen oder auch beim Diktieren im elementaren Deutschunterricht gelegentlich beobachtet werden kann“ (ebd.: 100) und die von den DAWB-Autor(inn)en ausdrücklich als nicht zur Standardaussprache gehörend eingestuft wird.

Dabei ist zu betonen, dass die drei o.g. phonostilistischen Ebenen im DAWB, anders als oben aufgeführt, in einer nicht linear-absteigenden Reihenfolge aufgezählt und beschrieben werden. Als erste erscheint nämlich nicht (wie es im GWDA, dem Vorgänger des DAWB, der Fall war) die *sehr hohe*, son-

dern die *hohe bis mittlere Artikulationspräzision*, die in phonostilistischer Hinsicht die mittlere Stufe bzw. Aussprachvariante darstellt. Die Setzung der hohen bis mittleren Sprechspannung und Artikulationspräzision als Kodifizierungsgrundlage ist als eine klare Prioritätenverschiebung im Verhältnis zum GWDA zu bewerten und sie findet nun zwangsläufig sowohl bei der Beschreibung der Ausspracheregeln im Einführungsteil des Wörterbuchs als auch bei den Ausspracheangaben im Wörterverzeichnis ihren Niederschlag, wo z. B. *geben*, *baden* und *legen* wegen Elision des Schwa-Lautes [ə] nach Lenis-Plosiven konsequent ausschließlich als [g'e:bm], [b'a:dŋ], [l'e:gŋ]², und nicht mehr wie im GWDA alternativ als ['ge:bm] od. [~ bən], ['ba:dŋ] od. [~ dən], ['le:gŋ] od. [~ gən] transkribiert werden.

Im DU-15 wird im Gegensatz zu allen Vorgängerauflagen, in denen verschiedene Formstufen innerhalb der Standardlautung aus Mangel an „einheitlichen und eindeutigen Ergebnissen“ (DUDEN 2000: 35) prinzipiell keine Beachtung fanden, der Differenzierung der deutschen Aussprache auffallend viel Aufmerksamkeit geschenkt. Bereits von den ersten Seiten an wird im einführenden Textteil hervorgehoben, dass „im Deutschen auch in der Standardaussprache (d. h. in formellen Situationen) verschiedene Aussprachevarianten gebräuchlich sind“ und die Autor(inn)en „es als eine wesentliche Aufgabe an[sehen], solche Varianten zu dokumentieren“ (DU-15: 10). Im Kapitel zur Standardaussprache des Deutschen, das mit der indiskutablen Feststellung „Variation ist ein Wesensmerkmal aller lebenden Sprachen“ (ebd.: 29) beginnt, wird die folgende Position der Verfasser diesbezüglich verkündet:

Die Autoren dieser Neuauflage des Duden-Aussprachewörterbuchs sind [...] der Auffassung, dass für die Dokumentation einer möglichst realistischen Beschreibung der Aussprache des Deutschen eine Erweiterung der als kodifikationsrelevant erachteten Situationen und Sprecher unabdingbar ist. Wir erachten es als sinnvoll und nützlich, Sprachformen, die teilweise von Millionen von deutschen Muttersprachlern ganz selbstverständlich in formellen Sprechsituationen (z.B. im Rahmen des Schulunterrichts von Lehrkräften wie Schülern gleichermaßen) verwendet und als situationsangemessen eingestuft werden, als standardsprachlich (im Sinne eines »Gebrauchsstandards«) anzusehen. Das im englischen Sprachraum schon längst etablierte Konzept, den Sprachgebrauch der »educated speakers« zur Grundlage standardsprachlicher Aussprachewörterbücher (und Grammatiken) zu machen, sehen wir darum auch für das Deutsche als zeitgemäßen Schritt zu einem erweiterten und damit realitäts- und gebrauchsnäheren Standard-sprachkonzept. (ebd.: 30f.)

Dieser Deklaration zufolge scheint den Autor(inn)en des DU-15 besonders daran gelegen zu haben, ein möglichst differenziertes und dabei realitätsnahes Bild der Aussprache des Deutschen in formellen Situationen zu liefern. Im Gegensatz zum DAWB wird hier keine hierarchische phonostilistische Schichtung postuliert; stattdessen erfolgt in einem separaten Kapitel eine konzise lineare Darstellung von relevanten Varianten und Variationsphänomenen innerhalb der Standardaussprache. Insgesamt werden rund vierzig verschiedene

Phänomene dieser Art im Bereich der Vokale, Konsonanten und schließlich in den Nebensilben diskutiert. Man erfährt beispielsweise, dass der Glottisschlag in Süddeutschland viel seltener als im übrigen Bundesgebiet vorkommt und in Österreich und der Schweiz weitgehend ungebräuchlich ist (vgl. DU-15: 63); dass die hohen Kurzvokale /ɪ ʏ ʊ/ in Österreich und in der Schweiz geschlossen und gespannt, dabei aber offener/zentralisierter als die entsprechenden Langvokale /i y u/ artikuliert werden; dass die halboffenen Kurzvokale /ɛ œ ɔ/ in Süddeutschland, Österreich und der Schweiz geschlossener, also wie /e ø o/ gesprochen werden; dass die beiden Vokale /a/ und /a:/ in Österreich oft als vordere und sehr offen artikuliert Laute realisiert werden (vgl. ebd.: 64); dass die Substitution von [ɛ:] durch [e:] in Nord- und Ostdeutschland sowie in Österreich allgemein ist, in der Schweiz aber konsequent zwischen [ɛ:] und [e:] unterschieden wird (vgl. ebd.: 64f.); dass in Nord- und Mitteldeutschland in vielen einsilbigen Wörtern mit theoretisch langem Vokal auch ein Kurzvokal gebräuchlich ist (*Rad* [rat], *Bad* [bat], *Tag* [tak] usw.; vgl. ebd.: 65); dass es in Ost- und Südösterreich oft zur Nasalisierung von Vokalen vor einem Nasal kommt, usw.

Im Bereich der Konsonanten werden v. a. Assimilationen, plosivische Verschlusslösung von auslautendem [ŋ] als [ŋk] in Norddeutschland, [ç]/[x]- oder [ç]/[ʃ]-Variation entsprechend in Österreich/Süddeutschland und in Mitteldeutschland, Nichteintreten der Auslautverhärtung in Österreich, Süddeutschland und der Schweiz, [s]/[ʃ]-Variation vor [t] und [p] im Silbenanlaut, Lenisierung von Fortiskonsonanten im Inlaut (vgl. ebd.: 67-72) usw. detailliert besprochen.

Bei dem Nebenton beziehen sich die Ausführungen u. a. auf variierende Realisationen des Schwa-[ə] in Süddeutschland, Österreich und der Schweiz, variierende Vokalquantität in einigen deutschen Suffixen (-tum, -sam, -ik, -it, -iz) oder Variation zwischen gespannten und ungespannten Vokalen in unbetonten Silben vor dem Hauptakzent (vgl. ebd.: 72f.).

Als Auswahlkriterien für die zu berücksichtigenden Variationsphänomene galten den Autor(inn)en primär „Vorkommenshäufigkeit und großregionale bzw. nationale Verbreitung“ (ebd.: 63), wobei der Gebrauch einer Variante in formellen Sprechsituationen ein zentraler Maßstab für die Einstufung als standardsprachlich war. Dass die Grenzen des phonetischen Standards hierbei weiter denn je zuvor gezogen sind, ist an einigen im besagten Kapitel aufgeführten Ausspracheformen zu erkennen, deren postulierte Richtigkeit — oder eben Standardzugehörigkeit — vielen gebildeten Muttersprachler(inne)n wohl ziemlich suspekt vorkommen wird, wie etwa [ta:x], [tak] und [tax] neben dem evidenten [ta:k]; [fli:çt] neben [fli:kt]; [dɪŋk] neben [dɪŋ]; [e:s] neben [ɛs]; [be'darf] und [ge'maxt] neben [bə'darf] und [gə'maxt]; [vəɪs] (als Farbbezeichnung) neben [vaɪs] u.a. Die Autor(inn)en fügen ergänzend hinzu:

Einige der hier aufgeführten Varianten werden von Berufssprecher(inne)n nicht oder nur selten verwendet, sind aber in der Bevölkerung weit verbreitet. Die Zusammenstellung enthält sprachliche Merkmale mit unterschiedlicher perzep-

tiver Auffälligkeit, unterschiedlicher geographischer Reichweite und (mit den beiden Aspekten verbunden) unterschiedlicher soziolinguistischer Akzeptanz. (ebd.: 63)

Die im Einführungsteil des Wörterbuchs aufgeführten diversen Varianten der Standardaussprache kommen allerdings nur begrenzt im Wörterverzeichnis des DU-15 zum Tragen, da sie hier „aus darstellungsökonomischen Gründen [...] weitgehend ausgeblendet werden“ (ebd.: 32) mussten. Einschlägige Angaben erscheinen vor allem in den sog. Infokästen, in denen mehrheitlich entweder statistische Ergebnisse einer Umfrage zur Akzeptanz alternativer Aussprachevarianten durch die Sprachbenutzer(innen) gezeigt oder solche alternativen Varianten kurz diskutiert und problematisiert werden (vgl. Abb. 1 und 2).

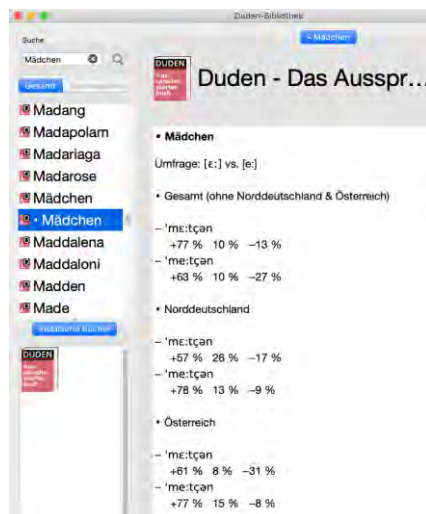


Abb. 1: DU-15: Infokasten zum Lemma *Mädchen*³

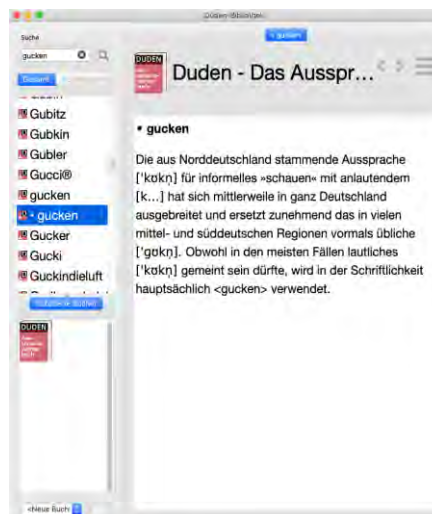


Abb. 2: DU-15: Infokasten zum Lemma *gucken*

Wie die meisten der oben angeführten Beispiele eindrücklich zeigen, geht die phonostilistische Differenzierung des Standards in der Darstellung des DU-15 fast immer mit einer territorialen Zuordnung der gegebenen artikulatorischen Variante einher, oft wird sie gar von der letzteren dominiert. Auf die plurizentrische Natur des Deutschen und die Darstellung dessen nationaler und regionaler Varietäten in den zur Debatte stehenden Aussprachewörterbüchern wird nun ausführlicher im nächsten Abschnitt eingegangen.

4. Reflexe des Plurizentrismus und der regionalen Variation der Standardaussprache

Der Plurizentrismus der deutschen Sprache bedeutet, dass sie zwar alle deutsch-

sprachigen Gebiete verbindet, dabei jedoch weder in geschriebener noch in gesprochener Form einheitlich ist (vgl. DAWB: 229)⁴. Vielmehr besteht sie aus territorialen Varietäten, die in Deutschland, Österreich, der deutschsprachigen Schweiz, Liechtenstein, Luxemburg, Ostbelgien und Südtirol gesprochen werden. Als nationale Vollzentren der deutschen Sprache gelten dabei Deutschland, Österreich und die Schweiz, weil dort „die standardsprachlichen Besonderheiten in eigenen Nachschlagewerken, vor allem Wörterbüchern, festgehalten und autorisiert sind“ (Ammon et al. 2016: XXXIX). Dessen ungeachtet stand in Aussprachewörterbüchern des Deutschen noch bis vor kurzem traditionell die bundesdeutsche Standardaussprache nicht nur im Mittelpunkt, sondern sie war gar der einzige Gegenstand der Beschreibung, ohne dass auf die anderen Varietäten überhaupt eingegangen wurde. Dies gilt auch für die Vorgänger der zur Debatte stehenden Aussprachewörterbücher (GWDA und sämtliche früheren Ausgaben des DUDEN-Wörterbuchs), in denen die Betrachtung der nationalen Varietäten ausgeblieben ist⁵.

Einen Durchbruch in dieser Hinsicht hat erst das 2009 erschienene DAWB gebracht, das der Standardaussprache in Österreich und der Schweiz zum ersten Mal gebührende Aufmerksamkeit schenkt. Den Wandel trägt dann auch das DU-15 mit, indem es Deutsch als eine pluri- bzw. polyzentrische Sprache anerkennt und die phonetischen Eigentümlichkeiten der oben genannten nationalen Varietäten, wenn auch in anderer Weise und in kleinerem Umfang als das DAWB, dokumentiert.

Im DAWB erfolgt die Beschreibung der österreichischen und der schweizerdeutschen Standardaussprache in zwei autonomen, relativ umfangreichen Hauptkapiteln im Einführungsteil (entsprechend Wiesinger 2009 und Haas und Hove 2009). Für das österreichische Deutsch werden zunächst — unter deutlichem Rückgriff auf die bereits für die Standardaussprache Deutschlands vorgenommene phonostilistische Dreiteilung — drei Register unterschieden:

- Register I: die gehobene Standardaussprache geschulter Sprecher(innen),
- Register II: die gemäßigte Standardaussprache geschulter Sprecher(innen),
- Register III: die Standardaussprache der Laien (vgl. DAWB: 235).

Die Register I und II werden „im Österreichischen Rundfunk und Fernsehen (ORF) von Ansager(inne)n, Nachrichtensprecher(inne)n und Moderator(inn)en verwendet“ (ebd.), wobei je nach Typ der Sendung Sprecher(innen) mit unterschiedlichen Registern eingesetzt werden können (z. B. Sprecher(innen) mit Register I meist in literarischen Sendungen und Hörspielen; Moderator(inn)en mit „oft nur geringer oder keiner Sprechausbildung“ (ebd.), d. h. mit Registern II/III, in spezifischen Sendungen). Was trotz der Unterschiede Akzeptanz bei der Bevölkerung garantiere, sei die österreichische Herkunft der Sprecher(innen) und ihre heimische Intonation (vgl. ebd.). Demgegenüber gilt das Register III als regionales Hochdeutsch, in das „individuell auch umgangssprachliche und teilweise sogar einzelne unmittelbare dialektale Lautungen einfließen [können]“ (ebd.). Dieses spiele in Rundfunk und Fernsehen insofern eine Rolle, als es von

Persönlichkeiten des öffentlichen Lebens aus dem politischen, kulturellen und wirtschaftlichen Umfeld verwendet wird.

Auf diese Gliederung folgt im besagten Kapitel eine — wie von dessen Autor erklärt — deskriptive, nicht präskriptive Darstellung von Aussprachephänomenen verschiedener Art. Die Bestandsaufnahme umfasst:

- (a) registerspezifische Realisierung von kurzen und langen Vokalen, u.a. Qualitätsunterschiede zwischen [i] – [u] – [y] und [i:] – [u:] – [y:], [ɛ] – [ɔ] – [œ] und [e:] – [o:] – [ø:], [a] und [a:], Realisierung von langem <ä>, Diphthongen und Nasalvokalen,
- (b) Besonderheiten im Bereich der Konsonanten, die v. a. die Intensität der Plosive und Frikative, Abweichungen in der Realisierung von <l>, verschiedene Artikulationen von <r> sowie Assimilierungen bzw. Tilgungen von Konsonanten betreffen,
- (c) abweichende Realisierung von den <e>-hältigen unbetonten Nebensilben und des Suffixes <-ig> sowie
- (d) Besonderheiten bei der Aussprache von Fremdwörtern und bei der Akzentuierung.

Darüber hinaus enthält das der Standardaussprache in Österreich gewidmete Kapitel ein alphabetisches Verzeichnis von insgesamt 90 Fremdwörtern bzw. zusammengesetzten und abgeleiteten Erbwörtern, die eine von dem bundesdeutschen Standard abweichende Akzentuierung aufweisen.

Anschließend wird die österreichische Standardaussprache — „da die Register I und II als gehobene und gemäßigte Standardaussprache mehr oder minder geschulter Sprecher breite Akzeptanz in ganz Österreich finden, während Register III als ‚regionales Hochdeutsch‘ der Laien in seinen Varianten auf Teilräume beschränkt ist“ — als ein „Kompromiss aus I und II“ (ebd.: 252) definiert und beschrieben. Die daraufhin formulierten „Empfehlungen für eine österreichische Standardaussprache“ (ebd.) betreffen sämtliche vorstehend erwähnten Ausspracheerscheinungen und haben in diesem Fall schon präskriptiven Charakter.

Das Kapitel zur Aussprache in der deutschsprachigen Schweiz ergibt indessen ein weitgehend anderes Bild. Gleich zu Beginn wird ausdrücklich betont, dass in der Schweiz im mündlichen Sprachgebrauch der Dialekt dominiert und die Deutschschweizer(innen) im Alltag ausschließlich ihren alemannischen Dialekt sprechen. Die schweizerische Standardsprache wird hingegen „im mündlichen Bereich in der Schule, in offiziellen Situationen und in künstlerischen Darbietungen verwendet und [...] mit Personen aus Deutschland oder Österreich und mit Fremdsprachigen gesprochen“ (ebd.: 259f.). Ferner wird darauf hingewiesen, dass „Übergangsformen zwischen dem Dialekt und der Standardsprache, also regional gefärbte Umgangssprachen, wie sie in anderen Teilen des deutschen Sprachraums vorkommen“ (ebd.), in der Schweiz nicht existieren. Dies mag auch der Grund dafür sein, dass im DAWB für die Aus-

sprache in der Schweiz keine innere phonostilistische Stratifizierung, die der zuvor beschriebenen Schichtung des bundesdeutschen und des österreichischen Standards analog wäre, vorgeschlagen wird.

Abgesehen von dem in der Kapitelüberschrift verwendeten Ausdruck (*Standardaussprache in der deutschsprachigen Schweiz*) ist in den weiteren Ausführungen auch kaum von einer schweizerischen Standardaussprache die Rede. Stattdessen wird der Terminus *Aussprachekonvention* benutzt:

Innerhalb der Deutschschweiz besteht eine recht weitgehende Übereinkunft darüber, welche Aussprachevarianten für die schweizerische Standardsprache angemessen sind und welche nicht. Es existiert eine Konvention, die nicht von jemandem festgelegt wird, sondern dadurch entsteht, dass die Mitglieder einer Gruppe — in diesem Fall die Deutschschweizerinnen und -schweizer — ganz allgemein gesehen zu Konformität im Verhalten neigen. Diese Konvention [...] umfasst ein Set von Varianten, die in der entsprechenden Gemeinschaft als angemessen und unmarkiert gelten. Die Übereinkunft darüber, welche lautlichen Varianten für die Schweizerinnen und Schweizer bei der Verwendung der Standardsprache angebracht sind, wird hier als *Aussprachekonvention* bezeichnet. (ebd.: 260)

Zugleich heben der Autor und die Autorin des betreffenden DAWB-Kapitels hervor, dass die so definierte Aussprachekonvention „keineswegs rigide“ sei und „in vielen Bereichen Variationen“ (ebd.) zulasse, ihre Existenz aber dazu führe, „dass sich die Deutschschweizer wesentlich einheitlicher verhalten, als sie es tun würden, wenn keine solche Konvention existierte“ (ebd.: 261). Die für die Schweiz empfohlenen Aussprachevarianten umfassen einerseits Formen, die mit dem bundesdeutschen und österreichischen Standard übereinstimmen, und andererseits solche, die von jenem Standard abweichen:

Gewisse Aussprachevarianten, die in Deutschland nicht üblich sind, werden in der Schweiz durchaus als standardsprachlich erachtet. Umgekehrt wird die Verwendung bestimmter Deutschland-spezifischer Varianten durch Deutschschweizer als unangemessen bewertet. (ebd.)

Bei der Einbeziehung zusätzlicher, für die Schweiz empfohlener Varianten dienen ihre Vorkommenshäufigkeit und Eindeutigkeit bzw. Verständlichkeit als Kriterien; berücksichtigt werden also Varianten, die „in der Standardsprache gebildeter Deutschschweizer eine gewisse Häufigkeit haben und welche die Verständlichkeit nicht beeinträchtigen“ (ebd.: 262).

Ähnlich wie im Kapitel zur Aussprache in Österreich folgen auf diese einführnden Überlegungen zur Sprachsituation in der deutschsprachigen Schweiz Aussprachehinweise für dieselbe. Die Darstellung umfasst zuerst Besonderheiten im Bereich der Vokale (u. a. das Ausbleiben des Glottisschlags im absoluten Wortanlaut, nach Präfixen und in Zusammensetzungen, landesspezifische phonetische Realisierung von Lang- und Kurzvokalen, a-Lauten und Diphthongen), dann der Konsonanten (intervokalische Konsonanten, Lenisplosive im absoluten Auslaut, vor einem Vokal bzw. vor Sonoranten, ferner die

s-, r-, *ich*- und *ach*-Laute und die Buchstabenverbindung <qu>), bei den Affixen und Fremdwörtern sowie Besonderheiten im Bereich der Wortakzentuierung. Zum Abschluss enthält das dem Schweizerdeutsch gewidmete Kapitel des DAWB eine transkribierte Wortliste mit gut 200 Einträgen, die die phonetische Eigenart dieser Varietät anschaulich illustrieren.

Eine derart komplexe und geordnete Darstellung der Standardaussprache in Österreich und in der deutschsprachigen Schweiz liegt im DU-15 hingegen nicht vor. Dafür gehen die Autor(inn)en dieses jüngsten Wörterbuches an einer Stelle explizit auf die beiden genannten Kapitel des DAWB ein, und zwar mit einem wohl als anerkennend zu deutenden Hinweis auf die Berücksichtigung der nationalen Varietäten und deren ausführliche Darstellung im DAWB sowie vor allem auf die Orientierung der Kodifikation am realen Sprachusus:

Kodifikationen dieser Art, die sich primär auf den tatsächlichen Sprachgebrauch von nicht in Sprechberufen Tätigen stützen, sind für das Deutsche im Bereich der Aussprache bereits von Walter Haas / Ingrid Hove (2009) für die Standardaussprache in der deutschsprachigen Schweiz sowie von Peter Wiesinger (2009) für die Standardaussprache in Österreich vorgelegt worden. Die beiden genannten Darstellungen sind selbstständige Beiträge im Einleitungsteil des von Krech et al. (2009) erarbeiteten »Deutschen Aussprachewörterbuchs«. (DU-15: 31)

Daraufhin wird unter Rückgriff auf Haas und Hove (2009: 259) als Fazit festgehalten, dass die Orientierung am wirklichen Sprachgebrauch „ein gewisses Maß an Variation [zulässt]“ und die Berücksichtigung der „Besonderheiten der deutschen, der österreichischen und der schweizerischen Standardsprache“ (DU-15: 31) zur Folge hat. Diese werden dann in der Tat im besagten Kapitel zur Variation in der Standardaussprache dokumentiert. Die Darstellung erfolgt hier allerdings nicht nach Regionen — wie dies im DAWB der Fall war —, sondern nach Aussprachephänomenen. Die einzelnen Aussprachvarianten werden dabei entweder einer **nationalen** (Deutschland, Österreich, Schweiz) oder einer **großregionalen** (Norddeutschland, Mitteldeutschland, Süddeutschland; Süden des deutschen Sprachraums) bzw. **regionalen** (Thüringen, Sachsen; Nordbayern, Südbayern, Ostbayern etc.) Varietät des Deutschen zugeordnet, bis hin zu kleinräumlich verbreiteten, **territorial sehr begrenzten (Sub-)Varietäten** (z. B. absoluter Nordwesten Deutschlands, Ostfriesland, Küstenregionen Norddeutschlands, Südrand Deutschlands; Süd- und Ostrand Bayerns u. Ä.). Auch innerhalb der österreichischen und der deutschschweizerischen Nationalvarietät sind immer wieder Hinweise auf den Status der jeweiligen Aussprachevariante in den regionalen Subvarietäten (Westschweiz, Westösterreich, Südösterreich, Tirol) zu finden⁶.

Auf konkrete regionale Differenzen wird ebenfalls — wie bereits im vorigen Abschnitt angedeutet — in den im Wörterverzeichnis des DU-15 verstreuten Informationsmodulen hingewiesen. Diese bringen zum einen prozentuale Ergebnisse einer Online-Umfrage zur Akzeptanz alternativer Aussprachevarianten ausgewählter Wortformen (wie z. B. *Mädchen*, *Mathematik*, *Nische*, *stilistisch*),

deren Teilnehmer(innen) verschiedene Regionen Deutschlands, Österreichs und der alemannischen Schweiz repräsentierten. Zum anderen werden in den Infokästen nationale bzw. (groß-)regionale Aussprachevarianten aufgeführt und kurz diskutiert (z.B. *höchst*, *vier*, *un-*, *-ig*; vgl. Abb. 3).

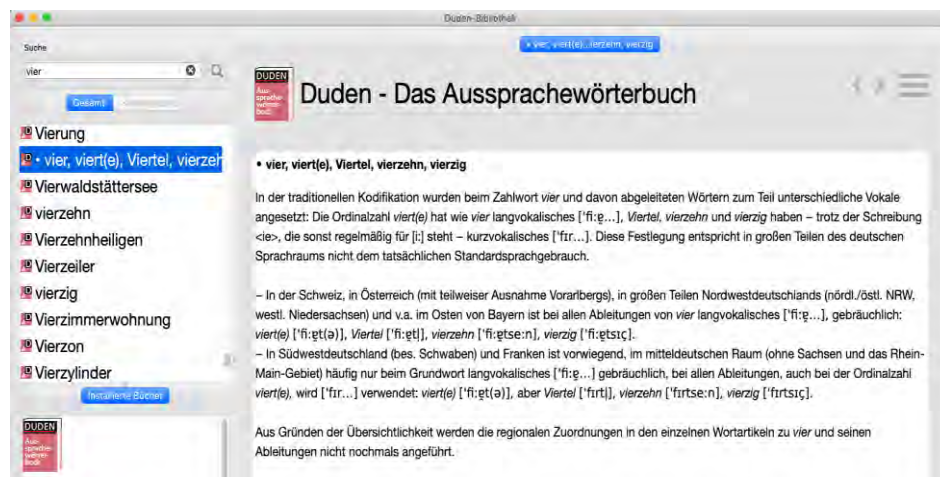


Abb. 3: DU-15: Infokasten zur Aussprache von *vier* im freien Gebrauch und als Wortbildungselement

Entsprechend markierte nationale, d. h. österreichische und schweizerische, sowie (groß-)regionale Aussprachevarianten finden sich darüber hinaus gelegentlich auch im Wörterverzeichnis innerhalb einzelner Lemmata. Vergleicht man z. B. Einträge wie *Herd* oder *Chemie* im DAWB, die jeweils lediglich aus dem in Alphabetbuchstaben geschriebenen Stichwort und dessen Transkription bestehen (**Herd** he:ʁt, **Chemie** ʃem'i:), so fallen deren Pendants im DU-15 wesentlich informativer aus (s. Abb. 4 und 5):



Abb. 4: DU-15: Das Lemma *Herd*

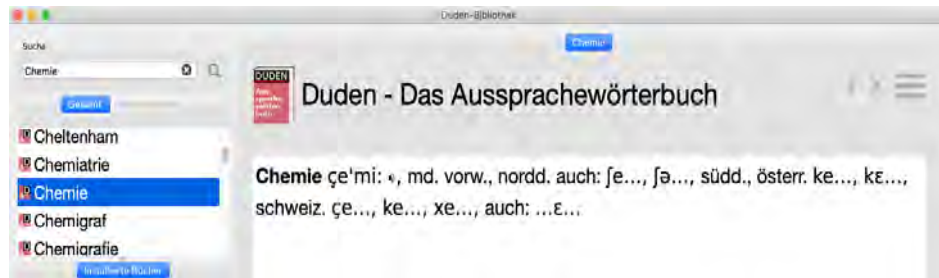


Abb. 5: DU-15: Das Lemma *Chemie*

Sowohl im einführenden Textteil als auch im Wörterverzeichnis wird zudem im DU-15 mit relativierenden Quantifikatoren vor Regionalangaben gearbeitet (*bes., vorw., oft, auch* und *selten*), um die relative Häufigkeit der betreffenden Variante allgemein oder in der spezifizierten Region zu bestimmen (vgl. Abb. 4 und 5; vgl. DU-15: 18).

Auch wenn also im DU-15 auf eine separate Darstellung der deutschen Standardaussprache Österreichs und der Schweiz verzichtet wird, trägt es mit seinen wiederholten Bezugnahmen auf nationale, oft aber auch subnationale und regionale Aussprachevarianten in besonderem Ausmaß der Tatsache Rechnung, dass Deutsch eine polyzentrische und polyareale Sprache ist.

5. Fazit

Das moderne, deutlich erweiterte Verständnis des phonetischen Standards, wie es in den beiden deutschen Aussprachewörterbüchern der dritten Generation — wenn auch in unterschiedlicher Weise und in verschiedenem Umfang — zum Tragen kommt, verdient ohne Zweifel Würdigung. Es schlägt sich zum einen in der größeren Realitätsnähe durch die (Mit-)Anerkennung niedrigerer phonostilistischer Register als Kodifizierungsgrundlage sowie, besonders im DU-15, (Mit-)Berücksichtigung vieler bisher als nicht-standardkonform angesehenen, jedoch in öffentlichen Alltagssituationen tatsächlich verwendeter Aussprachevarianten nieder. Zum anderen findet es seinen Ausdruck in der Sanktionierung der polyzentrischen und polyarealen Natur des Deutschen durch die eingehende systematische Beschreibung der drei nationalen Varietäten (DAWB) bzw. sachliche Bestandsaufnahme subnationaler, groß- und kleinregionaler Aussprachevarianten (DU-15).

Der letztgenannte Aspekt — die solide Auseinandersetzung mit der räumlichen Dimension der (Standard-)Aussprache — stellt auch im internationalen Maßstab eine bedeutende Leistung dar. Es reicht nur ein flüchtiger Blick auf die drei führenden Aussprachewörterbücher des Englischen — einer Sprache, deren Plurizentrismus und regionale Differenziertheit noch unvergleichlich stärker ausgeprägt sind und somit auch entsprechend mehr Beachtung ver-

dienen müssten –, um sich zu überzeugen, dass sie in dieser Hinsicht den beiden deutschen Nachschlagewerken erheblich unterlegen sind. In den Einführungskapiteln von Jones (2003), Wells (2008) sowie Upton et al. (2003), welche übrigens auch als Ganzes viel bescheidener als die von DAWB und DU-15 ausfallen, wird prinzipiell nur auf die signifikanten Unterschiede zwischen dem britischen Standard (*Received Pronunciation* bzw. *BBC English*) und dem allgemeinen US-amerikanischen Usus (*General American*) eingegangen. Einige Hinweise auf subnationale Spezifika innerhalb der beiden Nationalvarietäten finden sich zwar in Wells (vgl. 2008: XIXff.) und in Upton et al. (hauptsächlich zum Amerikanischen, vgl. 2003: XIVff.), doch sind sie insgesamt sehr knapp gehalten. In den Wörterverzeichnissen wird neben der jeweiligen britischen Aussprache auch die amerikanische angegeben (in Jones und Wells nur, wenn sie von der ersteren abweicht, in Upton et al. durchgehend), sonst gibt es aber keine weiteren Notationen mit geografischem Charakter. Werden bei einem Lemma mehrere britische oder mehrere amerikanische Aussprachevarianten verzeichnet, so geschieht es ohne jegliche räumliche Zuordnung innerhalb der betreffenden nationalen Varietät. Dieser Defizite sind sich die Wörterbuchautor(inn)en wohl bewusst:

A pronouncing dictionary that systematically presented the pronunciations of a range of regional accents would be very valuable, but it would be very much bigger than the present volume and the job of ensuring an adequate coverage which treated all accents as equally important would have taken many years. (Jones 2003: VI)

Wie nun aus der im vorliegenden Beitrag erfolgten Gegenüberstellung von DAWB und DU-15 ersichtlich sein dürfte, ist diese Arbeit für die deutsche Sprache schon in beachtlichem Maße geleistet worden.

Endnoten

1. Zur Auffassung des phonetischen Standards im GWDA und den älteren Ausgaben des DUDEN-Aussprachewörterbuches s. Tęcza und Nycz 2016: 384ff. sowie Nycz und Tęcza 2020b: 463ff.
2. Zu der untypischen Markierung des Wortakzents im DAWB vgl. z. B. Nycz und Tęcza 2020b: 481f.
3. Alle Abbildungen in diesem Beitrag stammen aus der digitalen Version des DU-15 (s. Literaturverzeichnis).
4. Nach Ansicht der DAWB-Autor(inn)en wird eine vollkommen einheitliche Sprache auch nicht angestrebt, diese sei „weder notwendig noch sinnvoll“, denn es entspreche „einem Bedürfnis der Deutschsprachigen, die in ihrer Sprachgemeinschaft übliche Varietät realisieren zu können“ (S. 259). Eine gewisse Variation schließt allerdings die Kodifizierung nicht aus — „(...) eine Kodifizierung [ist] notwendig, denn sie dient im Unterricht und in der Sprachverwendung in Beruf und Alltag als Richtlinie und trägt zur Legitimation von Varianten bei“ (ebd.).

5. Dieser Zustand wurde auch lange Zeit — bis zum Erscheinen des DAWB — verschiedentlich beklagt; vgl. z. B. Back 1995: 282, Berend 2004: 147, Lameli 2005: 500, Ehrlich 2009: 8ff.
6. So bekommt man beispielsweise zu wissen, dass die mittelhohen Langvokale /e: ø: o:/ vor allem in der Westschweiz häufig offen, ähnlich wie /ɛ: œ: ɔ:/ gesprochen werden; dass das lange /o:/ in Tirol zum Teil zentralisiert gesprochen wird; dass es in Ost- und Südösterreich oft zur Nasalisierung von Vokalen vor einem Nasal kommt usw. (vgl. DU-15: 63ff.).

Literaturverzeichnis

I. Aussprachewörterbücher (chronologisch, mit Siglen)

- GWDA:** *Großes Wörterbuch der deutschen Aussprache*. 1982. Hrsg. von E.-M. Krech, E. Kurka, H. Stelzig, E. Stock, U. Stötzer und R. Teske unter Mitw. von K. Jung-Alsen. Leipzig: Bibliographisches Institut.
- DUDEN 2000:** *Duden Aussprachewörterbuch. Wörterbuch der deutschen Standardaussprache* (= DUDEN Bd. 6). 2000. Bearb. von M. Mangold in Zusammenarbeit mit der Dudenredaktion. 4., neu bearbeitete und aktualisierte Auflage. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- DAWB:** *Deutsches Aussprachewörterbuch*. 2009. Von E.-M. Krech, E. Stock, U. Hirschfeld und L. Ch. Anders, mit Beitr. von W. Haas, I. Hove und P. Wiesinger. Berlin/New York: De Gruyter.
- DUDEN 2015 (DU-15):** *DUDEN. Das Aussprachewörterbuch* (= DUDEN Band 6). 2015. Bearb. von St. Kleiner und R. Knöbl in Zusammenarbeit mit der Dudenredaktion. 7., komplett überarbeitete und aktualisierte Auflage. Berlin: Dudenverlag.
- DUDEN 2015 (DU-15) digital:** Duden-Bibliothek Version 6.4.0.0. *DUDEN. Das Aussprachewörterbuch*. 7., komplett überarbeitete und aktualisierte Auflage. Berlin: Dudenverlag.

II. Selbstständige Beiträge im Aussprachewörterbuch

- Haas, Walter und Ingrid Hove.** 2009. Die Standardaussprache in der deutschsprachigen Schweiz. Krech, E.-M. u. a.: 2009. *Deutsches Aussprachewörterbuch*: 259-277. Berlin/New York: De Gruyter.
- Wiesinger, Peter.** 2009. Die Standardaussprache in Österreich. Krech, E.-M. u. a.: 2009. *Deutsches Aussprachewörterbuch*: 229-258. Berlin/New York: De Gruyter.

III. Sekundärliteratur

- Ammon, Ulrich, Hans Bickel und Alexandra N. Lenz (Hrsg.).** 2016. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen*. (2. Aufl.). Berlin/Boston: De Gruyter.
- Back, Otto.** 1995. Überlegungen zu einer österreichischen Standardlautung des Deutschen. Muhr, Rudolf, Richard Schrodtt und Peter Wiesinger (Hrsg.). 1995. *Österreichisches Deutsch: linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen*: 282-288. Wien: Holder-Pichler-Tempsky.

- Berend, Nina.** 2004. Regionale Gebrauchsstandards — Gibt es sie und wie kann man sie beschreiben? Eichinger, Ludwig M. und Werner Kallmeyer (Hrsg.). 2004. *Standardvariation: Wie viel Variation verträgt die deutsche Sprache?*: 143-170. Berlin/New York: De Gruyter.
- Ehrlich, Karoline.** 2009. *Die Aussprache des österreichischen Standarddeutsch — umfassende Sprech- und Sprachstandserhebung der österreichischen Orthoepie* (Diss.). Wien: Universität Wien.
- Jones, Daniel.** 2003. *English Pronouncing Dictionary*. (16. Aufl.). Roach, P., J. Hartman und J. Setter (Hrsg.). Cambridge: Cambridge University Press.
- Lameli, Alfred.** 2005. Standard und Regionalsprache — Konstanz und Wandel. Eggers, Eckhard, Jürgen Erich Schmidt und Dieter Stellmacher (Hrsg.). 2005. *Moderne Dialekte, neue Dialektologie. Akten des 1. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGGD) am Forschungsinstitut für deutsche Sprache «Deutscher Sprachatlas» der Philipps-Universität Marburg vom 5.-8. März 2003*: 495-513. Stuttgart: Steiner.
- Nycz, Krzysztof und Zygmunt Tęcza.** 2020a. On the Pronunciation Dictionaries of Contemporary German: Lexicographic Construction and Multimedia Components. *International Journal of Lexicography* 33(1): 73-89. URL: <https://academic.oup.com/ijl/article-abstract/33/1/73/5580320?redirectedFrom=fulltext>.
- Nycz, Krzysztof und Zygmunt Tęcza.** 2020b. On the Pronunciation Dictionaries of Contemporary German: The Concepts of Phonetic Standard and Differences in Specific Phonetic Issues. *International Journal of Lexicography* 33(4): 463-487. URL: <https://academic.oup.com/ijl/article-abstract/33/4/463/5682462?redirectedFrom=fulltext>.
- Tęcza, Zygmunt.** 2018. Die /r/-Allophone in der modernen deutschen Phonolexikographie. *Lublin Studies in Modern Languages and Literature* 42(1): 103-123. URL: <https://journals.umcs.pl/lsmll/article/downloadSuppFile/6037/2318>.
- Tęcza, Zygmunt und Krzysztof Nycz.** 2016. Deutsche Aussprachewörterbücher im Vergleich: Die Auffassung des phonetischen Standards und Divergenzen in phonetischen Einzelfragen. *Germanica Wratislaviensia* 141: 383-398. URL: <https://wuwr.pl/gwr/article/view/2289>.
- Upton, Clive, William A. Kretzschmar und Rafal Konopka.** 2003. *Oxford Dictionary of Pronunciation for Current English*. Oxford/New York: Oxford University Press.
- Wells, John C.** 2008. *Longman Pronunciation Dictionary*. (3. Aufl.). Harlow, Essex: Pearson.

Frequency or Keyness?

Zorica Đurović, *Faculty of Maritime Studies Kotor,
University of Montenegro, Kotor, Montenegro*
(zoricag@ucg.ac.me)

Abstract: The possibility of compiling electronic corpora, as of the second half of the last century, has provided new opportunities for vocabulary research. This has also resulted in the development of a series of computer software solutions for the lexical analysis of texts and the building of vocabulary lists for language learners. In this article, the differences in building technical vocabulary lists according to their frequency and keyness in corpora of English for Specific Purposes (ESP) are discussed. Both criteria have been discussed in terms of their benefits and limitations, and the possibilities of the most convenient combination of both. Finally, the word frequency list has been upgraded with keywords to provide a more comprehensive, but still very attainable, word list suitable for building a bilingual glossary or to be extended into a dictionary.

Keywords: WORD LIST, FREQUENCY, KEYNESS, KEYWORDS, CORPUS, MARINE ENGINEERING, ENGLISH, LEXIS, VOCABULARY

Opsomming: Frekwensie of sleutelstatus? Die moontlikheid om elektroniese korpora sedert die tweede helfte van die laaste eeu saam te stel, het nuwe geleenthede vir woordeskatnavorsing geskep. Dit het ook gelei tot die ontwikkeling van 'n reeks rekenaarsagteware-oplossings vir die leksikale ontleding van tekste en die saamstel van woordeskatlyste vir taalaanleerders. In hierdie artikel word die verskille in die samestelling van tegniese woordeskatlyste volgens hul frekwensie en sleutelstatus in korpora van Engels vir Spesifieke Doelwitte (ESD) bespreek. Albei kriteria word in terme van hul voor- en nadele bespreek, asook die moontlikhede van die gerieflikste kombinasie van beide. Laastens is die woordfrekwensielys aangevul met sleutelwoorde om 'n omvattender, maar steeds heel haalbare woordelys te verskaf wat geskik is om 'n tweetalige glossarium saam te stel of wat uitgebrei kan word tot 'n woordeboek.

Sleutelwoorde: WOORDELYS, FREKWENSIE, SLEUTELSTATUS, SLEUTELWOORDE, KORPUS, SKEEPSINGENIEURSWESE, ENGELS, LEKSIS, WOORDESKAT

1. Introduction

The possibility of creating and analysing electronic corpora has provided course designers and lexicographers with new opportunities in designing vocabulary (teaching and learning) material. The first corpus-based dictionaries were General English (GE) ones, such as *Monolingual Learner's Dictionaries* and *Collins COBUILD English Language Dictionary*, providing for a justified selection of words and original corpus-based examples (cf. Hanks 2020). Today, corpora and corpus-based tools are considered almost a conventional approach

to building lexicographic materials (Sinclair 1992; Abdelzaher 2022). Therefore, it comes as a surprise that it has generally not been adequately acknowledged and precisely defined by technical dictionaries and glossaries, especially as the vocabulary volume in technical and scientific texts is not as large as in GE texts, thus having higher frequency (density) of core vocabulary (Chung and Nation 2004; Kovalev et al. 2019; Kruse and Heid 2021). This may be the case because of frequently and ad hoc developed technical (often bilingual) specialized glossaries (e.g. of some medical, business or nautical terms) which are often not compiled by language professionals, thus not receiving significant attention from lexicographers (cf. Nkomo and Madiba 2011). The aim of this article is therefore to tackle two possible methods of computer-based headwords selection for a technical English, to contrast them, but also to combine the advantages of both.

The first method taken into account is a frequency count, used to identify the most frequent target vocabulary and, in that way, build a frequency-based word list. Frequency has been a primary criterion for headword selection, initially with GE dictionaries, glossaries and word lists. Following the needs of language learners who are at the same time professionals in technical areas, software tools have been developed for producing frequency vocabulary lists, starting with GE ones, but also for more and more of those related to specialized areas and pertaining to specific professional corpora.

Keyness, on the other hand, aims to detect the key vocabulary for a specific area by comparing its vocabulary frequencies with those in a reference GE corpus. The two methods are tested and discussed with reference to a professional corpus of marine engineering instruction books, with English for Marine Engineering Purposes (EMEP) generally considered extremely demanding, vocabulary-wise (Hsu 2014; Đurović et al. 2021). Adding to this the globality of the seafaring profession and the fact that English is the official means of communication of this complex discourse community, as formally established after World War II, technical vocabulary has been a mandatory requirement, but also a major challenge, for non-native speakers of English, as well as for language instructors. Specifically, this article stems from previous research on a word frequency list for marine engineering purposes. Thus, this study is a continuation with a general overview of the previous research and findings and, beyond that, we are presenting further investigation regarding applicable methodology and the possibilities of improvement when it comes to building effective ESP word lists.

2. Theoretical background and previous research

Analysing the most practical language needs of the target language learners — in this case, future and active marine engineers attending English for Marine Engineering Purposes courses — we embarked on the ambitious project of seeking the most effective and practical vocabulary tool(s) for technical lan-

guage learning, but also for the overall marine engineering profession. Furthermore, software solutions were tested that could assist in determining a practical and successful methodology.

The project started with the collection and selection of the most technical and professional marine engineering corpus, i.e. ship instruction books. In targeting these research objectives, expert advice and extensive teaching experience in the area was followed, but, even more importantly, the official requirements and recommendations set out by the International Maritime Organization (IMO) as the global standard-setting authority for the international shipping industry was also adhered to. The IMO's International Convention on the Standards of Training, Certification and Watchkeeping for Seafarers (STCW, Part 2.2) and the IMO Model Course 3.17 — Maritime English, notably the part on Specialised Maritime English dedicated to marine engineering courses of English, were particularly used as guidance. Apart from general communication skills in terms of using internal communication systems, the majority of the language skills requirements (about 90% of the anticipated course and self-study hours) are dedicated to "adequate knowledge of the English language to use engineering publications" (IMO Model Course 3.17 2015: 153). Guided by these clear instructions, the area of interest has been reading comprehension of marine engineering publications, specifically instruction books (Đurović 2021).

Following previous research findings and recommendations, the lexical profiling methodology and some of the most up-to-date software for the creation of a specialised marine engineering word list was applied (Đurović 2021). A frequency count was the starting point, which anticipates those words that appear most frequently across the corpora. Keyness, as a corpus linguistics method, on the other hand, refers to the frequency of the words in a special-purpose corpus compared to their frequency in a reference GE corpus. In addition to numerous other authors dealing with similar methodologies for building word lists (abundantly referred to in e.g. Archer et al. 2016 and Nation 2016), this article directly relies and builds upon the author's previous research with the same professional corpus, briefly presented below with reference to the marine engineering word frequency list.

2.1 Corpus

Following certain expert advice and experiences, primarily those of Chief Engineers, we sought to create a relevant selection of instructional engineering material of the utmost practical importance to marine engineers, ranging from familiarisation with a ship's systems and machinery, to regular maintenance, repairs and overhauling. The selection comprised technical manuals (most frequently referred to as instruction books) from a container ship, a tanker ship, a cruise ship and an offshore vessel. Additional material was added to enhance the diversity and bring the technologies up to date. The final corpus material comprises thousands of pages of electronic material related to ship machinery,

devices and systems, converted, additionally "cleaned" and prepared (Nation 2016: 224) to accommodate the software requirements. The Corpus of Ship Instruction Books (CSIB) was, in this way, finalised with 1,769,821 running words (tokens). Bearing in mind the composition and size, we may say that our corpus is of representative importance to the discipline-specific genre so as to guarantee the validity of the results and conclusions produced (Đurović 2021). Further details on the corpus can be found in the author's referenced research article.

2.2 Word frequency lists

Modern research into "specialized or technical vocabulary has focused primarily on producing a word list of technical vocabulary in professional fields of expertise in English for Specific Purposes" (Coxhead and Demecheleer 2018; Đurović 2021). As both native speakers and language learners tend to acquire vocabulary according to its frequency, both in general language and in specialized areas of their interest or (business) activity (Nation 2006), frequency has been the main criterion for extracting core vocabulary.

The first corpus-based frequency word lists were, naturally, GE word lists (e.g. Fries and Traver 1950; West 1953). Since then, depending on the specific needs of the (English) language learners and non-native speaking language users, computer tools and methods have been developed to build specialized or technical word lists. Owing to the availability of electronic GE corpora and modern software possibilities, there has been a growing number of technical word lists aimed at early specialisation in the target professional and technical areas. The main presumption in the process is that the language learners have mastered at least 2,000–3,000 GE words, which are considered the most frequent GE words and expected to cover about 80% of texts (Nation 2006; Dang and Webb 2016; Van Zeeland and Schmitt 2013; Web and Rodgers 2009). Therefore, the designated software solutions, such as RANGE (Heatley et al. 2002) and, more recently, AntWordProfiler (<https://www.laurenceanthony.net/software/antwordprofiler/>), provide the possibility of eliminating the assigned word lists from being further counted. In the case of building a technical word list, the eliminated GE word lists would be those containing the 2,000–3,000 most frequent GE words (as those anticipated to have been already mastered). In determining the size of the list and the cut-off point in terms of the frequency count, the main purpose and evaluation criterion of the produced word lists is to reach the level of 95% (as adequate) and/or 98% (as the ideal threshold) of the text coverage, together with the 2,000–3,000 most frequent GE words (Dung and Web 2016; Nation 2016; Laufer 1992; Đurović 2021).

Although the notions of frequency and word lists can refer to other criteria, such as keyness, we will henceforth refer to word frequency lists as those built upon the frequency count only (usually accompanied by a determined cut-off point), and keyword lists will be the ones built upon the keyness metrics.

2.3 Keyness and keywords

The lexical units of a language are generally considered equal in status, but, when it comes to text, their significance and role vary (Bondi 2010). The new interest in "words" has been gaining in importance when it comes to lexical analysis of texts (genres) and related corpus linguistics areas of research. Generally, the notion of keywords has not been defined by official linguistics; keywords have rather earned by themselves their rising importance and attraction, especially in Englishes for Specific Purposes. First of all, authors, more or less intentionally or spontaneously, still use both written forms: *key word* as a collocation and/or *keyword* as a compound, as is the case in this article. Also, keywords have often been used as general, and more or less provisional, markers of "aboutness" and of the style of a text, e.g. papers and articles (Scott and Tribble 2006: 59). However, not many have been aware of the new possibilities of eliciting a statistically justified ranking of keywords, which is now available by using contemporary software solutions.

Unlike frequency counts, the keyness of a word does not necessarily anticipate a high, but rather an unusual, frequency of that word as compared to the general language — in our case General English. They are "key" because they capture the essence of particular types of discourses (Culpeper and Demmen 2015: 1). Their importance also signifies the cultural importance of lexical items (Li and Tarp 2022), as they relate in "culturally significant ways", and would "provide a representation of socially important concepts" (Gabrielatos 2018). The statistical software possibilities provide us with an insight into these particular words of "special status" (Stubbs 2010: 21). Based on their significantly increased frequency as compared to referent types of general texts, keywords point to the very nature of the text, i.e. the genre itself, and enable its easier comprehension (Baker 2004; Gabrielatos 2018). Keyness is one of the highlights in professional corpora such as, specifically, maritime genres. Generally, when there is a corpus of a very demanding lexical load — such as marine engineering publications — this demands special attention as regards adequate comprehension and mastering of specialised vocabulary. Therefore, our intention was to test both metrics criteria (frequency count and keyness) for eliciting specialised vocabulary to be focused on during language courses and professional work.

3. Methodology

The relevant methodology for a frequency count was based upon the use of the AntWordProfiler 1.4.0w software (<https://www.laurenceanthony.net/software/antwordprofiler/>) which is an upgraded version of the previously used RANGE programme (Heatley et al. 2002). Since the focus here is a highly technical branch of ESP, the general tendency and recommendation of e.g. Coxhead (2000), Hsu (2014), Yang (2015), Nation (2016), Kwary and Artha (2017) and Vuković Stamatović (2020), was followed to upgrade the first 2,000 or 3,000 GE words with specialised vocabulary lists, which together aim to reach the adequate

reading comprehension threshold in the most efficient way. The referent General English (GE) word lists used for the process were Nation's word lists produced from the British National Corpus and Corpus of Contemporary American English (BNC/COCA, <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists>). The total of 25 lists contain about 1,000 word families each and, for this kind of research, they are usually accompanied by additional lists of the most frequent proper names, abbreviations, transparent compounds and marginal words developed by the same author (Nation 2004; 2006).

For formatting the lists into headwords only, lemmas, or expanding them into an all-family-members form, we used the Familizer + Lemmatizer program (<https://www.lex tutor.ca/familizer/>). For the preparation of the corpus and converting it into "plain text" format, we used AntFileConverter (<https://www.laurenceanthony.net/software/antfileconverter/>). The detailed methodology and procedure for building a specialised word list (in this case a marine engineering one), as well as the final results, were meticulously presented in previous research (cf. Đurović 2021). Therefore, the specifics of the methodology and results will be briefly summarised here as is needed and as is relevant.

Considering the keyness metrics, this will be carried out using the AntConc software, version 3.5.8. (<https://www.laurenceanthony.net/software/antconc/>). Although there are more and more online programs offered for the selection of words, the methodology of frequency count lists and recognised scientific research based upon it was followed, including that of Laurence Anthony as the software designer. His software solutions were adopted since they are complementary and readily available. The advantage of these programs is that they are free-of-charge, they have been regularly updated, they provide comparable results, and can be used for any language.

Investigating the keyness of words is an important research method of corpus linguistics where comparative analysis is conducted between corpora, i.e. our target corpus and the reference one. As the comparison is usually done with a GE corpus, the FLOB corpus of contemporary British English was used. FLOB was created as a contemporary version of the Lancaster-Oslo/Bergen corpus (LOB) from 1961 and contains about a million words from various British genres. With the aim of creating an up-to-date British English corpus methodologically, resembling the Brown University Standard Corpus of Present-Day American English (Kučera and Francis 1967) in size, the Freiburg version of the LOB corpus was published by the University of Albert-Ludwig in Freiburg in 1999 under the acronym FLOB. Considering its up-to-dateness and the size being close to the target corpus, this GE corpus was opted for as the reference one. Comparing sufficiently large corpora of a similar size is convenient for ensuring similar frequency opportunities, thus providing for comparability of the results (Nation 2016).

The target methodology tested here is actually a "hybrid" one, aiming to combine the benefits of both ones mentioned above — the former based upon frequency and the latter focused on keyness. The intention is to provide a more comprehensive and effective word list that could still be attainable and practi-

cal for ESP (English for Marine Engineering Purposes — EMEP) classes and courses. The hope is to provide a solid recommendation for combined corpus linguistics methods applicable to other ESP areas and cases.

4. Marine engineering word frequency list

Aiming to provide our target learners of English for Marine Engineering Purposes (EMEP) with a practical vocabulary tool to help them reach an adequate reading comprehension text coverage of 95%, the methodology recommended by recognised authors from the area (e.g. those summarised in Nation 2016) were followed and applied for the purpose of comparison, evaluation and recognition. All the necessary decisions and interventions made on the way, as well as the specifics, limitations and further possibilities, were meticulously presented as a part of previous research (cf. Đurović 2021). Finally, a marine engineering word list of 337 word families was developed, accompanied by a list of 73 transparent compounds, which were derived from the corpus of marine engineering instruction books consisting of 1,769,821 running words. For practical reasons, the list is not provided in the addendum as it is readily available in the previous research cited. Nevertheless, it makes up an integral part of the final glossary list given in Addendum 3.

Since the produced word lists are evaluated by the adequate level of (cumulative) coverage in the target corpus, we are here briefly referring to the evaluation of our marine engineering word list (Table 1). Considering the specifics of the corpus and possibilities of extension, we refer to it here as the Word List of Ship Instruction Books (WLSIB).

Table 1: Coverage of Word List of Ship Instruction Books (WLSIB) in the corpus of marine engineering instruction books (Đurović 2021)

Word lists	Tokens	Coverage (%)
BNC/COCA 3,000 + proper nouns, abbreviations and marginal words	1,547,067	87.41
Transparent compounds	12,783	0.72
WLSIB without compounds	130,994	7.41
Outside of the lists	78,977	4.46
Total	1,769,821	100

In total, together with the first 3,000 GE words, proper nouns, abbreviations and marginal words, the level of 95.54% (87.41% + 0.72% + 7.41%) is reached, thus reaching the goal of the adequate reading comprehension threshold, as recommended by Laufer (1989) and supported by e.g. Laufer and Ravenhorst-

Kalovski (2010) and Van Zeeland and Schmitt (2013). Taking into consideration that the desired level of coverage can be attained with no fewer than 12,000 general English words (only), as tested in Đurović et al. (2021), the final results perfectly fit the findings of Laufer and Ravenhorst-Kalovski (2010), by which the threshold of 95% is expected to be reached through the use of 4,000–5,000 word families (Đurović 2021).

Nevertheless, as human intervention and expertise are required and are indispensable throughout the process, we would not readily exclude the other valid criteria for vocabulary selection. We generally wanted to explore both criteria, compare them and possibly combine them to obtain more comprehensive results which would still be an attainable task both for students and trainees in marine engineering.

5. Keyword list

In reaching the positive evaluation of the list and the desired 95% coverage, we were further inspired by the possibilities of corpus linguistics in the selection of the most useful and most effective vocabulary for our target group of language learners. In the case of a specific technical corpus and genre of marine technical manuals, first we wanted to explore the keywords and the range of their keyness in terms of frequency when compared to General English (the FLOB corpus). For illustrative purposes, in Figure 1 the first 20 words are presented according to their keyness, i.e. the frequency ranking, resulting from comparison to their frequency in the reference GE corpus (FLOB) by means of the keyword metric.

Rank	Frequency	Keyness	Word
1	13977	+13629.13	oil
2	10887	+11146.53	valve
3	144202	+8269.42	the
4	10280	+8233.94	water
5	7642	+7712.58	pump
6	8635	+7597.54	pressure
7	6471	+6361.53	engine
8	7548	+6061.34	air
9	5990	+5791.6	fuel
10	5996	+5603.31	check
11	5685	+4953.51	operation
12	6859	+4866.11	system
13	6279	+4790.83	control
14	3723	+3611.38	manual
15	3779	+3513.01	bearing
16	3449	+3492.11	cylinder
17	3728	+3461.48	operating
18	3996	+3320.54	temperature
19	3719	+3183.09	step
20	3522	+3139.43	ring

Figure 1: Frequency and keyness of keywords in CSIB as compared to the reference FLOB corpus

From the table and numerical presentation of the results above it is clear that the frequency and keyness of the vocabulary are similar at the beginning, with keyness, as expected, having a far more rapid decrease in values than the frequency. Also, as expected, the keywords reflect the extremely technical character of marine engineering. Here it must also be noted that the AntConc programme presents results in the form of word types (not word families), as seen in the example of *operation* and *operating*, which are here given as separate units. This is another notion that should be borne in mind when combining methodologies and comparing the results.

In order to overcome the limitations of the AntConc programme in terms of the (im)possibility of using various word lists in the analysis, the initial keyword list obtained from ship instruction books (2,437 word types/1,172,171 tokens) were subjected to further analysis through the AntWordProfiler software. The convenience of the AntWordProfiler comes from its option to eliminate designated words/word lists from further counting and analysis. In that way the first 3,000 GE (BNC/COCA) words could be eliminated, as well as the lists of the most frequent proper names, abbreviations and marginal words (Nation 2004; 2006). Additionally, in order to obtain vocabulary that would be distinctive in relation to the obtained frequency list of marine engineering vocabulary, one of the lists assigned to the AntConc software was also our initially produced frequency list (WLSIB), including the obtained list of the most frequent transparent compounds (Đurović 2021). The newly obtained (additional) keyword list, accompanied by the list of key transparent compounds, was further analysed and "purged" of abbreviations and typos, converted into word families and supplemented by "unclassified" words, i.e. those not recognised by the programs, thus not automatically classed into family or lemmatised categories (e.g. *arrester*, *retighten*, *feedwater*). Furthermore, the previously obtained frequency list (WLSIB) was also supplemented with additional members of WLSIB word families that had been detected in the new keyword list, and the same was done with the three GE word lists and the lists of the most frequent proper names, marginal words and abbreviations. In particular, the initial WLSIB was supplemented by some words that excelled in terms of their keyness but that were "missed" by the related word families in the initial word frequency list, such as *actuation*, *igniter* and *emulsify*. Finally, a list of 124 marine engineering keywords (Addendum 1) and an addition of 43 key compounds (Addendum 2) was obtained. Considering the size of the list, we opted to keep it in full.

Owing to the results newly obtained through the application of this combined method seeking the benefits of both software solutions, a joint list was created that can serve as a glossary of marine technical manuals. In addition, following practical procedures generally favoured by engineers, all the words were placed into an integrated list arranged alphabetically (Addendum 3). The units presented are word families (again, for practical reasons of presentation) although lemmas are preferred and recommended when it comes to glossaries,

especially dictionaries. This would provide for a separate presentation of word types within a word family, e.g. *alter*, *alternate*, *alternator*. In particular, the expanded glossary word list initially comprised 1,500 units, but, as is usually the case with word list presentation, they are condensed into a word family list (Addendum 3). This can be further expanded to lemmas or all family members with adequate programmes, such as Familizer + Lemmatizer (Cobb 2018), as used here.

Finally, by integrating the frequency and keyness lists obtained from the Corpus of Ship Instruction Books, a total list of 577 words was created.

Table 2: The frequency and keyness word lists from ship instruction books

Word lists	Number of word families
WLSIB	337
Frequency list of transparent compounds	73
List of keywords	124
List of key transparent compounds	43
Total	577

6. Pedagogical and lexicographical implications

Taking into consideration that the total number of word families obtained through both criteria — frequency and keyness — including transparent compounds, is "only" 577 (Table 2), i.e. still below 1,000 (Nation 2004) or below 800, as is deemed a realistically attainable task for a language learning period of two years (Dang and Web 2016: 174), a glossary obtained this way could have a very practical application in ESP classes and courses, especially throughout one's professional career. Another reason for adding keywords would be that keywords would further reflect the style and specificity of the genre of ship instruction books and manuals. Therefore, both criteria should be considered and included in the optimisation of the produced technical vocabulary tool. In this way, there can be provided for the inclusion of all the keywords, i.e. the words that are the most specific ones for the marine engineering genre in comparison with the GE genres.

A glossary based upon such a word list can be monolingual or bilingual. Considering the globality of the seafaring profession and English as its lingua franca, bilingual variations would be the most useful and practical ones for marine engineers. Another advantage would be that, once formed in English, the glossary can be used in combination with any other language.

Furthermore, and if needed, the glossary can also be expanded by the first 3,000 BNC/COCA (GE) words, thus comprising the total vocabulary required

for adequate reading comprehension of marine engineering technical manuals. In addition, it can also be supplemented with lower-frequency technical words by lowering the initial cut-off point of 50 (Đurović 2021) in order to obtain an expanded glossary or dictionary foundation.

7. Discussion and limitations

In order to test the significance and validity of the integrated vocabulary list, i.e. the glossary of ship instruction manuals, its coverage was tested in our corpus in the same way we did for our primary WLSIB word list.

Table 3: Coverage of the glossary in the corpus of ship instruction books

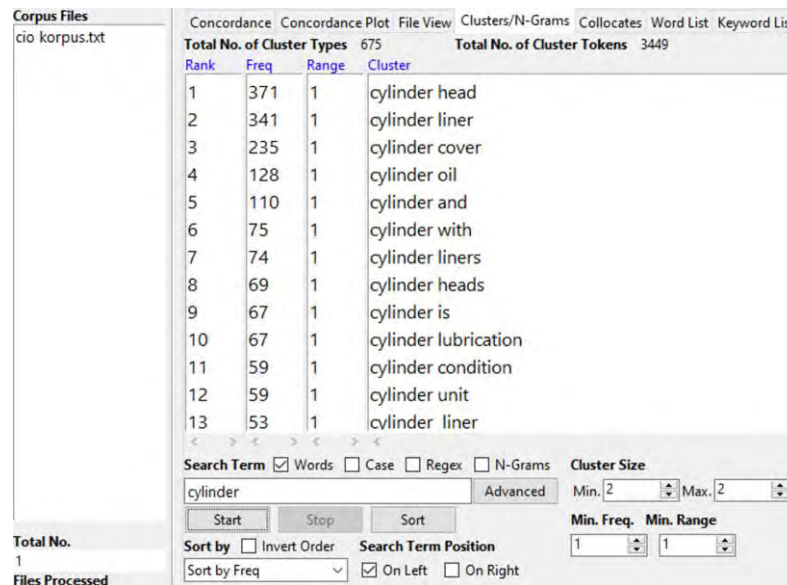
Word lists	Tokens	Coverage (%)
BNC/COCA 3,000 + proper names, abbreviations and marginal words	1,547,071	87.41
Glossary list (with transparent compounds)	151,135	8.54
Outside of the lists	71,615	4.05
Total	1,769,821	100

Based on the results presented in Table 3, a somewhat higher coverage of the glossary list (8.54%) was noticed as compared to the initial list with transparent compounds (8.13%). The total coverage of the glossary list with compounds is 95.95% (87.41% + 8.54%), which exceeds the coverage of WLSIB with transparent compounds by 0.41% (or about 7,500 corpus words) (see Table 1). It was expected not to make a drastic difference, based upon earlier research findings and conclusions related to the length of the word lists.

From the results, we can also confirm an earlier determined regularity whereby an additional extension to the list, i.e. inclusion of additional words with decreasing frequencies, is also accompanied by a rapid decrease in additional coverage in the corpus (Dang et al. 2017; Coxhead 2018; Nation 2016; Zipf 1935; 1949). Considering that in our specific case the difference does not significantly affect the final results, the initial methodology can be supplemented with this "hybrid" model. It would include additional key vocabulary as compared to the reference general one and would therewith enrich the frequency list and upgrade it to a more comprehensive and effective one.

In addition, we are well aware that words do not hold standalone meaning, but acquire their meaningfulness through combinations with other words. Therefore, collocations, n-grams and similar word combinations have been a recurrent topic of interest for ESP learners and instructors (Chen 2022). For this

purpose, the AntConc software can further be used for examining word relations such as collocations (and/or n-grams), which can be of use for additional development of glossary and dictionary items (Figure 2).



The screenshot shows the AntConc software interface. On the left, under 'Corpus Files', 'cio korpus.txt' is listed. The main window displays a concordance search for the term 'cylinder'. The search results are shown in a table with columns: Rank, Freq, Range, and Cluster. The results list 13 clusters, all with a frequency of 1 and a range of 1. The clusters are: cylinder head, cylinder liner, cylinder cover, cylinder oil, cylinder and, cylinder with, cylinder liners, cylinder heads, cylinder is, cylinder lubrication, cylinder condition, cylinder unit, and cylinder liner. Below the table, there are search options: 'Search Term' (cylinder), 'Words' (checked), 'Case' (unchecked), 'Regex' (unchecked), 'N-Grams' (unchecked), 'Cluster Size' (Min: 2, Max: 2), 'Sort by' (Sort by Freq), 'Invert Order' (unchecked), 'Search Term Position' (On Left checked, On Right unchecked), and 'Advanced' (unchecked).

Rank	Freq	Range	Cluster
1	371	1	cylinder head
2	341	1	cylinder liner
3	235	1	cylinder cover
4	128	1	cylinder oil
5	110	1	cylinder and
6	75	1	cylinder with
7	74	1	cylinder liners
8	69	1	cylinder heads
9	67	1	cylinder is
10	67	1	cylinder lubrication
11	59	1	cylinder condition
12	59	1	cylinder unit
13	53	1	cylinder liner

Figure 2: AntConc presentation of collocation and n-grams search

With the goal of building an effective and comprehensive, but still practical, vocabulary list for the target language learners (future and active marine engineers), we tested, contrasted, and combined the two methodologies presented above. On the way, the recommendations and previous findings and experiences we followed, aiming to contribute to the final product, but also to possible future methodologies. As metrics tools for the purpose, two software solutions developed by Laurence Anthony were mainly used: AntWordProfiler and AntConc. Lexical profiling using the AntWordProfiler software provides us with accurate information on the lexical characteristics and load of the target corpus. This software surpasses the AntConc software (and numerous others) in one very important aspect. It provides the opportunity to eliminate the available lists from further analysis, but it also measures the coverage of each list, as well as their cumulative coverage in the corpora. The keyness method (AntConc), however, provides us with the most specific vocabulary for the particular type of text by counting not the simple frequency, but rather the unusual frequency as compared to General English. As we can see in the example of our corpus, it provided us with additional specific technical vocabulary (e.g. *alloy*, *funnel*, *plumbing*, etc.) which do not belong to the most frequent English vocabulary (Nation 2004) but would certainly come in handy for marine engineers who are

non-native English speakers. One more reason for expanding the initial frequency list is the fact that the additional keyword list is not too long and the joint glossary list totals 577 (head)words (Table 2, Addendum 3), which is considered an attainable task for ESP courses (Dang and Webb 2016). On the other hand, when building a keyword list, the AntConc software does not have the option to eliminate any other list members from the count, thus by itself it cannot serve to upgrade the existing word lists. This is the reason why, for example, the AntWordProfiler was used with the obtained keyword list so that the members of the first GE word lists and the WLSIB frequency lists could be eliminated, which provided additional vocabulary only (with no repetition or overlapping).

Still, however statistically accurate the count, the methodology "does not work" without human expertise and intervention. In the software processing, especially in the case of technical corpora, there are always some "unrecognised" words which are presented as unclassified. They also need to be focused on and added to a certain word family or word list or eliminated in the case of a typo or similar error. Additional attention should also be paid to various spelling options coming from different publications (e.g. *manoeuvre* vs. *maneuver*, *minimise* vs. *minimize*, etc.) and putting them into the same families with cumulative frequency.

Further analysis of the list(s), as more or less statistical products, would open up some new possibilities and questions, such as those of a semantic nature. The software solutions do not recognise polysemous or cryptotechnical words, as referred to by Fraser (2009), which we should especially have in mind in the case of building a bilingual glossary. Another phenomenon has also been confirmed here, and that is the fact that the most frequent content words are also the most polysemic ones (Ravin and Leacock 2000). These words can formally belong to the first 3,000 GE words, but gain new meanings in marine engineering, either individually or in collocations. As recommended by previous research findings, they were added to the GE word families, although special attention should be paid to them in language courses. This also relativises the statistical results, as their frequency is added to the GE words only.

There is always a possibility of including, at least partially, the most frequent GE words that the ESP list has been built upon. In the example of the target corpus, words such as *actuator*, *mess*, *skirt*, *pin* and similar have been added to the first 3,000 BNC/COCA word families, as suggested by the established methodology. However, in English for Marine Engineering, the terms relate to specific parts of a ship or propulsion machinery and can have various translations in different languages. This goes a step further with collocations. For example, *arm* and *rock* also belong to the most frequent GE words. However, *rocker arms* are very important parts of the valve opening/closing mechanisms that have different translations in different languages. Thus, here again, the importance of expert intervention and attentiveness must be emphasised, regardless of the detailed methodology and previous recommendations and findings.

As concerns the corpus selection, another highlight should be put forward here. When building a word list, it is always related and should refer to the specific corpus. In this case, it was the professional corpus of marine engineers. However, in other possible research that could be, for example, dedicated to English language learners undertaking marine engineering studies in English, the corpus could comprise marine engineering course books in English, or they could be an upgrade to the corpus of ship instruction books, which may, again, be of different compositions and sizes.

8. Conclusion

Aiming to provide the target language learners with a concentrated and specialised word list elicited from their professional corpus of marine engineering publications, we initially followed the established methodology for building a word list from ship instruction books and manuals. This specialised word list comprised 348 words with 75 transparent compounds. Tempted to try out other criteria, specifically the keyness of words in the texts, we explored the possibilities of combining the two methodologies, building on the benefits of one over the other, while at the same time overcoming the limitations of both. Ultimately, we finalised our recommended glossary list at a total of 577 words (word families), which is still sufficiently practical in size to be used in EMEP courses or in building effective bilingual glossaries for members of this challenging discourse community with English as a non-native language.

In building technical vocabulary lists, we have been able to attest that the AntWordProfiler software especially comes in handy since it provides us with the opportunity to exclude lists of the most frequent GE words (or any other type of word list) from further processing, thus focusing the frequency count on the technical vocabulary to be mastered in order to reach the adequate (reading) comprehension of professional genre(s). On the other hand, it can also be used (and we wholeheartedly recommend it) with the keywords list obtained through AntConc in order to obtain additional vocabulary of key significance to ESP learners and users. In this way, the methodologies build upon each other and, if attentively conducted, the resulting word list can serve as a good basis for a glossary that can be made bilingual in combination with any other language. We hope that the presented methodology and exemplar results can inspire other researchers and ESP teachers to use them either individually, or in combination, as presented here. Furthermore, the methodologies can also be further extended to build frequency and/or keyness dictionaries.

References

- Abdelzaher, E.M. 2022. An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years. *Lexikos* 32: 162-179. <https://doi.org/10.5788/32-1-1714>.
- Anthony, L. 2014. *AntWordProfiler* (Version 1.4.1.0) [Computer Software]. Tokyo: Waseda University. Available from <https://www.laurenceanthony.net/software/antwordprofiler>.

-
- Anthony, L.** 2017. *AntFileConverter* (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Anthony, L.** 2019. *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Archer, D. (Ed.).** 2016. *What's in a Word-list? Investigating Word Frequency and Keyword Extraction*. London: Routledge. <https://doi.org/10.4324/9781315547411>.
- Baker, P.** 2004. Querying Keywords: Questions of Difference, Frequency and Sense in Keywords Analysis. *Journal of English Linguistics* 32(4): 346-359.
- Bondi, M.** 2010. Perspectives on Keywords and Keyness. Bondi, M. and M. Scott (Eds.). 2010. *Keyness in Texts*: 1-18. Amsterdam/Philadelphia: John Benjamins.
- Chen, Y.** 2022. The Effect of Learning Conditions on Collocation Gains: A Case Study of Task-based Dictionary Use Instruction. *Lexikos* 32: 1-30. <https://doi.org/10.5788/32-1-1679>.
- Chung, T.M. and I.S.P. Nation.** 2004. Identifying Technical Vocabulary. *System* 32(2): 251-263. <https://doi.org/10.1016/j.system.2003.11.008>.
- Cobb, T.** 2018. From Corpus to CALL: The Use of Technology in Teaching and Learning Formulaic Language. Siyanova-Chanturia, A. and A. Pellicer-Sanchez (Eds.). 2018. *Understanding Formulaic Language: A Second Language Acquisition Perspective*: 192-211. New York: Taylor & Francis.
- 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/New York: Routledge.
- Coxhead, A. and M. Demecheleer.** 2018. Investigating the Technical Vocabulary of Plumbing. *English for Specific Purposes* 51: 84-97.
- Culpeper, J. and J. Demmen.** 2015. Keywords. Biber, D. and R. Reppen (Eds.). 2015. *The Cambridge Handbook of English Corpus Linguistics*: 90-105. Cambridge: Cambridge University Press.
- Dang, T.N.Y. and S. Webb.** 2016. Making an Essential Word List for Beginners. Nation, I.S.P. (Ed.). 2016. *Making and Using Word Lists for Language Learning and Testing*: 153-167. Amsterdam: John Benjamins.
- Dang, T.N.Y., A. Coxhead and S. Webb.** 2017. The Academic Spoken Word List. *Language Learning* 67(4): 959-997.
- 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. <https://doi.org/10.5788/31-1-1647>.
- Đurović, Z., Vuković Stamatović, M. and M. Vukičević.** 2021. How Much and What Kind of Vocabulary do Marine Engineers Need for Adequate Comprehension of Ship Instruction Books and Manuals? *Círculo de lingüística aplicada a la comunicación* 88: 123-133. <https://dx.doi.org/10.5209/clac.78300>.
- Fraser, S.** 2009. Breaking Down the Divisions Between General, Academic, and Technical Vocabulary: The Establishment of a Single, Discipline-based Word List for ESP Learners. *Hiroshima Studies in Language and Language Education* 12: 151-167.
- Fries, C.C. and A.A. Traver.** 1950. *English Word Lists*. Ann Arbor: George Wahr.
- Gabrielatos, C.** 2018. Keyness Analysis: Nature, Metrics and Techniques. Taylor, C. and A. Marchi (Eds.). 2018. *Corpus Approaches To Discourse: A Critical Review*: 225-258. Oxford: Routledge.
- Hanks, P.** 2020. English Dictionaries and Corpus Linguistics. Ogilvie, S. (Ed.). 2020. *The Cambridge Companion to English Dictionaries*: 219-239. Cambridge: Cambridge University Press.

-
- Heatley, A., I.S.P. Nation and A. Coxhead. 2002. Vocabulary Analysis Programs. The Range Program (online) <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>.
- Hsu, W. 2014. Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates. *English for Specific Purposes* 33: 54-65.
- IMO. 2015. *Maritime English*. Model Course 3.17. London: International Maritime Organization.
- IMO. 2017. *STCW Convention and STCW Code*. London: International Maritime Organization.
- Kovalev, I.V., S. Yu Piskorskaya, M.V. Karaseva and A.A. Voroshilova. 2019. Computer-aided Approach to Synthesis of the Frequency Dictionary on System Analysis in Electronic Machinery, Aviation and Space Industry. *IOP Conference Series: Materials Science and Engineering* 537(4): 1-7. <https://doi.org/10.1088/1757-899X/537/4/042085>.
- 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. Gavrilidou, 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.
- Kučera, H. and W.N. Francis. 1967. *Computational Analysis of Present-day American English*. Providence, RI: Brown University Press.
- Kwary, D.A. and A.F. Artha. 2017. The Academic Article Word List for Social Sciences. *MEXTESOL* 41(4): 1-11.
- Laufer, B. 1989. What Percentage of Text-Lexis is Essential for Comprehension? Lauren, C. and M. Nordman (Eds.). 1989. *Special Language: From Human Thinking to Thinking Machines*: 316-323. Clevedon: Multilingual Matters.
- Laufer, B. 1992. How Much Lexis Is Necessary for Reading Comprehension? Arnaud, P.J.L. and H. Bejoing (Eds.). 1992. *Vocabulary and Applied Linguistics*: 129-132. London: Macmillan.
- Laufer, B. and G.C. Ravenhorst-Kalovski. 2010. Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language* 22(1): 15-30.
- Li, Q. and S. Tarp. 2022. Differentiated Treatment of Cultural Items in Lexicographical Products: A Necessary Adaptation to the Digital Environment. *Lexikos* 32: 90-117. <https://doi.org/10.5788/32-1-1706>.
- Nation, I.S.P. 2004. A Study of the Most Frequent Word Families in the British National Corpus. Bogaards, P. and B. Laufer (Eds.). 2004. *Vocabulary in a Second Language: Selection, Acquisition and Testing*: 3-13. Amsterdam: John Benjamins.
- Nation, I.S.P. 2006. How Large a Vocabulary is Needed for Reading and Listening? *Canadian Modern Language Review* 63(1): 59-82.
- Nation, I.S.P. 2016. *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins.
- Nkomo, D. and M. Madiba. 2011. The Compilation of Multilingual Concept Literacy Glossaries at the University of Cape Town: A Lexicographical Function Theoretical Approach. *Lexikos* 21: 144-168. <https://doi.org/10.5788/21-1-41>.
- Ravin, Y. and C. Leacock (Eds.). 2000. *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press.
- Scott, M. and C. Tribble. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Philadelphia: John Benjamins.
- Sinclair, J. 1992. *Collins COBUILD English Language Dictionary*. London: HarperCollins.

- Stubbs, M.** 2010. Three Concepts of Keyness. Bondi, M. and M. Scott (Eds.). 2010. *Keyness in Texts*: 21-42. Amsterdam/Philadelphia: John Benjamins.
- Van Zeeland, H. and N. Schmitt.** 2013. Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics* 34(4): 457-479. <http://doi.org/10.1093/applin/ams074>.
- Vuković Stamatović, M.** 2020. Vocabulary Complexity and Reading and Listening Comprehension of Various Physics Genres. *Corpus Linguistics and Linguistic Theory* 16(3): 487-514. <https://doi.org/10.1515/cllt-2019-0022>.
- West, M.** 1953. *A General Service List of English Words*. London: Longman, Green & Co.
- Yang, M.-N.** 2015. A Nursing Academic Word List. *English for Specific Purposes* 37: 27-38. <https://doi.org/10.1016/j.esp.2014.05.003>.
- Zipf, G.K.** 1935. *The Psycho-biology of Language*. Boston: Houghton Mifflin.
- Zipf, G.K.** 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, Mass.: Addison-Wesley Press.

Addendum 1: List of additional keywords from ship instruction books (without transparent compounds)

acetylene	expel	pest
adhere	extern	plumbing
affixed	finned	portable
alloy	fountains	poultry
armature	funnel	prerequisites
arrester	galvanize	radiator
ascertain	garbage	readout
baffle	goggles	receptacle
beverage	grams	recess
breather	graphite	reciprocating
buffet	grease	rectify
bulb	hydrazine	refrigerant
burrs	hydroxide	reportable
buzzer	hysteresis	retention
chassis	illuminate	ridges
chock	increment	scuffing
circlip	ingress	scum
citric	inhalation	serum
clicking	insoluble	setter
coalescer	ion	shim
compensators	kerosene	sling
conformity	lamellar	slotted
contactor	lapping	slushing
coupler	lateral	sterilizer
deficiencies	locker	strap
descaling	lowing	swabs
dew	magnifying	swirler
diaper	malfunctions	synopsis
diarrhea	mandrel	tapered
diffuser	micron	tappet
dimensioned	modulating	Teflon
diode	molluscan	template
disengaged	mop	turbidity
disulfide	nitrite	unitor
dongle	notification	vanadium
dowel	ohm	vapor
duplex	opacity	Vaseline
elapsed	osmosis	vomit
elysator	pallet	wedges
emery	pantries	wobb
encoder	pentane	
erection	pertaining	

Addendum 2: List of additional key transparent compounds from ship instruction books

aftmost	framebox	startup
backpressure	gearbox	staybolts
backup	gearwheel	testbed
backwash	hereby	tubesheet
burnertype	inline	undercooked
carryout	logout	underside
checkbag	lowermost	undersize
convertbox	manhole	upwards
deckhead	microorganisms	usefor
download	overpressure	wastewater
downtime	pushbutton	website
feedwater	salinometer	workcard
flowmeter	shellstock	worktable
foodborne	shipyard	
foreword	startstop	

Addendum 3: Glossary list of headwords from ship instruction books

aboard	backup	changeover	crankpin
abrasive	backwash	chassis	crankshaft
accessory	baffle	checkbag	crankthrow
accord	barge	chlorine	crosshead
acetylene	barrel	chock	crosswise
acid	batch	circlip	cylinder
actuate	bedplate	citric	datalogger
acute	bellow	clamp	debris
adhere	beverage	classification	decant
adhesive	bilge	clicking	deckhead
adjacent	blade	clockwise	default
adsorb	blink	clog	defect
affixed	bolt	clutch	deficiencies
aft	bonnet	coalesce	deflect
aftmost	bracket	cock	deform
align	brass	coil	descaling
alkaline	breakdown	collar	detergent
alloy	breather	combustion	deteriorate
alternate	bronze	communicable	deviate
aluminum	buffer	compartment	dew
ambience	buffet	compatible	diagnosis
amplify	bulb	compensators	diagram
analog	bulkhead	comply	dial
annex	bunker	compress	diameter
annular	burnertype	con	diaper
anode	burrs	condense	diaphragm
anti	buzzer	cone	diarrhea
appendix	bypass	configure	diesel
appliance	cabin	conformity	differential
armature	calibrate	console	diffuser
arrester	calorific	contactor	digit
arrow	cam	contaminate	dimensioned
ascertain	camshaft	contouch	din
ash	carrieout	convertbox	diode
assemble	cartridge	copper	dip
astern	casing	copyright	dipstick
automate	caterpillar	corrode	discard
automobile	caution	countdown	discrete
auxiliary	cavity	coupler	disengaged
axis	centrifuge	crane	disinfect
backflow	centripetal	crank	dismantle
backpressure	certify	crankcase	dispense

displace	flue	galvanize	impulse
dissolve	fluid	garbage	incinerate
distillate	flush	gasket	increment
disulfide	flywheel	gastroenteritis	inert
dongle	foodborne	gastrointestinal	ingress
dowel	fore	gauge	inhalation
download	foreword	gearbox	inhibit
downstream	foul	gearwheel	inlet
downtime	fountains	generator	inline
drip	framebox	geometry	insoluble
droop	freshwater	gland	insulate
dual	friction	globe	intact
duct	funnel	glove	intake
duplex	furnace	goggles	integral
durable	fuse	grams	intercept
duration	galvanize	graphite	interface
dynamic	garbage	grease	interlock
effluent	gasket	grease	intermediate
eject	gastroenteritis	grind	interval
elapsed	gastrointestinal	groove	ion
electrode	gauge	gudgeon	jacked
elysator	gearbox	halogen	kerosene
emery	gearwheel	hammer	keyboard
emulsion	generator	handwashing	kit
enclose	geometry	handwheel	knob
encoder	gland	harness	lamellar
erection	globe	hereby	lance
erosion	glove	hexagon	lapping
evaporate	goggles	hoist	lateral
ex	grams	hood	layout
expel	flue	horizontal	lever
expire	fluid	hose	linear
extern	flush	hub	linen
eyebolts	flywheel	hull	liner
fasten	foodborne	humid	liter
fax	fore	hydraulic	locker
fecal	foreword	hydrazine	login
feedback	foul	hydroxide	logout
feedwater	fountains	hysteresis	loop
finned	framebox	icon	lowermost
fixture	freshwater	identical	lowing
flange	friction	idle	lube
flap	funnel	ignite	lubricate
flotation	furnace	illuminate	magnifying

flowmeter	fuse	impel	malfunctions
mandrel	overpressure	rack	setpoint
manhole	override	radial	setter
manifold	overspeed	radiator	setup
maneuver	overview	ram	shaft
manometer	oxidation	ramp	shellfish
membrane	oxygen	readout	shellstock
mesh	pallet	receptacle	shim
micro	pantries	recess	shipbuilding
micron	parameter	reciprocating	shipyard
microorganisms	particle	recreation	shutdown
millimeters	password	rectify	silicon
mineral	paste	refract	sketch
minimize	pentane	refrigerant	sleeve
mist	permissible	relay	sling
modulating	pertaining	reportable	slotted
moisture	pest	residue	sludge
molluscan	pinion	resilience	slushing
molybdenum	pipelines	retention	socket
mop	piston	ridges	sodium
mount	pliers	rim	soiled
multi	plumbing	rinse	solenoid
nameplate	plunge	rod	solvent
needle	pneumatic	rotate	soot
nipple	polyamide	rubber	sootblower
nitrite	polymer	rudder	spa
node	portable	rust	span
nominal	potable	saline	spanner
notification	potentiometer	salinometer	spark
nozzle	poultry	sanitize	specimen
offset	precaution	satisfactory	spindle
ohm	preface	sauer	spiral
onboard	preliminary	scavenge	splash
opacity	prerequisites	scrape	spool
optimum	prescribe	screwdriver	stack
opus	preset	scrubber	standby
orifice	prolong	scuffing	standstill
osmosis	propel	scum	startstop
outbreak	propulsion	seawater	startup
outlet	proximity	seizure	static
overboard	pulley	selfjector	staybolts
overflow	pulse	sensor	steer
overhaul	puncture	serial	sterilizer
overheating	purge	serum	stool

overlay	pushbutton	servo	strap
overload	quarantine	servomotor	stud
stuffed	thermometer	turbocharger	ventilate
suction	thermostat	tween	verify
sulphur	thread	undercooked	vertical
sump	threshold	underside	vibrate
surge	throttle	undersize	viscous
surveillance	throughput	unitor	volt
swabs	thrust	upstream	vomit
swirler	tiller	uptake	warewashing
synopsis	tilt	upward	warranty
synthetic	tolerance	upwards	wastewater
tab	torch	usage	website
tag	torque	usefor	wedges
tapered	torsion	utensil	weld
tappet	touchscreen	vacuum	whirlpool
Teflon	toxic	valve	wobble
telescope	transducer	vanadium	workcard
template	troubleshooting	vane	worktable
terminal	tubesheet	vapor	wrench
testbed	turbidity	Vaseline	yoke
thermal	turbine	velocity	

The Treatment of Argument Structure Constructions in Online English Learners' Dictionaries: The Case of *V N ADJ* ASCs

Yongfang Feng, *School of English Education, Guangdong University of Foreign Studies, Guangzhou, China*
(fengyf@gdufs.edu.cn)

and

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

Abstract: Despite their diversity and complexity in structure and meaning, little research has been conducted into the systematic presentation of argument structure constructions (ASCs) in English learners' dictionaries (ELDs). To fill in the gap, this paper focused on the treatment of *V N ADJ* ASCs in the "Big Five" online ELDs against usage data. First a list of 40 target verbs was obtained using two measures of contingency (i.e., faithfulness and collocation strength) through collocational analysis of *V N ADJ* ASCs in the British National Corpus, and then the related ASCs regarding their macro- and micro-structural presentation in selected dictionaries were examined. It was found that most ELDs attach importance to ASCs, notably LDOCE, which assigns many of them the status of phrases. Nevertheless, the treatment of ASCs in current ELDs is inadequate. First, the inclusion of ASCs in most ELDs is far from sufficient. Second, in some cases, there is a mismatch between a valency pattern illustration and an illustrative example. Last but not least, most ELDs are not consistent with respect to their policy of listing and encoding ASCs, and assigning them the status of phrases. We proposed, in the framework of the usage-based construction grammar, some suggestions for optimizing the treatment of ASCs.

Keywords: ARGUMENT STRUCTURE CONSTRUCTIONS, ONLINE ENGLISH LEARNERS' DICTIONARY, PHRASEOLOGY, COLLOCATIONAL ANALYSIS, USAGE-BASED CONSTRUCTION GRAMMAR

Opsomming: Die leksikografiese hantering van argumentstruktuurkonstruksies in aanlyn Engelse aanleerderswoordeboeke: Die geval van *V N ADJ* ASK's.

Ten spyte van hul diversiteit en kompleksiteit in struktuur en betekenis is daar nog min navorsing oor die sistematiese voorstelling van argumentstruktuurkonstruksies (ASK's) in Engelse aanleerderswoordeboeke (EAW'e) gedoen. Om hierdie gaping te vul, is daar in hierdie artikel gefokus op die hantering van *V N ADJ* ASK's in die "Groot Vyf" aanlyn EAW'e teenoor gebruiksdata. Eers is 'n lys van 40 doelwerkwoorde verkry deur twee gebeurlikheidswaardes (d.i. betroubaarheid en kollostruksionele sterkte) met behulp van kollostruksionele analise van *V N ADJ* ASK's in die Britse

Nasionale Korpus te gebruik, en daarna is die verwante ASK's ten opsigte van hul makro- en mikrostrukturele voorstelling in geselekteerde woordeboeke ontleed. Daar is bevind dat die meeste EAW'e ASK's belangrik ag, veral die LDOCE, wat frasestatus aan baie van die ASK's toeken. Die hantering van ASK's in huidige EAW'e is nogtans ontoereikend. Eerstens is die insluiting van ASK's in die meeste EAW'e glad nie voldoende nie. Tweedens is daar in sommige gevalle 'n wanverhouding tussen 'n valensiepatroonillustrasie en 'n illustratiewe voorbeeld. Laastens, maar nie die minste nie, is die meeste EAW'e nie konsekwent rakende hul beleid van opname en enkodering van ASK's nie asook nie ten opsigte van die toekenning van frasestatus aan hierdie ASK's nie. Ons doen binne die raamwerk van die gebruiksgebaseerde konstruksiegrammatika enkele voorstelle vir die optimalisering van die hantering van ASK's aan die hand.

Sleutelwoorde: ARGUMENTSTRUKTUURKONSTRUKSIES, AANLYN ENGELSE AANLEERDERSWOORDEBOEKE, FRASEOLOGIE, KOLLOSTRUKSIONELE ANALISE, GEBRUIKSGEBASEERDE KONSTRUKSIEGRAMMATIKA

1. Introduction

Traditionally, dictionaries are meaning-driven: they deal with lexicon, and focus on the different senses of individual words. But they reveal little about the phraseology that separates one sense from another. By comparison, grammars are structure-driven: they contain the productive rules for constructing utterances, and describe how words combine to form sentences (Hanks and Može 2019).

However, the advent of corpus linguistics since the 1980s has provided ample evidence that there is no clear dividing line between grammar and lexicon, and that they are merely different ends of the same continuum. The notion in traditional lexicography that meanings can be identified in individual words has been increasingly challenged. The viability of the traditional model of the dictionary as an ordered listing of individual words and senses has been undermined, pointing towards a radically different model where meanings are located through and within phraseology (Sinclair 1991; Kilgarriff 1997; Hunston and Francis 2000; Sinclair 2004; Hanks 2008; Hoey and O'Donnell 2008; Moon 2008; Herbst and Klotz 2009; Stubbs 2009; Hanks 2012, 2013; Hunston 2014; Rundell 2018; Hunston 2019; Hunston and Su 2019; Hanks and Ma 2020). The existence of phraseological dictionaries demonstrates that the distinction between lexicon and grammar is by no means as clear-cut as was often assumed (Herbst and Klotz 2009).

In the usage-based construction grammar, the boundary between lexicon and grammar is blurred. Constructions, like traditional words, are construed as conventional, learned pairings of form and meaning/function at different levels of schematicity. Language is a network of constructions, ranging from substantive word construction to highly schematic argument structure constructions (ASCs). The acquisition of language is the learning of constructions (Goldberg 1995, 2006; Hoffmann and Trousdale 2013; Goldberg 2019). These tenets of construction

grammar highlight the commonality between words and phrasal units, and make it viable to characterize all conventional constructional units in a similar way to the representation of lexical units in traditional lexicography.

All these insights have challenged the prevailing notion of headwords in traditional dictionaries with their linear organization, and motivated a transfer from the conventional focus on lexical semantics and morphology to a new trend towards phraseological units of meaning. Following this trend, some linguists and lexicographers have proposed that construction grammar can be applied to broaden the scope of phraseological description for practical lexicography (Przepiórkowski et al. 2017; Croft and Vigus 2017).

Similar to collocations, ASCs are conventionalized but unpredictable, concerning item-specific knowledge with respect to the co-occurrence of one word with a particular grammatical construction (Herbst et al. 2014). Recent research on ASCs has shown that it is implausible to focus on the semantics of verbs alone to explain the syntactic realization of verbs, and that both verbs and constructions contribute meanings and arguments to the whole construction (Goldberg 1995; Boas 2003; Goldberg et al. 2004; Goldberg and Jackendoff 2004; Goldberg 2006; Boas 2009; Faulhaber 2011; Boas 2014; Herbst et al. 2014; Perek 2015; Goldberg 2019). Verbs used in a construction share common constructional meanings and yet demonstrate idiosyncrasies at the same time. For instance, the syntactic frame *V N ADJ* can be the English resultative construction "[X MAKE [Y BECOME Z]]" (Goldberg 1995; Boas 2003; Goldberg and Jackendoff 2004), or the English attributive construction "[X THINK [Y BE Z]]" (Quirk et al. 1985; Hampe 2011).

However, a brief glimpse at *V N ADJ* ASCs in current English learners' dictionaries (ELDs) reveals that dictionary compilers tend to focus mainly on the semantics of verbs, unaware of the possible role of constructions in the argument realization of ASCs. Most ELDs only use labels like [+ adj], [~ sth + adj], or [*v n adj*] to indicate the resultative construction and attributive construction. No further device is used to differentiate the two constructions. As Goldberg and Jackendoff (2004: 563) point out, "having a phrase structure rule *V NP AP/PP* does not tell us enough about the resultative construction". In addition, the great variety and differences in the use of labels to indicate the valency patterns of *V N ADJ* ASCs in the ELDs imply the diversity and complexity of these constructions. Thus, ASCs deserve more attention from lexicographers. Furthermore, for some verb-specific constructions, the words used in each slot may demonstrate idiomatic preferences or restrictions, but current ELDs fail to convey such information. For example, in the OPINION sense of *hold*, words that can be used in the *ADJ* slot in the construction *HOLD N ADJ* are restricted to *accountable*, *liable*, and *responsible* (Francis et al. 1996: 280). However, *Oxford Advanced Learner's Dictionary* (10th ed.) (OALD10) only offers an example (viz., *Parents will be held responsible for their children's behaviour.*) after the pattern ~ **sb/sth adj**, and does not further indicate its restrictions in usage.

Compared with the abundant studies on the treatment of collocations in

dictionaries, little attention has been given to ASCs in lexicography. Though Cheng and Xu (2022) have illustrated how lexicographers can implement the idea of construction grammar to optimize the entry of the *way*-construction in learners' dictionaries, their focus is on a nominal ASC, and their investigation is mainly based on elicitation tasks rather than usage (viz., quantitative analysis of corpus evidence).

To fill in the gap, this study aims to investigate how online ELDs present one of the typical verb-based constructions — *V N ADJ* ASCs — within the framework of the usage-based construction grammar. Specifically, it explores to what extent the description of a language that a dictionary provides reflects the reliable empirical evidence regarding the way in which the language is actually used, and how to implement the basic tenets of the usage-based construction grammar into lexicographic practice.

2. Methodology

2.1 The "Big five" online ELDs

The online dictionaries analyzed in this study include five leading ELDs, known as the "Big Five" ELDs, namely *Oxford Advanced Learner's Dictionary* (OALD), *Longman Dictionary of Contemporary English* (LDOCE), *COBUILD Advanced English Dictionary* (COBUILD), *Cambridge Advanced Learner's Dictionary* (CALD), and *Macmillan English Dictionary* (MED). These dictionaries were chosen because they are widely acknowledged as the most popular ELDs of admirably high standards, and represent the most recent developments in pedagogical lexicography. Additionally, they are all corpus-based and phraseology-oriented, reflecting the tendency in current pedagogical lexicography towards a more explicit representation of formulaic language. The usage patterns found in them are believed to largely correspond to the most frequent lexical and syntactic paradigms in the British National Corpus (BNC). An investigation into the "Big Five" online ELDs helps to gauge precisely the significant and distinctive features of leading ELDs, thus lending more credence to future improvements in lexicography.

2.2 List of Target *V N ADJ* ASCs

2.2.1 Retrieval of instances of *V N ADJ* ASCs from the BNC

V N ADJ ASC, which is also called complex pattern or complex transitive complementation, refers to a clause pattern in which an object is followed by an adjective phrase as its object complement (Quirk et al. 1985; Francis et al. 1996). Given the bottom-up approach adopted in the COBUILD project and its emphasis and reliance on corpus evidence, the complex pattern with an adject-

tive phrase as its object complement (i.e., the *V N ADJ* pattern), and the 97 verbs identified in this pattern (Francis et al. 1996), provided an ideal starting point for our analysis. Since Francis et al. (1996) did not indicate how frequent each of the listed verb types is, our focus upon usage makes it necessary to reconstruct the data.

V N ADJ ASC consists of syntactic categories such as a nominal phrase and an adjective phrase. Such categories are very complicated and diverse, and some involve combinations of more than two words. Therefore, it is not easy to capture them through a simple corpus search with any degree of precision. Considering this, the BNC was chosen for the quantitative analysis because it is a balanced corpus with a size of 100 million words. We can obtain instances of *V N ADJ* ASCs in the BNC via Sketch Engine, whose distinctive feature is Word Sketch which can be used as a one-page summary of a word's syntactic and collocational behavior (Kilgariff et al. 2014; Thomas 2014). In the results of Word Sketches, *V N ADJ* ASCs can be identified by "adjectives after 'VERB' and noun". Figure 1 presents information about the adjective complements of *confess*.

adjectives after "confess" and noun		
	227	0.00
ignorant	15	6.54
guilty	47	6.05
unable	37	5.18
daily	12	2.93
first	14	0.77
due	18	0.40

Figure 1: Adjective complements of *confess* in the BNC

In accordance with previous studies (Boas 2003; González-García 2009; Hampe 2011), an expression will be counted as an instance of *V N ADJ* ASCs if the complement is (1) object-related, (2) non-deletable, and (3) not deletable without a change in the semantics of the VP. A close examination of the data in Figure 1 shows that *daily*, *first*, and *due* do not belong to *V N ADJ* ASCs, because these adjectives are adjuncts instead of object-complements (e.g., *I will <confess> my sins daily to God ...*).

Considering the noise in the results of the Word Sketch for the adjective complements of *confess* in the BNC, a pilot study was conducted to estimate the precision of Sketch Engine. First, we randomly sampled 10 out of the 97 verbs.

Second, each verb in the BNC was queried via Word Sketch in Sketch Engine, and all the concordances for each query were downloaded. Then, errors were manually discarded, and we conducted a precision analysis. Results show an average of 97.7% in terms of precision. This suggests that Sketch Engine is a reliable tool for this research. Note that precision¹ was prioritized over recall in the analysis, because the research aims to provide evidence for lexicographic presentation, in which typical instances of usage are provided. Therefore, we decided to use the data provided by Sketch Engine. We searched the 97 verbs one by one, downloaded all the instances of *V N ADJ* ASCs, detected errors manually, and identified their token frequency. It was found that 13² out of the 97 verbs do not occur in *V N ADJ* ASCs, and that the token frequency of *have* in *V N ADJ* ASCs is difficult to calculate. Thus, we discarded them from our analysis, and finally obtained a list of 83 verbs.

Thereafter the type and token frequencies of verbs that occupy *V N ADJ* ASCs were calculated. To determine whether they fall into a coherent meaning group, we measured their contingency of associations with *V N ADJ* ASCs and the semantic associations of these verbs. The words used in each slot in each verb-based *V N ADJ* ASC were retrieved as well.

2.2.2 Sampled verbs and verb-based *V N ADJ* ASCs

To make the analysis more reliable, we had to ensure that *V N ADJ* ASC is the typical and common usage pattern for each verb sampled. Following the instructions of collocation analysis (Stefanowitsch and Gries 2003)³, *R* scripts were used (Gries 2014) to calculate two measures of contingency — faithfulness and collocation strength — for the verb types occupying *V N ADJ* ASCs in our searches.

Faithfulness refers to the proportion of tokens of total verb usage that appear in a particular construction. It is the simplest measure of contingency. The high value of faithfulness implies the high proportion of a verb occurring with a particular construction. Collocation strength measures the association strength between a particular construction and the lexical elements filling certain slots in the construction. The higher the value of collocation strength is, the stronger the association between the verb and the construction is. By using both faithfulness and collocation strength, we can ensure that *V N ADJ* ASC is one of the most common and typical usages of the chosen verbs, thus deserving to be described in ELDs.

Therefore, based on the results of the collocation analysis, the top 30 verbs in order of faithfulness, and 10 verbs from the top 20 verbs in order of collocation strength were picked. In other words, the 40 verbs sampled are among either the top 30 verbs in order of faithfulness or the top 20 verbs in order of collocation strength. Table 1 lists the 40 verbs. It is worth noticing that some verbs (i.e., *hold*, *find*, *judge*, and *jerk*) belong to more than one group.

Table 1: Verbs sampled for investigating *V N ADJ* ASCs in the "Big Five"

Verb groups	Members
MAKE	<i>make, render, leave, keep, hold</i> <i>drive, knock, send, scare</i> <i>set, push, wrench, prise, slam, yank, jerk, clamp, lever</i> <i>shoot, paint, shake, stuff, wipe, cram, scrub, towel, jerk</i>
CONSIDER	<i>consider, think, find, hold, deem, prove, judge, rate, believe</i>
CALL	<i>call, declare, pronounce, presume, certify, judge, find, profess</i>

A close examination of these verb groups shows that the general constructional meanings of *V N ADJ* ASCs can be generated from the lexical semantics of the matrix verbs that belong to at least two relatively distinct semantic groups, viz. the "MAKE" group, and the "CONSIDER/CALL" group. The "MAKE" group consists of a broad range of verbs that can have an effect on someone or something when used in *V N ADJ* ASCs. Those verbs instantiate the resultative ASCs with the semantics "X CAUSES Y to BECOME Z" for the syntactic pattern $NP_x V NP_y AdjP_z$. The "CONSIDER/CALL" group comprises a number of verbs expressing opinions, feelings, or facts. Thus, these verbs provide a solid usage foundation for the predicative ASCs with the semantics "X THINKS/FEELS/DECLARES Y to BE Z" for the syntactic pattern $NP_x V NP_y AdjP_z$.

2.3 Examination of target verb entries in the "Big Five"

We then looked up all the target verb entries, and examined how the target *V N ADJ* ASCs are presented in the "Big Five" through a detailed analysis of their macro- and micro-structures. Macrostructurally, we explored the extent to which constructions are treated as multiword expressions like a phrase/idiom, or placed under the entry verb to illustrate one of its senses. Microstructurally, we adapted previous frameworks for microstructural classification and examination (e.g., Atkins and Rundell 2008: 203-246), and concentrated on the following devices: definition, valency pattern label by means of syntactic code (see CALD and COBUILD) or pattern illustration (see OALD, LDOCE, and MED), and illustrative example. We investigated all the entry components in the "Big Five", and calculated the frequencies each device is used for presenting verb-based *V N ADJ* ASCs.

3. Results

Table 2 displays the overall coverage of *V N ADJ* ASCs in the "Big Five". It

demonstrates that the inclusion of the 40 instances in each dictionary differs. COBUILD includes all of them: while three (i.e., *consider*, *set*, and *believe*) are solely labeled with syntactic codes, the others are provided with examples as well, and some are further illustrated in definitions. OALD also has an extensive coverage, with only the TOWEL *N ADJ* ASC excluded. In contrast, the number of *V N ADJ* ASCs covered in the other three ELDs is relatively low: LDOCE presents 30, CALD shows 28, and MED illustrates 25. The three ELDs also vary markedly in terms of *V N ADJ* ASCs that are not presented. A Chi-square test indicates that the coverage varies significantly across the "Big Five" online ELDs ($\chi^2 = 29.435$, $df = 4$, $p < .05$).

Table 2: Coverage of the 40 instances of *V N ADJ* ASCs in the "Big Five"

	OALD	LDOCE	CALD	COBUILD	MED
inclusion	39	30	28	40	25
phrase	0	21	3	2	8
idiom	4	0	2	0	0
definition	54	60	37	54	36
valency pattern	53	58	31	77	33
illustrative example	83	167	70	64	43

3.1 Macrostructure

The rise of corpus linguistics and the feasibility of studying language data in quantitative dimensions open up entirely new possibilities for presenting phraseology in dictionaries (Rundell 1999; Dobrovol'skij 2015; Steyer 2015). Multi-word expressions (MWEs) such as phrases and idioms demonstrate the collocational preferences of a particular lexical unit in a prominent way. The five ELDs differ significantly in treating *V N ADJ* ASCs as MWEs.

As indicated in Table 2, OALD gives four instances of *V N ADJ* ASCs the status of idioms, and lists them in the IDIOMS column, which is located at the bottom of the main entry. They are *make something good*, *scare somebody shitless*, *wipe the slate clean*, and *knock somebody dead*. Among them, *make something good* and *knock somebody dead* are provided with not only definitions but also example sentences, whereas the other two are only indicated in definitions. Similarly, CALD treats two instances (viz., *make it quick/fast* and *scare sb shitless*) as idioms, places them in the IDIOMS column at the bottom of the entry, and uses hyperlinks to direct the constructions to the idioms, which are illustrated with definitions and examples. In addition, CALD presents three instances (i.e., *keep*

sth quiet, drive sb mad, crazy, etc., and drive sb wild) as phrases in the main entry, and provides them with definitions and examples.

In LDOCE, 21 instances of *V N ADJ* ASCs enjoy the status of being phrases in two different ways: (a) six are treated as **"PHRASES" in the COLLOCATIONS column**, which is placed at the end of an entry; and (b) fifteen are displayed as phrases with hyperlinks, which are listed within the main entry and numbered in the same way as other sense divisions. The six **"PHRASES" in the COLLOCATIONS column** are provided with examples, and most of them are glossed in brackets. Almost all the hyperlinked phrases are supplied with both definitions and examples except for *keep something quiet* and *not be as black as you are painted*, which are offered with definitions.

COBUILD displays two instances as phrases, namely *wipe the slate clean* and *knock them/'em dead*. They are treated as a separate sense for the headword, and hyperlinked with the phrases which are illustrated with both definitions and examples. MED treats eight instances as phrases, places them in the PHRASES column at the end of each entry, and hyperlinks them with the phrases equipped with definitions and examples.

3.2 Microstructure

3.2.1 Definition

As shown in Table 2, each dictionary splits a different number of senses for the 40 instances of *V N ADJ* ASCs. LDOCE gives the most, viz., 60 with a mean of 1.50 sense divisions, and CALD and MED offer the least, being 37 (0.93) and 36 (0.90) respectively. A one-way ANOVA test indicates that the differences in sense divisions among the "Big Five" are statistically significant ($F(4, 195) = 2.657$, $p < .05$).

Results of post hoc tests further indicate that there are considerable differences between LDOCE and CALD ($p < .05$), and between LDOCE and MED ($p < .05$), whereas the differences between other ELDs are not statistically significant. The differences between LDOCE and CALD/MED may stem from the policy of the former giving many ASCs the status of phrases.

We further compared the treatment of the *HOLD N ADJ* ASC in the "Big Five" against the usage data in the corpus. As discussed in Francis et al. (1996), the verb *hold* in the *V N ADJ* ASC belongs to at least two general meaning groups, namely the "MAKE" group and the "CONSIDER" group. The distributional patterns of the *HOLD N ADJ* in the BNC reveal that it conveys the meaning of "keeping something or somebody in a particular state or position" when it co-occurs with adjectives such as *open, high, close, firm, and steady*, "to keep somebody in a particular place" when combined with adjectives like *captive* and *incommunicado*, and "an opinion or belief" when followed by adjectives such as *responsible, liable, and accountable*.

As far as the sense divisions are concerned, both CALD and COBUILD

split the *HOLD N ADJ ASC* into two senses, whereas the other three ELDs provide no less than three senses. In addition to three sense divisions, LDOCE and MED respectively give *hold someone/something dear* and *hold someone responsible/accountable/liable* the status of phrases, supply them with definitions and examples, and hyperlink them to the phrasal entries.

A closer look at the meanings explained in the "Big Five" reveals that these ELDs differ in the specificity of senses assigned to the *HOLD N ADJ ASC*. All the five ELDs give the meaning of "keeping something or somebody in a particular position or state". The differences lie in that LDOCE provides two senses for *hold something open*, MED differentiates between *hold something steady/shut/still* and *hold something tight/close*, while the other three ELDs do not make such a clear-cut distinction. All of them include its *OPINION* sense, but they indicate it in a markedly different style: CALD explains it in a traditional way, offering the meaning "to believe an idea or opinion", and illustrating it with examples; OALD divides it into two senses; LDOCE and MED treat it as a phrase; and COBUILD uses a full-sentence definition to explain its meaning. Additionally, only LDOCE and MED take into account the meaning "to keep somebody in a particular place", and include *hold somebody/something dear* as a separate phrase.

3.2.2 Valency pattern illustration

The "Big Five" online ELDs vary in the way they convey valency information about *V N ADJ ASCs*.

(1) In OALD, valency patterns of *V N ADJ ASCs* are presented in bold type, for example, ***find somebody/something + adj.*** and ***render something + adj.*** Most pattern illustrations of this type precede example sentences.

(2) Slightly different from OALD, LDOCE and MED use specific words rather than "adj." in the adjective slot, for instance, ***render somebody/something impossible/harmless/unconscious etc.***

(3) CALD encodes valency patterns of *V N ADJ ASCs* with both formal (i.e., [+ adj]) and functional categories (i.e., [+ obj]). Patterns used in CALD are somewhat inconsistent and vary slightly for different verb-based *ASCs*, as displayed below:

- (a) [+ adj] (*render, towel, make, set, prove, and send*);
- (b) [+ noun/adj] (*deem and certify*);
- (c) [+ obj + adj] (*find, keep, leave, shoot, scrub, paint, hold, presume, knock, shake, and believe*);
- (d) [+ obj + noun/adj] (*pronounce and deem*);
- (e) [+ obj + (to be) + noun/adj] (*declare*).

(4) COBUILD devises two ways to present valency information of *V N ADJ* ASCs: (a) specific pattern illustrations embodied in full-sentence definitions, and (b) general pattern illustrations in terms of pure formal categories. The first type of pattern illustrations is designed for verbs such as *render*, *towel*, *prise*, *wrench*, *find*, *make*, and *hold*, while the second is provided for every verb-based *V N ADJ* ASCs and placed after the corresponding examples. In addition, COBUILD makes a distinction between verb-based *V N ADJ* ASCs in the active form (e.g., [*V n adj to-inf*] for *make*) and those in the passive form (e.g., [*be V-ed adj/n*] for *deem* and *set*), and provides them with different patterns. In contrast, the other four ELDs fail to make such a distinction.

COBUILD offers an exhaustive list of valency patterns for *V N ADJ* ASCs. However, in COBUILD, verbs in the same class are sometimes not given a consistent pattern. Take *cram* and *stuff* as an example. While the valency pattern provided for the illustrative example *I crammed my bag full of swimsuits and T-shirts ...* is [*V n full of*], the pattern for the sentence example *He grabbed my purse, opened it and stuffed it full, then gave it back to me ...* is [*VERB noun adjective*]. Moreover, previous research (Bogaards and Van der Kloot 2001) has questioned the usefulness of the rich syntactic information indicated after illustrative examples, because most learners are unable or unwilling to take note of such information (Bogaards 2003).

The "Big Five" vary in the number of valency pattern illustrations (see Table 2). A one-way ANOVA test shows that the differences are statistically significant ($F(4, 195) = 4.533, p < .05$). Post hoc tests further indicate that there are considerable differences between COBUILD and CALD, between COBUILD and MED, and between LDOCE and CALD, while the differences between other ELDs are not statistically significant. The differences between COBUILD and CALD/MED may be attributed to the policy of COBUILD providing every example with a corresponding pattern code on the one hand, and the lower coverage of *V N ADJ* ASCs in CALD and MED on the other hand.

Even though there is considerable variation among the "Big Five", their coding systems share a common feature that they assume very little grammatical knowledge on the part of users. They aim to satisfy users' needs in a much more explicit and self-explanatory way.

3.2.3 Illustrative example

The "Big Five" always give at least one example to illustrate *V N ADJ* ASCs they include, except for COBUILD which fails to provide one for *CONSIDER N ADJ*, *SET N ADJ*, and *BELIEVE N ADJ*. As Table 2 indicates, the "Big Five" vary considerably in the number of examples used to illustrate *V N ADJ* ASCs. A one-way ANOVA test reveals that the differences are statistically significant ($F(4, 195) = 9.455, p < .05$). Post hoc tests further indicate that LDOCE substantially differs from the other four ELDs, while the differences between other

ELDs are not statistically significant. The differences may be due to the policy of LDOCE providing plentiful examples in the extra column "Examples from the Corpus" on the one hand, and giving many ASCs the phrase status and equipping them with abundant examples on the other hand.

4. Discussion

This section discusses whether the treatment of *V N ADJ* ASCs in the "Big Five" online ELDs systematically reflects the language in usage, and proposes some suggestions for optimizing the presentation of ASCs in ELDs.

4.1 Strengths of the treatment of ASCs in each ELD

Based on the analysis above, it can be concluded that OALD, LDOCE and COBUILD provide much richer and more comprehensive constructional information for *V N ADJ* ASCs than CALD and MED do. Specifically, as far as the 40 instances of *V N ADJ* ASCs are concerned, OALD and COBUILD include almost all of them, and provide not only illustrative examples but also valency pattern illustrations. For *V N ADJ* ASC typical of a verb, they highlight the construction as a whole either by giving a pattern illustration before its definition (OALD) or by explaining it in a full-sentence definition (COBUILD).

LDOCE is relatively richer in providing constructional information. Despite the fact that the total number of *V N ADJ* ASCs included in LDOCE is lower than in OALD and COBUILD, LDOCE offers abundant examples for each instance of *V N ADJ* ASCs, not only in the main entry but also in the "Examples from the Corpus" column. It is worth noticing that all the five ELDs exhibit a phraseology-orientation towards constructions, but it is LDOCE that gains prominence, because it gives a much larger number of *V N ADJ* ASCs the status of phrases, and treat them as a separate sense-division.

In contrast to the other three ELDs, CALD and MED present *V N ADJ* ASCs in a simpler way. They include a relatively lower number of ASCs. In addition, the number of definitions, illustrative examples, and valency pattern illustrations they provide for *V N ADJ* ASCs is far lower than the other three, and the constructional information they offer is inadequate.

4.2 Inadequate treatment of ASCs in current online ELDs

The treatment of *V N ADJ* ASCs in the "Big Five" is inadequate in the following respects.

4.2.1 Insufficient inclusion of ASCs in ELDs

As explained in Section 2.2.2, the 40 instances of *V N ADJ* ASCs examined are

prototypical ones. Thus, they deserve due attention, and are supposed to be presented in ELDs. However, a close investigation into the coverage of them in the "Big Five" online ELDs reveals that they are inadequately included in current ELDs.

The reasons might be as follows. Firstly, whether a construction for a particular verb is included in an ELD might be related to the frequency of the verb as well as the frequency of the construction in usage. Verbs in the MAKE group such as *make*, *render*, *keep*, *leave*, and *hold* have a very high token frequency in usage, and the *V N ADJ* ASCs for them are prototypical ones. Thus, they attract a great deal of attention from dictionary compilers, and their various constructions, including *V N ADJ* ASCs, are adequately presented in dictionaries. In contrast, the fact that instances of some *V N ADJ* ASCs like *CRAM N ADJ*, *STUFF N ADJ*, and *PRISE N ADJ* are unduly treated in ELDs might be due to the relatively lower token frequency of such verbs as *cram*, *stuff*, and *prise*. Nevertheless, the collostructional analysis shows that their *V N ADJ* ASCs are faithful and prototypical for each verb. Therefore, all of them should be covered from the perspective of the usage-based construction grammar. Secondly, it might be attributed to the different types of constructions they belong to. *V N ADJ* ASCs for verbs like *make*, *render*, *keep*, *leave*, etc. are instances of Type A resultative ASCs, and the words in every slot in the construction are obligatory. In contrast, in Type B resultative ASCs such as *CRAM N ADJ*, *STUFF N ADJ*, *SLAM N ADJ*, *WIPE N ADJ*, and *TOWEL N ADJ*, the adjective result phrase is optional, and it can be omitted without affecting the grammaticality of the sentence. The result is implied by the verbal event, and can be conveyed by a restricted range of adjectives. Generally, there is little difference between the sentences with and without the adjective phrase. Given that, some dictionary compilers might think it unnecessary to include both of them in ELDs. As a result, some of the Type B resultative ASCs are presented in ELDs, whereas some are not included. However, based on the construction grammar approach, a sentence without an adjective phrase focuses on the verbal event while a sentence with an adjective phrase not only conveys the verbal event but also specifies the resultative event. Therefore, they belong to two different constructions, and deserve to be fully presented and differentiated in dictionaries.

Although lexicographers have attached importance to frequently-used headwords and to constructions of words with a high frequency, little attention has been given to constructions of words with a low frequency. In addition, there is no consistent policy for including ASCs in the "Big Five". Therefore, with respect to the coverage of ASCs in ELDs, there is still room for improvement.

4.2.2 Mismatch between valency pattern illustrations and illustrative examples

A scrutiny of the treatment of ASCs in specific entries indicates that the "Big Five" sometimes inconsistently encode the constructional information in valency

pattern illustrations and illustrative examples. As the following shows, for some ASCs, there is no one-to-one match between pattern illustrations and sentence examples.

(1) In some cases, there is not an adequate number of examples to match a pattern illustration. For example, OALD labels **WRENCH N ADJ** with the pattern illustration *wrench (something/somebody/yourself) + adj.*, and provides two illustrative examples (viz., *They wrenched the door open.* and *She managed to wrench herself free.*), but it does not exemplify whether the pattern without the object **something/somebody/yourself** has a similar meaning. In the same vein, OALD supplies only one example (i.e., *Scrub the vegetables clean.*) to the pattern *scrub something/yourself + adj.*

(2) Sometimes the examples provided do not match the patterns to illustrate. For instance, in the "Examples from the Corpus" column in LDOCE, the example sentence *Once a soldier has been certified medically deaf he was always shipped home.* is not matched with the illustrative pattern *certify (that)*, and the sentence example *The defendant is presumed innocent until proved guilty.* is located under the pattern *presume* instead of the pattern *be presumed dead/innocent etc* (see Figure 2). Nevertheless, it is worth noticing that most examples in the "Examples from the Corpus" column in LDOCE are arranged according to their patterns, while the examples in the "Extra examples" column in OALD are not sorted.

Examples from the Corpus

presume

- The price includes all your transportation and hotels, I **presume**?
- Was this the treatment Roman meted out to any female who **presumed** a little too much, grew a little too possessive?
- They say Mind **presumed** an inquiry would involve the families and those advising them and we were astonished when it did not.
- One **presumes** his wife Eimear knew that when she married him.
- As a 19 year-old student, I **presume** I am one of these.
- The defendant is **presumed** innocent until proved guilty.

be presumed dead/innocent etc

- An aerial search of the area proved fruitless, and they **were presumed dead**.
- Nineteen were killed and two went missing and **are presumed dead**.
- The mysterious Mr Kipper has never been traced and Susie **is presumed dead**.
- He **was presumed dead**, but where?

Figure 2: "Examples from the Corpus" for *presume* in LDOCE

It is also untenable for LDOCE to place these illustrative examples at the end of the whole entry, because they are not matched with the senses and patterns

they are intended to illustrate on the one hand, and they might not be noticed by users on the other hand. Instead, they can be directly placed under relevant senses and patterns they aim to illustrate by means of "icons for collapsing and expanding them" (Rundell 2015: 320).

(3) In some cases, some ELDs fail to distinguish the active form from the passive form. In other words, the valency pattern illustrations provided are in the active form, while the corresponding illustrative examples are in the passive form. For example, OALD provides the example *Three people were shot dead during the robbery.* for the illustrative pattern **shoot somebody/something + adj.**; LDOCE supplies the example *The driver was certified dead at the scene.* to the pattern **certify somebody dead**; and CALD provides the examples *The boat's captain is missing, presumed dead.* and *In British law, you are presumed innocent until you are proved guilty.* for the syntactic code [+ obj + adj]. COBUILD is an exception: it makes a distinction between the active and passive forms, and offers a one-to-one match between a syntactic code and an illustrative example (see Section 3.2.2).

4.2.3 Unsystematic treatment of ASCs of the same type

As an important reference tool, ELDs are expected to treat constructions of the same type in a consistent way. However, an examination of the constructional information presented in the "Big Five" demonstrates that some ELDs fail to meet this criterion.

Firstly, the policy for including ASCs is not consistent in some ELDs. Of the same type of constructions, some are listed, and some are not. For example, CALD includes PROVE *N ADJ* and THINK *N ADJ*, but fails to cover RATE *N ADJ*, PROFESS *N ADJ*, JUDGE *N ADJ*, and CALL *N ADJ* despite the fact that the degree of prototypicality of the latter four is higher than that of the first two. A similar problem is also found in LDOCE and MED.

Secondly, the valency pattern illustrations used to encode the same type of ASCs are not consistent, and sometimes are even confusing. For example, in CALD, constructions like RENDER *N ADJ*, MAKE *N ADJ*, KEEP *N ADJ*, LEAVE *N ADJ* and HOLD *N ADJ* belong to the resultative ASCs, but they are labeled with [+ adj] for the first two, and [+ obj + adj] for the latter three. Even instances of the same verb-based ASCs are encoded differently. For example, the sentence example *The area has now been deemed safe.* is labeled with [+ obj + noun/adj], while *We will provide help whenever you deem it appropriate.* is encoded as [+ noun/adj]. A similar problem exists in COBUILD.

The semantic differences between the resultative and the predicative *V N ADJ* ASCs are not negligible (see Section 2.2). Thus, it is necessary to devise an effective way to highlight their differences. However, ELDs such as CALD and COBUILD still resort to the same syntactic label to indicate them.

Thirdly, there is a lack of a clear rule with respect to the specificity of valency pattern illustrations. The usage-based analysis of the distributional

characteristics indicates that the co-varying collexemes for different verbs in *V N ADJ* ASCs vary greatly. In some cases, a wide range of words can be used in each slot of an ASC, for example, *MAKE N ADJ*, and *CONSIDER N ADJ*. In other cases, there are some collocational preferences and semantic prosodic restrictions in the adjective slot. For example, the adjectives occurring in *RENDER N ADJ* usually convey negative prosody, meaning something bad or unexpected. The adjectives used in most Type B resultative ASCs are also restricted. In the case of *confess*, *count*, *profess*, *pronounce*, and *prove*, the nominal group following the verb is always a reflexive pronoun. In indicating the restrictions of such constructions, some ELDs are too general, and some are too specific.

Last but not least, the criteria for assigning an ASC the status of phrase or idiom are not explicit. In most cases, the "Big Five" do not pay due attention to the role of ASCs. The constructional information they provide in syntactic codes or sentence examples is mainly used to illustrate the typical context and usage of the entry word. As for ASCs, some ELDs treat them as phrases or idioms, while some present them in valency pattern illustrations and/or illustrative examples. Specifically, LDOCE gives a large number of *V N ADJ* ASCs the status of phrases, equipping each of them with definitions and examples, and hyperlinking them to the main entry. In other ELDs, most *V N ADJ* ASCs are placed in the main entry and presented in the traditional way: being encoded either in valency pattern illustrations or illustrative examples. For ASCs like *wipe the slate clean* and *knock them/'em dead*, COBUILD displays them as phrases, while OALD treats them as idioms.

4.3 Presentation of ASCs in construction-driven ELDs

Following the usage-based construction grammar approach, we put forward some suggestions for the design of a new type of construction-driven ELD.

Firstly, a construction-driven dictionary should characterize all conventional constructional units in a similar way to the traditional representation of lexical units. Construction refers to a conventional, learned pairing of form and meaning/function at different levels of schematicity, ranging from substantive word construction to highly schematic ASC. Phrasal units share commonalities with words in the traditional sense. Therefore, in the new type of ELD, the traditional headwords are "no more than access points" (Moon 2008: 253), and can be substituted by constructions. Whether a particular construction will be included in an ELD or not is, in principle, contingent upon its frequency and prototypicality in language usage. Since constructions gain the status of traditional headwords, they can be placed at the beginning of each entry, highlighted in bold type, and followed by corresponding definitions and illustrative examples.

Secondly, context predetermines meaning (Moon 2008). Corpus linguistic research over the past four decades has undermined the traditional notion of words as the autonomous bearer of meaning, stimulated the development of

the idea that meanings are mostly constructed through context, and revealed that recurrence and regularity are essential features of the language system (Rundell 2018). In the new type of ELD, what is defined is the extended constructional unit of meaning as a whole instead of the headword in isolation. As far as the defining style is concerned, it is advisable to make reference to existing definition models as adopted in the "Big Five", insights gained from users' research, and innovations in valency databases such as *A Valency Dictionary of English*, *FrameNet* and *Pattern Dictionary of English Verbs*. In addition, different constructions with the same pattern (e.g., the predicative construction *hold somebody responsible* and the resultative construction *hold something open*) should be distinguished, and listed in separate entries.

Thirdly, illustrative examples offered in the new type of ELD are intended to help advanced learners with productive activities. Corpus-based examples adapted for lexicographic purposes are preferred in most cases. Consistency should be maintained between examples and the patterns/definitions they are intended to illustrate. In the digital age, space is no longer a major concern as of printed dictionaries. The quantity of examples provided will not have an impact on the cost of the e-dictionary. Nevertheless, too many examples could be distracting, resulting in an information overload that is actually detrimental to learning. Hyperlinks can be utilized to relate entries to more illustrative examples. Encoding examples provided in the extra column are supposed to be sorted by patterns, and put directly under relevant senses (rather than at the end of the whole entry). In addition, it is advisable to highlight the constructions embodied in examples in bold type, thus increasing users' awareness of the target construction as a conventional unit.

5. Conclusion

This article examined the treatment of *V N ADJ* ASCs in the "Big Five" against the results of the usage-based constructional analysis of them in the BNC. It was found that the treatment of constructions in existing ELDs is inadequate, and needs improvement. Some suggestions for presenting ASCs in a construction-driven ELD were also proposed.

This study has some implications for the development of a new type of ELD. Traditional headwords should be substituted with constructions. It is constructional units rather than words in isolation that are supposed to be defined. In addition, different devices (e.g., definition, pattern illustrations, and sentence examples) can be combined to present constructions at different levels of generality and schematicity.

It is impossible for one article to adequately address all the issues relevant to the presentation of constructions in ELDs. Future research can enlarge the sample to consider more ASCs and ELDs, further elaborate on how to implement the basic tenets of the usage-based construction grammar in practical lexicography, and investigate learners' reference needs for a corpus-based online ELD.

Acknowledgements

This study was supported by a grant from the National Philosophy and Social Sciences Foundation of China (Grant No. 22BYY010), and the Humanities and Social Sciences Foundation of Ministry of Education of China (Grant No. 23YJC740013). We are grateful to Mr. André du Plessis and two anonymous adjudicators for their valuable comments and suggestions. Correspondence should be addressed to Dr. Hai Xu.

Endnotes

1. Precision is defined as the quotient of the number of accurate matches returned by a search divided by the number of all matches returned by a search, whereas recall is defined as the number of accurate matches returned by a search divided by the number of all possible accurate matches in the data (Gries 2009: 16).
2. The 13 verbs that do not occur in *V N ADJ* ASCs in the BNC are *label, shove, tape, batter, crank up, turn down, turn up, spray, slice, capture, be born, picture, and serve*. Because our research aims to provide usage evidence for lexicographic presentation that prioritizes prototypical usage, we excluded these verbs from our analysis although some of them (e.g., *shove, tape, batter, slice, and capture*) exhibit characteristics of *V N ADJ* ASCs in our searches of the English Web 2015 (enTenTen 15) via Sketch Engine.
3. See also <https://www.stgries.info/teaching/groningen/index.html>.

References

A. Dictionaries

- CALD.** *Cambridge Advanced Learner's Dictionary*. <https://dictionary.cambridge.org>.
COBUILD. *COBUILD Advanced English Dictionary*. <https://www.collinsdictionary.com/>.
LDOCE. *Longman Dictionary of Contemporary English*. <https://www.ldoceonline.com/dictionary>.
MED. *Macmillan English Dictionary*. <https://www.macmillandictionary.com/>.
OALD. *Oxford Advanced Learner's Dictionary*. <https://www.oxfordlearnersdictionaries.com/>.

B. Other literature

- Atkins, S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Boas, H.C.** 2003. *A Constructional Approach to Resultatives*. Stanford: CSLI Publications.
- Boas, H.C.** 2009. Verb Meanings at the Crossroads between Higher-level and Lower-level Constructions. *Lingua* 120(1): 22-34.
- Boas, H.C.** 2014. Lexical and Phrasal Approaches to Argument Structure: Two Sides of the Same Coin. *Theoretical Linguistics* 40(1-2): 89-112.

- Bogaards, P.** 2003. MEDAL: A Fifth Dictionary for Learners of English. *International Journal of Lexicography* 16(1): 43-55.
- Bogaards, P. and W.A. van der Kloot.** 2001. The Use of Grammatical Information in Learners' Dictionaries. *International Journal of Lexicography* 14(2): 97-121.
- Cheng, G. and H. Xu.** 2022. Treatment of the Way-construction in English–Chinese Learners' Dictionaries: A Construction Grammar Approach. *International Journal of Lexicography* 35(1): 107-128.
- Croft, W. and M. Vigus.** 2017. Constructions, Frames and Event Structure. *Papers from the 2017 AAAI Spring Symposium on Computational Construction Grammar and Natural Language Understanding, 27–29 March 2017, Stanford University*: 147-153. Stanford, CA: Stanford University.
- Dobrovolskij, D.** 2015. Introduction. *International Journal of Lexicography* 28(3): 275-278.
- Faulhaber, S.** 2011. *Verb Valency Patterns: A Challenge for Semantics-based Accounts*. Berlin: Mouton de Gruyter.
- Francis, G., S. Hunston and E. Manning.** 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
- Goldberg, A.E.** 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A.E.** 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, A.E.** 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Goldberg, A.E., D.M. Casenhiser and N. Sethuraman.** 2004. Learning Argument Structure Generalizations. *Cognitive linguistics* 15(3): 289-316.
- Goldberg, A.E. and R. Jackendoff.** 2004. The English Resultative as a Family of Constructions. *Language* 80(3): 532-568.
- González-García, F.** 2009. The Family of Object-related Depictives in English and Spanish: Towards a Usage-based Constructionist Analysis. *Language Sciences* 31(5): 663-723.
- Gries, S.T.** 2009. *Quantitative Corpus Linguistics with R: A Practical Introduction*. London: Routledge.
- Gries, S.T.** 2014. Coll.Analysis 3.5. A Script for R to Compute Perform Collostructional Analyses.
- Hampe, B.** 2011. Discovering Constructions by Means of Collostruction Analysis: The English Denominative Construction. *Cognitive Linguistics* 22(2): 211-245.
- Hanks, P.** 2008. The Lexicographical Legacy of John Sinclair. *International Journal of Lexicography* 21(3): 219-229.
- Hanks, P.** 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.
- Hanks, P.** 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, Mass.: MIT Press.
- Hanks, P. and W. Ma.** 2020. Meaning and Grammar in the Light of Corpus Pattern Analysis. *International Journal of Lexicography* 34(1): 135-149.
- Hanks, P. and S. Može.** 2019. The Way to Analyse 'way': A Case Study in Word-specific Local Grammar. *International Journal of Lexicography* 32(3): 247-269.
- Herbst, T. and M. Klotz.** 2009. Syntagmatic and Phraseological Dictionaries. Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography*: 234-243. Oxford: Oxford University Press.
- Herbst, T., H.-J. Schmid and S. Faulhaber (Eds.).** 2014. *Constructions Collocations Patterns*. Berlin/Boston: De Gruyter Mouton.
- Hoey, M. and M.B. O'Donnell.** 2008. Lexicography, Grammar, and Textual Position. *International Journal of Lexicography* 21(3): 293-309.

-
- Hoffmann, T. and G. Trousdale (Eds.).** 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
- Hunston, S.** 2014. Pattern Grammar in Context. Herbst, T., H.-J. Schmid and S. Faulhaber (Eds.). 2014. *Constructions Collocations Patterns*: 99-120. Berlin/Boston: De Gruyter Mouton.
- Hunston, S.** 2019. Patterns, Constructions, and Applied Linguistics. *International Journal of Corpus Linguistics* 24(3): 324-353.
- Hunston, S. and G. Francis.** 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Studies in Corpus Linguistics 4. Amsterdam/Philadelphia: John Benjamins.
- Hunston, S. and H. Su.** 2019. Patterns, Constructions, and Local Grammar: A Case Study of 'Evaluation'. *Applied Linguistics* 40(4): 567-593.
- Kilgarriff, A.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý and V. Suchomel.** 2014. The Sketch Engine: Ten Years On. *Lexicography* 1(1): 7-36.
- Moon, R.** 2008. Sinclair, Phraseology, and Lexicography. *International Journal of Lexicography* 21(3): 243-254.
- Perek, F.** 2015. *Argument Structure in Usage-Based Construction Grammar. Experimental and Corpus-based Perspectives*. Constructional Approaches to Language 17. Amsterdam/Philadelphia: John Benjamins.
- Przepiórkowski, A., J. Hajič, E. Hajnicz and Z. Urešová.** 2017. Phraseology in Two Slavic Valency Dictionaries: Limitations and Perspectives. *International Journal of Lexicography* 30(1): 1-38.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik.** 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-53.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Rundell, M.** 2018. Searching for Extended Units of Meaning — and What to Do When You Find Them. *Lexicography* 5(1): 5-21.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.** 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stefanowitsch, A. and S.T. Gries.** 2003. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.
- Steyer, K.** 2015. Patterns. Phraseology in a State of Flux. *International Journal of Lexicography* 28(3): 279-298.
- Stubbs, M.** 2009. The Search for Units of Meaning: Sinclair on Empirical Semantics. *Applied Linguistics* 30(1): 115-137.
- Thomas, J.** 2014. Discovering English with the Sketch Engine. Jager, S., L. Bradley, E.J. Meima and S. Thouësny (Eds.). 2014. *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands*: 363-367. Dublin: Research-publishing.net.

Mit Wortschatz und lexikografischen Ressourcen handeln: kritische Überlegungen zur Anwendung lexikalischer, lexikografischer und digitaler Kompetenzen im virtuellen Raum beim akademischen Schreiben im DaF-Bereich

Meike Meliss, *Universidad de Santiago de Compostela, Spanien*
(meike.meliss@usc.es — ORCID: 0000-0003-0380-8905)

Abstract: Der Beitrag beschäftigt sich kritisch mit dem Wortschatzerwerb beim akademischen Schreibprozess im fortgeschrittenen DaF-Bereich in Zusammenspiel mit der Entwicklung einer angebrachten Sprachbewusstheit und einer aktuellen und umfassenden digitalen Kompetenz. Ausgangspunkt ist die Beobachtung, dass im akademischen L2-Schreibprozess das umfangreiche Informationspotenzial sowohl der ein-, zwei- und mehrsprachigen lexikografischen und lexiko-grammatischen Online-Ressourcen sowie anderer sprachtechnologischer Werkzeuge nur ungenügend bekannt ist. Dementsprechend werden die vielzähligen digitalen Sprachressourcen nicht optimal für Methoden der Vermittlung von Wortschatz und der Handlung mit ihm, z.B. beim akademischen Schreiben herangezogen. Andererseits nehmen die digitalen Medien immer mehr Raum in unserem Alltag ein und stellen auch sprachtechnologisch hoch entwickelte Werkzeuge für verschiedene L2-Kompetenzbereiche, wie z.B. das Übersetzen bereit. Ziel des Beitrages ist es daher, sowohl Wege für die sinnvolle Nutzung der vorhandenen Internetressourcen aufzuzeigen als auch ganz neue auf künstlicher Intelligenz beruhende Wege zu skizzieren. An ausgewählten Ressourcen soll exemplarisch das umfangreiche Informationspotenzial aufgezeigt sowie Möglichkeiten der Nutzung beim akademischen Schreiben diskutiert werden. Ein Ausblick verweist auf die Notwendigkeit, sich auf revolutionäre technologische Veränderungen im Bereich der Fremdsprachdidaktik einzustellen.

Schlüsselwörter: LERNERLEXIKOGRAFIE, WORTSCHATZERWERB, AKADEMISCHES SCHREIBEN, MULTILATERALITÄTEN, ONLINE-LEXIKOGRAFIE, SPRACHTECHNOLOGIEN, DEUTSCH ALS FREMDSPRACHE, KORPUSLINGUISTIK, SCHREIBASSISTENT, KÜNSTLICHE INTELLIGENZ

Abstract: Acting with Lexicon and Lexicographical Resources: Critical Reflections on the Use of Lexical, Lexicographic and Digital Literacy in Virtual Environments in Academic Writing in GfL.

The paper deals with vocabulary acquisition in the academic writing process in advanced GfL in interaction with the development of appropriate language awareness and wide-ranging up-to-date digital literacy. The starting point is the observation that in the academic L2 writing process the extensive information potential of monolingual, bilingual and multilingual lexicographic and lexico-grammatical online resources as well as other language technology tools is not adequately known. Accordingly, the many digital language resources are not optimally used for methods of teaching vocabulary and acting with it, e.g. in academic writing. On the other hand, digital media occupy more and more space in our everyday life and also provide highly developed tools based on linguistic technology for various L2 competence areas, such as translation. The aim of this paper is therefore both to show ways of making sensible use of existing internet resources and to outline entirely new ways based on artificial intelligence. Selected resources will be used as examples to demonstrate the extensive information potential and to discuss possibilities for use in academic writing. An outlook points to the necessity of adapting to revolutionary technological changes in the field of foreign language didactics.

Keywords: LEARNER'S DICTIONARIES, VOCABULARY ACQUISITION, ACADEMIC WRITING, MULTILITERACIES, ONLINE-LEXICOGRAPHY, LANGUAGE TECHNOLOGIES, GERMAN AS FOREIGN LANGUAGE, CORPUSLINGUISTICS, WRITING ASSISTANT, ARTIFICIAL INTELLIGENCE

1. Einleitung

Die aktuelle digitale Transformation betrifft alle Bereiche unseres Alltages (vgl. Pressemitteilung ARD/ZDF-Onlinestudie 2020) und so u.a. auch das Fremdsprachenerlernen und -lehren (Biebighäuser und Feick 2020; Biebighäuser et al. 2021; Biebighäuser 2022; Braun und Klimaszyk 2022; Vuorikari et al. 2022)¹. Dies setzt Multiliteralitäten² voraus, um sich im digitalen Kontext bewegen und entsprechend handeln zu können (Küster 2014; Bothma und Gouws 2020). Der immer stärker werdende Einfluss der Künstlichen Intelligenz (KI) auf das Fremdsprachenerlernen sowie die sprachliche Interaktion allgemein (Strasser 2020; Schmidt und Strasser 2022) erfordert eine aktuelle Auseinandersetzung mit den erforderlichen Kompetenzen aller Akteure im Fremdsprachenerwerbsbereich. Nur so können auf KI basierende aktuelle und zukünftige Sprachtechnologien als Herausforderungen und Chancen für den DaF-Bereich und nicht als „desruptiv“ (Hartmann 2021) verstanden werden.³

Im virtuellen Raum, der in diesem Beitrag als digitaler Handlungsbereich verstanden wird, über den im Internet sowohl Lehr- und Lernaktivitäten als auch andere berufliche und alltägliche Aktivitäten stattfinden, gewinnt das schriftbasierte Handeln immer mehr an Bedeutung⁴. Diese Verlagerung auf neue Handlungsräume wird auch im L2-Schreibprozess wahrgenommen. Dort kann u.a. das Schreiben unter Benutzung der existierenden technischen Hilfsmittel und digitalen Tools (u.a. lexikografische Ressourcen, Schreibassistenten)

erfolgreich gelingen und die Schreibkompetenz bzw. Schreibfähigkeit aktiv gefördert werden. Besonders im universitären DaF-Umfeld versteht sich das akademische Schreiben mit seinen wissenschaftssprachlichen Strukturen und Merkmalen (Thielmann 2017) als Teilbereich einer wichtigen handlungsorientierten sprachlichen Aktivität und verlangt besondere Aufmerksamkeit (Ehlich 2000, 2003, 2018; Fandrych und Thurmair 2011; Nardi und Farroni 2022).

Sowohl aus der Lehr- als auch aus der Lernperspektive steht die Beschäftigung mit lexikalischen Fragestellungen in Verbindung mit fremdsprachigen Produktionssituationen wie dem akademischen Schreiben im universitären DaF-Bereich eine spezielle Herausforderung dar. Für den Erwerb von Wortschatz und dessen handlungsorientiertem Gebrauch sind neben dem Einsatz von diversen lexikalischen und lexikografischen Kompetenzen auch digitale Kompetenzen erforderlich, um den zahlreichen nicht nur textsortenspezifischen Anforderungen gerecht zu werden (Kretzenbacher 1998; Jaworska 2011; Graefen und Moll 2011; Graefen 2014; Andresen und Knorr 2017). Eine schon seit geraumer Zeit sehr aktive Wörterbuchbenutzungsforschung versucht aus benutzerorientierter Perspektive die unterschiedlichen Bedürfnisse zu eruieren, die lexikografischen Angebote und ihren Nutzen zu untersuchen und die diversen Kompetenzen zu definieren (Müller-Spitzer et al. 2018; Meliss et al. 2018; Nied Curcio 2020, 2022; Wolfer et al. 2021). Allerdings wird spätestens mit der von Nied Curcio (2014) formulierten Frage, ob im Zeitalter des Smartphones überhaupt noch Vokabeln gelernt werden, deutlich, dass z.B. das Handeln mit Wortschatz im DaF-Bereich im virtuellen Raum teilweise auf nicht kontrollierbare Bereiche verlagert wird. Obwohl das Schreiben von Texten heute schon zu einem großen Teil unter Einbezug von computerbasierten und digitalen Technologien stattfindet, die auf künstlicher Intelligenz (KI) basieren, liegen entsprechende Erhebungen zu der Nutzung der entsprechenden Technologien, Ressourcen und Werkzeuge bei dem Wortschatzerwerb und konkret bei dem handlungsorientierten Gebrauch von Wortschatz im akademischen Schreibprozess bis jetzt kaum vor.

In Anlehnung an diese Vorbemerkungen, lassen sich die Ziele dieses Beitrages wie folgt beschreiben: vor dem Hintergrund von unterschiedlichen sprachdidaktischen und lernpsychologischen Aspekten zum Wortschatzerwerb (Abschnitt 2) sollen am Beispiel einer konkreten Aufgabenstellung aus dem Bereich „akademisches Schreiben“ einige ausgewählte Möglichkeiten aufgezeigt werden, für den Wortschatzerwerb und dessen handlungsorientierten Gebrauch den virtuellen Raum zu nutzen. Das Eindringen der Künstlichen Intelligenz (KI) in den Bereich des Fremdspracherwerbs (Hartmann 2021) führt zu einer kritischen Auseinandersetzung mit den nötigen vielschichtigen Kompetenzen zum handlungsorientierten Wortschatzerwerb mit Hilfe von diversen lexikografischen Sprachressourcen und Sprachtechnologien und dessen konstruktiver Nutzung (Abschnitt 3) und zu der Formulierung einiger Desiderata für unsere neue digitale und vernetzte Lern- und Lehrrealität (Abschnitt 4).⁵

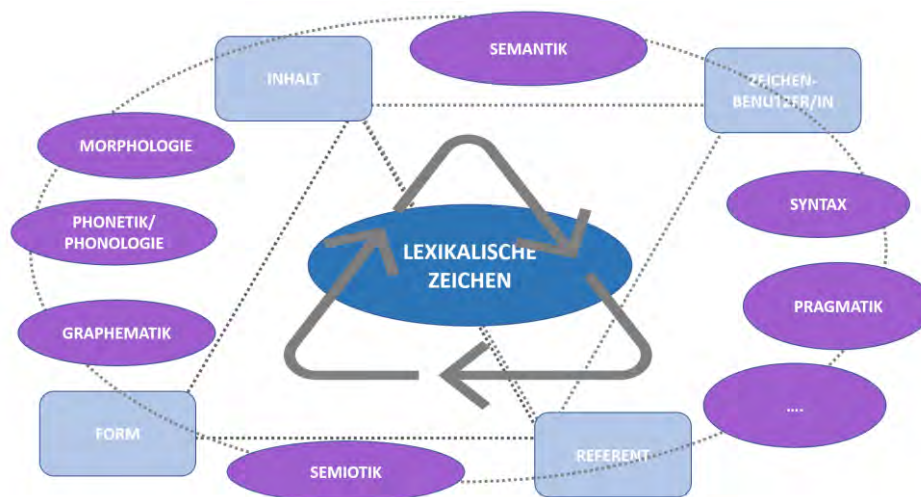
2. Wortschatz: lexikalische Kompetenz, Erwerb und Handeln

Unterschiedlichste Faktoren und methodologische Zugriffe beeinflussen die Entwicklung unseres mentalen Lexikons und die Ausbildung unserer lexikalischen Kompetenz (Leimbrink 2015; Altmayer et al. 2021). Aktuelle sprachtechnologische Entwicklungen rund um die KI stellen uns vor neue Herausforderungen im Umgang mit dem Wortschatz. In Verbindung mit einigen Vorüberlegungen zu dem Begriff "lexikalische Kompetenz" stehen u.a. Fragen danach, welche Information an den Wortschatz gekoppelt ist, wie wir welchen Wortschatz erwerben, wie wir den "Wortschatz im Kopf" definieren und wie wir mit Wortschatz handeln. Ausgangspunkt für mögliche Antworten ist die Auffassung von Wortschatz "als integrative Komponente des Sprachsystems", welche Schippan geprägt hat.

[...] Beim Erlernen einer Sprache spielt die Aneignung der Lexik eine fundamentale Rolle: zuerst erwerben wir die Wörter, erst dann erfolgt die Entfaltung der Grammatik. [...] Daraus kann man die Schlussfolgerung ziehen, dass das Lexikon eine integrative Rolle spielt, dass viele Bereiche unseres sprachlichen Wissens lexikalisch organisiert sind. [...] Wir bezeichnen den Wortschatz als eine integrative Komponente des Sprachsystems. [...] Beim Erlernen der Muttersprache, wie auch später von Fremdsprachen erlernen wir mit den Wörtern/Vokabeln nicht allein die Bedeutungen, sondern auch die Regeln der Aussprache, der grammatischen Formung und Verbindung, die Regeln der Verwendung in bestimmten Kommunikationssituationen [...] (Schippan 1992: 2)

Wenn wir die Auffassung von Schippan teilen, dann hat dies folgende Konsequenzen. An jedes Element aus dem Wortschatz ist eine Vielzahl von Information grammatischer, semantischer, pragmatischer und semiotischer Natur geknüpft. In Abbildung 1 wird versucht, diese integrativen Komponenten zu visualisieren. In dem Erwerbsprozess gilt es, diese unterschiedlichen Informationen kontextadäquat sowohl zu erkennen als auch zu lehren und zu erlernen. Je nach Lern- und Lehrkontext tritt das eine oder andere Informationsmodul mehr oder weniger in den Fokus des konkreten, meist angeleiteten Wortschatzerwerbs. In realen Kommunikationssituationen und bei der handlungsorientierten Nutzung des virtuellen Raumes aber wird idealerweise das Wissen über alle lexikalischen Informationsmodule aktiviert und außerdem ergänzt durch weitere (außersprachliche) Wissensmodule in Verbindung mit der individuellen Rechercheaktivität des jeweiligen Lernenden.

Abbildung 1: Der Wortschatz als integrative Komponente (nach Schippan 1992)



Es ist daher unabdingbar, den Wortschatz in kommunikativen Kontexten zu erwerben, um mit ihm entsprechend handeln zu können. Für den "integrativen" Erwerb von Wortschatz und dessen handlungsorientiertem Gebrauch scheint es daher auch notwendig, u.a. über adäquate Hilfsmittel und Tools zu verfügen, die diesen Ansatz ermöglichen bzw. verfolgen und dem "integrativen" Erwerb von Wortschatz und dessen Gebrauch dienlich sein können. Der virtuelle Raum ist durch das bedarfsorientierte Abrufen von Information in unterschiedlichen Ressourcen ein idealer individualisierbarer Handlungsraum. So verweisen z.B. Internet-Wörterbücher neben der zentralen Information zu dem gesuchten Stichwort teilweise auch auf Flexionsparadigmen, andere digitale Ressourcen und/oder Korpora (vgl. DWDS). Die Nutzung dieser erweiterten Recherchemöglichkeiten über das jeweils existierende Verlinkungsangebot ist abhängig von den jeweiligen individuellen Bedürfnissen und der Situation.

Die neueren sprachtechnologischen Entwicklungen in Verbindung mit der künstlichen Intelligenz verschieben die Perspektive vom Erwerb auf das direkte Handeln mit Sprache und Wortschatz mittels unterschiedlicher Werkzeuge, die beim Sprachproduktionsprozess assistieren (z.B. Schreibassistenten) und/oder Sprache teilweise automatisch generieren (maschinelle Übersetzungen und Schreibprozessunterstützung, Sprachgeneratoren, Spracherkennung, adaptierte Lernprogramme etc.) (u.a. Strasser 2020; Hartmann 2021). Die Nutzung von lexikografischen Ressourcen bildet in diesem Szenario nur eine unter vielen anderen Recherchehandlungen und somit geht die Ausbildung entsprechender vielschichtiger Kompetenzen weit über die lexikografische hinaus. Der Erwerb des mentalen Lexikons kann zwar unterschiedlich fokussiert werden, wird aber allgemein als kumulativer, sukzessiver Prozess verstanden (Henrik-

sen 1999; Nation 2001; Tschirner 2010). Neben methodologischen Ansätzen rund um die Kognitionspsychologie, die Zweitspracherwerbsforschung und Wortschatzdidaktik sowie die Korpuslinguistik (Kühn 2007; Lüdeling und Walter 2009; Leimbrink 2015; Lotze 2018; Strank 2020; Aguado und Warneke 2021; Altmayer et al. 2021; Meissner 2021; Perkuhn 2021) stehen die besagten Ansätze⁶ in Verbindung mit neuere Sprachtechnologien in der aktuellen Diskussion⁶ und werden daher in den folgenden Ausführungen besonders berücksichtigt.

3. Wortschatz im virtuellen Raum: Kompetenzen zum Handeln beim akademischen Schreiben

Der virtuelle Raum, so wie er oben definiert wurde, bietet interessante Möglichkeiten sowohl für den gesteuerten und ungesteuerten Erwerb von Wortschatz als auch für das Handeln mit ihm (Biebighäuser et al. 2021; Biebighäuser und Feick 2020). Der Wortschatz ist als Teil von sprachlicher Äußerung im Internet in den unterschiedlichsten Textsorten und Medialitäten in unserem Alltag überall präsent. In strukturierter Form liegt uns der Wortschatz hauptsächlich in digitalen Sprachressourcen und Werkzeugen vor. Die digitale Transformation hat schon lange den DaF-Bereich erfasst, aber der Zugriff von DaF-Lernenden auf diesen digitalen "Materialpool" erfordert in jedem Fall verschiedene zielgruppenorientierte Kompetenzen für die diversen handlungsorientierten sprachlichen Aktivitäten. Dieser Beitrag will sich genauer mit der Nutzung einiger der Sprachressourcen und Tools beschäftigen, die den Wortschatz als strukturiertes und strukturierbares Inventar präsentieren und für den Wortschatzerwerb und das Handeln mit ihm im virtuellen Raum bereitstellen. Es geht hier sowohl um Internetwörterbücher, Datenbanken und Korpora aber auch um die Nutzung von diversen Apps, die auf der Grundlage von aktuellen KI-unterstützten Sprachtechnologien für das Lehren und Lernen im DaF-Bereich ganz neue Dimensionen eröffnen. In Anknüpfung an die Frage ob Wörterbücher wirklich nützliche und adäquate Werkzeuge bei der Textproduktion und dem Überarbeiten von Texten sind (Wolfer und Müller-Spitzer 2022), stehen die allgemeinen Kompetenzen für die handlungsorientierte Nutzung der besagten Ressourcen im DaF-Bereich in Verbindung mit folgenden Faktoren: (i) der anvisierten Zielgruppe, (ii) der quantitativen und qualitativen Wortschatzauswahl (Geht es um Allgemeinwortschatz oder z.B. Fachwortschatz?), (iii) der Kommunikationssituation (Geht es um Rezeption, Produktion, Interaktion oder Mediation?) und (iv) den Hilfsmitteln und ihrer adäquaten Auswahl und optimalen Nutzung.

Unter der Berücksichtigung dieser Faktoren sind lexikografische, lexikalische, korpuslinguistische und digitale medienspezifische Kompetenzen unabdingbar, um im virtuellen Raum Wortschatz zu erwerben und mit ihm contextadäquat zu handeln. Es gilt, aus dem breiten und uferlosen Informationspool die nötigen Daten über eine adäquate Strategie zu finden und erfolgreich anzuwenden. In den folgenden Abschnitten 3.1 und 3.2 werden exemplarisch

anwendungsorientierte Überlegungen für das akademische Schreiben anhand von unterschiedlichen Strategien diskutiert.

3.1 Anwendungsorientierte Überlegungen: Möglichkeiten und Grenzen

Das Schreiben ist ein Handlungsfeld zum Erwerb und Gebrauch von Wortschatz. Die globalen Kannbeschreibungen für die schriftliche Produktion im B2 Niveau nach dem Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) lautet wie folgt:

Kann mit einem relativ umfangreichen Wortschatz längere und detaillierte Texte zu verschiedenen allgemeinen oder aktuellen Themen schreiben und dabei einen bestimmten Standpunkt darlegen. Kann in Texten zu Themen aus seinem/ihrer Fach- und Interessengebiet eine Argumentation aufbauen und die einzelnen Argumente aufeinander beziehen. Kann bei relativ guter Beherrschung der Grammatik eine Reihe von Konnektoren und anderen Mitteln der Textverknüpfung anwenden, um seine/ihre Ausführungen zu einem klaren, zusammenhängenden Text zu verbinden [...] (Glaboniat et al. 2002: 130)

Wie kann der virtuelle Raum für den DaF-Lernenden effizient und zielgruppenorientiert genutzt werden, um Wortschatz zu erwerben und mit ihm adäquat zu handeln? Die folgenden Überlegungen zu Möglichkeiten und Grenzen beschäftigen sich kritisch mit der Nutzung von Sprachressourcen und Sprachtechnologien für den handlungsorientierten Erwerb und Gebrauch von Wortschatz beim akademischen Schreiben im fortgeschrittenen DaF-Bereich. Dazu wird exemplarisch folgende konkrete sprachliche Handlungssituation mit zielgruppen- und kontextspezifischem Charakter ausgewählt: „Erstellen einer schriftlichen Zusammenfassung in der L2 von einem fachsprachlichen Text/Vortrag zum Thema „Sprachwissenschaft“ im universitären Kontext“.

Ab dem B2 Niveau gehört das Schreiben einer Zusammenfassung von Artikeln oder Beiträgen zu unterschiedlichen Themen von allgemeinem Interesse zu den Kannbeschreibungen der schriftlichen Produktion (Glaboniat et al. 2002: 130). Der GER beschreibt dies wie folgt für B2 und C1:

- C1 Kann lange, anspruchsvolle Texte zusammenfassen.
- B2 Kann ein breites Spektrum von Sachtexten und fiktiven Texten zusammenfassen und dabei die Hauptthemen und unterschiedliche Standpunkte kommentieren und diskutieren. Kann Auszüge aus Nachrichten, Interviews oder Reportagen, welche Stellungnahmen, Erörterungen und Diskussionen enthalten, zusammenfassen.
Kann die Handlung und die Abfolge der Ereignisse in einem Film oder Theaterstück zusammenfassen. (zitiert nach GER-online)

Besonders im universitären Kontext steht das akademische Schreiben im Mittelpunkt der Aktivitäten (Ehlich 2018; Dengscherz 2018, 2020). So ist es z.B. erforder-

lich, schriftliche Zusammenfassungen von fachsprachlichen Texten oder einem wissenschaftlichen Vortrag zu erstellen. Wenn die Studierenden keine L1-Kenntnisse der deutschen Sprache aufweisen, ist die Realisierung dieser schriftlichen Produktionshandlung je nach DaF-Niveau eine nicht einfach zu lösende Herausforderung, die alle sprachlichen Ebenen umfasst. Dieser schriftliche Diskurs erfolgt meistens eigenständig, kann aber geplant und vorbereitet und mit diversen Hilfsmitteln unterstützt werden. Zur Bewältigung dieser Aufgabe müssen u.a. folgende zwei Voraussetzungen gegeben sein: (i) eine adäquate schriftliche/mündliche Rezeption des Ausgangstextes und (ii) Besitz von Kenntnissen über die entsprechenden textsortenspezifischen Merkmale der Textsorte "Zusammenfassung". Mögliche konkrete lexikalische Herausforderungen existieren u.a. in der themenspezifischen Auswahl und dem adäquaten Gebrauch der Lexik — teilweise aus dem fachsprachlichen Bereich — und konkret bezüglich der Vielzahl von Kommunikationsverben (z.B. *berichten*, *darstellen*, *diskutieren*, *erörtern*) für die Redewiedergabe der schriftlichen bzw. mündlichen Information.

Wie können fortgeschrittene DaF-Lernende im universitären Umfeld der Aufgabe gerecht werden und sich den genannten Herausforderungen u.a. im lexikalischen Bereich stellen? Einer der vielen Funktionen von lexikografischen Ressourcen ist die der Assistenz im fremdsprachigen Schreibprozess (Tarp, Fisker und Sepstrup 2017: 495). Voraussetzung dafür ist, dass bestimmte digitale Sprachressourcen bekannt sind und sie genutzt werden, um die Aufgabenstellung zu bewältigen.

Im Folgenden werden daher sieben mögliche Strategien zur Bewältigung der Aufgabe „Erstellung einer schriftlichen Zusammenfassung“ vorgestellt. Die Strategien bezeichnen mögliche Wege, adäquate Hilfestellung zur Bewältigung einer konkreten Aufgabenstellung für das akademische Schreiben zu erzielen und umfassen die Nutzung sowohl von unterschiedlichen lexikografischen Ressourcen (ein-, zwei- und mehrsprachige Lernerwörterbücher, einsprachige Universalwörterbücher, Spezialwörterbücher) als auch von Apps und Korpora. Dabei wird hauptsächlich die Auswahl und der Gebrauch von unterschiedlichen Kommunikationsverben zur Redewiedergabe fokussiert. Exemplarisch wird dabei v.a. von dem deutschen Verb *diskutieren* und/oder dem spanischen *discutir* ausgegangen. Die kritischen Kommentare beschäftigen sich sowohl mit dem Bekanntheitsgrad der Ressourcen als auch mit den erforderlichen Kompetenzen zur erfolgreichen Nutzung und dem zu erwartenden Ergebnis.

Strategie 1: Nutzung gängiger Textverarbeitungsprogramme und externer Schreibtools bzw. Schreibassistenten

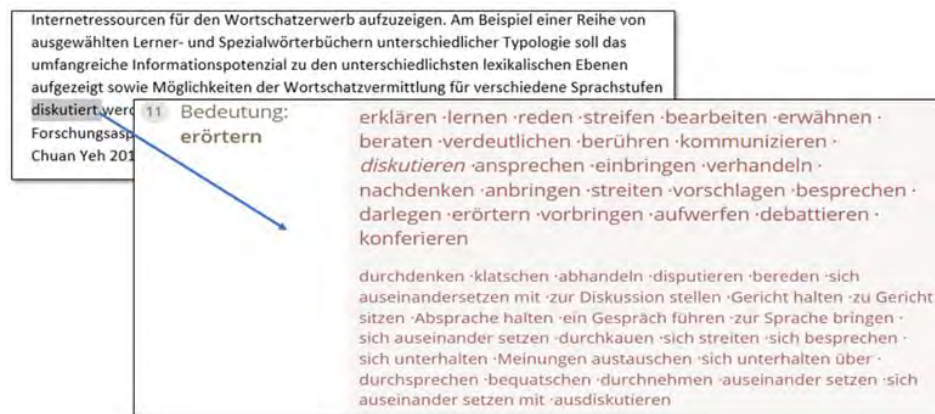
Im Idealfall wird eine Produktionsaufgabe, wie die die für diese Studie formuliert wurde, am PC erledigt. Es sollten also die allgemeinen Voraussetzungen für den adäquaten Arbeitsplatz (Internet, Textverarbeitungsprogramm) existieren. Gängige Textverarbeitungsprogramme bieten nicht nur über die entsprechende Sprachauswahl Korrekturen zu Orthografie und Interpunktion an. Sogar Mög-

lichkeiten der Übersetzung und barrierefreie Kommentare sind inkorporiert. Auch ein Angebot von bedeutungsähnlichen Wörtern kann für stilistische Korrekturen beim Verfassen eines Textes genutzt werden. Ob aber die Zielgruppe wirklich die vielseitigen Möglichkeiten der gängigen Textverarbeitungsprogramme kennt und damit umgehen kann, ist bisher kaum untersucht worden. Anzunehmen ist, dass die Studierenden dieses Angebot nur eingeschränkt in Anspruch nehmen, weil sie eventuell nicht genug Sprachkompetenz besitzen, um z.B. den Unterschied zwischen dem Verb *diskutieren* und den vorgeschlagenen bedeutungsähnlichen Lexemen der Liste (z.B. hier: *erörtern* oder *besprechen*) nicht kennen und sich daher nicht "trauen", Wörter durch andere bedeutungsähnliche in einem konkreten Text (Zusammenfassung eines sprachwissenschaftlichen Artikels) zu ersetzen. Außerdem stehen inzwischen eine ganze Reihe von digitalen Schreibtools zur Verfügung, mit denen die Aufgabe der Erstellung einer Zusammenfassung in einer L2 gut sprachlich bewältigt werden kann. Neben den schon erwähnten Textverarbeitungsprogrammen mit Korrekturoptionen werden auch umfangreiche Tools für den Schreibprozess angeboten. Für das Deutsche sind u.a. der *Duden-Mentor* (Premium-Funktionen: Synonyme, Stil, Vokabular) und der Schreibassistent *Language Tool*, der Grammatik-, Stil- und Rechtschreibprüfung für viele Sprachen anbietet, zu nennen. Studien zu ihrem Einsatz und positiven Ergebnissen stehen noch aus. Die Entwicklung von anspruchsvollen *Writing Assistants* ist jedoch nicht mehr wegzudenken und wird, laut Tarp et al. unsere akademische Aktivität zukünftig entscheidend beeinflussen: „High-tech tools designed to assist the writing, reading and translation of texts will be an integrated part of our lives in the years to come.“ (Tarp et al. 2017: 519).

Strategie 2: Nutzung von paradigmatischen Spezialwörterbüchern

Im Hintergrund von Textverarbeitungs- und Schreibprogrammen steht, wie oben erwähnt (vgl. Strategie 1), u.a. Information zu paradigmatischen Sinnrelationen der Lexik. Im Schreibprozess könnten daher zu Fragen der Auswahl des einen oder anderen bedeutungsähnlichen lexikalischen Elements auch paradigmatische Spezialwörterbücher, wie z.B. Synonymwörterbücher online konsultiert werden, um Zweifel zu der Bedeutung und dem Gebrauch bedeutungsähnlicher Lexeme zu klären. Zu ihrem Bekanntheitsgrad, Nutzen und ihrem adäquaten Gebrauch im Schreibprozess liegen allerdings kaum aussagekräftigen Daten vor. Wenn z.B. für die oben formulierte Aufgabe *Woxikon* zu Rate gezogen wird (Abb. 2), liegt zwar ein Angebot an bedeutungsähnlichen Wörtern vor, aber ohne weitere Informationen zu Bedeutung und Gebrauch, die die Auswahl des einen oder anderen Elements favorisiert, bieten Ressourcen dieser Art keine konkrete Hilfestellung im Schreibprozess an. Die Sprachkompetenz der Zielgruppe ist für entsprechende Disambiguierungen oft nicht ausreichend ausgebildet.

Abbildung 2: Textausschnitt und Ausschnitt aus dem Wörterbucheintrag zu *diskutieren* (aus *Woxikon*)



Strategie 3: Nutzung von syntagmatischen Spezialwörterbüchern

Um genauere Information zum Kombinationspotenzial und damit zum Gebrauch der ausgewählten Lexeme zu erhalten, können z.B. weitere Konsultationen in syntagmatischen Spezialwörterbüchern dienlich sein. Folgende Ressourcen werden exemplarisch ausgewählt, um zu erörtern, inwiefern das vorhandene Informationsangebot zur Syntagmatik beim Schreibprozess hilfreich sein kann: (i) das elektronische Valenzwörterbuch *E-VALBU* und (ii) die elektronische Version des Kollokationenwörterbuchs von Buhofer et al. (2014).

(i) *E-VALBU* bietet Information zum syntaktischen Valenzpotenzial an und dokumentiert z.B. für das Verb *diskutieren* einen Satzbauplan mit einer Alternanz zwischen einem Akkusativ- und einem Präpositivkomplement mit "über". Unterschiede im Kombinationspotenzial, wie hier z.B. im Satzbauplan bei den bedeutungsähnlichen Verben *diskutieren* und *erörtern* werden allerdings nur für die linguistisch geschulten Benutzenden sichtbar (Abb. 3).

(ii) Kollokationenwörterbücher bzw. Stilwörterbücher bieten häufigkeitsbasierte Informationen zu möglichen Kollokationspartnern an. Bei gebrauchorientierten Fragestellungen in Verbindung mit der Auswahl aus dem Angebot bedeutungsähnlicher Lexeme erweist sich die Information aber als ungenügend, wie hier an dem Beispiel zu *diskutieren* und *erörtern*⁷ aus Buhofer et al. (2014) ersichtlich wird. Zu beiden Verben, die teilweise in Verbindung mit demselben Nomen (*Angelegenheit*, *Frage*, *Lösungen* etc.) stehen (Abb. 4), wird keine ausreichende Information zu möglichen Distributionsbeschränkungen angeboten.

Abbildung 3: Gegenüberstellung der Einträge zu *erörtern* und *diskutieren* aus *E-VALBU* (Ausschnitte) mit Hervorhebungen durch die Autorin dieses Beitrages

erörtern (Lesart 1)	diskutieren <über> mit (Lesart 1)
Strukturbeispiel [bildungsspr.] jemand erörtert etwas	Strukturbeispiel jemand diskutiert etwas bzw. über etwas mit jemandem
Im Sinne von jemand analysiert und stellt etwas argumentativ dar; erläutern	Im Sinne von jemand erörtert etwas mit jemandem bzw. führt ein Gespräch über etwas, wobei unterschiedliche Standpunkte dargelegt werden
Satzbauplan K _{subj} + K _{obj}	Satzbauplan K _{subj} + K _{obj} / K _{obj} + K _{obj} + K _{obj}
Beispiele	Beispiele
(1) Eine Stunde lang sprach der Norddeutsche über demografische Entwicklung, Globalisierung und Staatsverschuldung, wobei der Politiker das Dilemma anhand von Schautafeln erörterte. (Rhein-Zeitung, 22.03.2011, S. 21)	(1) Diesen Punkt möchte ich noch mit Ihnen diskutieren.
(2) Um 19 Uhr wird Marlene Leuz aus Dietrich unter dem Titel "Kefir, Quark & Co." die Vielfalt der Milchprodukte erörtern. (Rhein-Zeitung, 01.02.2011, S. 19)	(2) Der Verein will nun seine Vorschläge mit den Landtagsparteien diskutieren. (Mannheimer Morgen, 21.07.2012, S. 6)
(3) Der Juryvorsitzende erörterte die Stärken und Schwächen des Vereins bis ins kleinste Detail. (Rhein-Zeitung, 28.04.2011, S. 24)	(3) Thomas ist ein leidenschaftlicher Diskutierer, mit dem man über alles diskutieren kann: über Kanzlerin Merkel, den Euro, das Ozonloch, die Fußball-WM, die Jugendkriminalität, die Rechtschreibreform und expressionistische Lyrik.

Abbildung 4: Kollokationenwörterbuch für den Alltag (elektronische Version) (Ausschnitte): Gegenüberstellung der Einträge zu *diskutieren* und *erörtern* mit Hervorhebungen durch die Autorin dieses Beitrages

diskutieren	erörtern (Stichwort nicht vollständig bearbeitet)
□ weitere Verbindungen (alphabetisch)	NOMEN
NOMEN	eine Angelegenheit erörtern ge • eine Frage
eine Angelegenheit diskutieren ge •	erörtern ge • eine Lage erörtern ge • eine Lösung
eine Frage (ausführlich) diskutieren ge •	erörtern ge • einen Punkt erörtern ge • ein Thema
eine Lösung diskutieren Probleme und Lösungen	erörtern ty
werden ausführlich diskutiert. ge •	
ein Modell diskutieren Wochenlang schon	
diskutieren sie die abenteuerlichsten Modelle zur	
Gesundheitsreform. ge •	
(verschiedene) Möglichkeiten diskutieren ge •	
ein Problem diskutieren ge •	
über den Sinn v. etw. diskutieren ge •	
Studenten/Studentinnen diskutieren Die	

Strategie 4: Nutzen von unterschiedlichen Bedeutungswörterbüchern

Um den lexikalischen Herausforderungen beim Textschreiben gerecht zu werden, können auch Vertreter unterschiedlicher ein- und zweisprachiger Bedeutungswörterbücher ausgewählt werden. Im Folgenden wird jeweils am Beispiel eines Vertreters der Nutzwert im Schreibprozess erörtert.

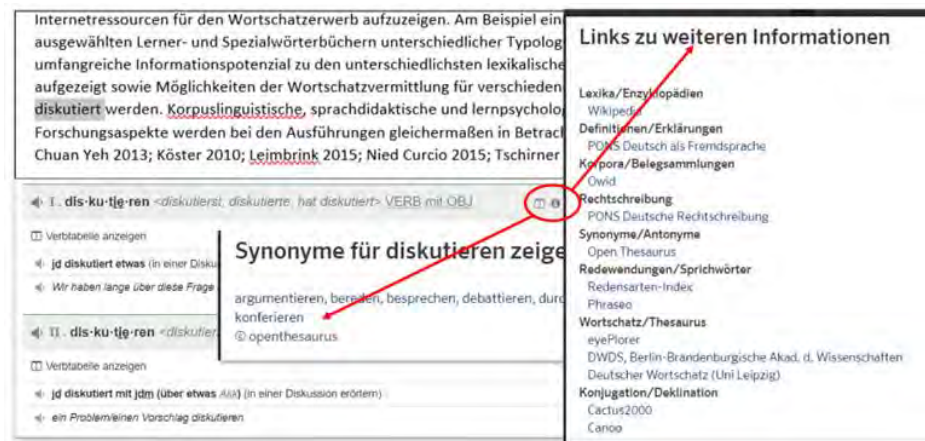
(i) Im L2-Recherchekontext führt ein Weg zu den zweisprachigen Wörterbüchern, die laut unterschiedlicher Benutzerstudien am häufigsten zur Konsultation im fremdsprachigen Lernkontext zu Rate gezogen werden (Meliss et al. 2018; Nied Curcio 2020). Das Wörterbuch für das Sprachenpaar Deutsch/Spanisch⁸ von *LEO* bietet über die syntaktische Disambiguierung verschiedene Entsprechungen für das spanische Verb *discutir* an (Abb. 5). Es ist aber ein hohes Maß an Sprachbewusstsein notwendig, um zwischen den angebotenen deutschen Entsprechungen *debattieren*, *erörtern*, *besprechen*, *sich auseinandersetzen* etc. auszuwählen und im Kontext korrekt verwenden zu können.

Abbildung 5: Wörterbucheintrag aus *Leo* Spanisch→Deutsch (Ausschnitt): *discutir* und Entsprechungen im Deutschen

Verbos	
discutir algo	etw. ^{acus} diskutieren diskutierte, diskutiert
discutir con alguien	sich. ^{acus} mit jmdm. streiten stritt, gestritten
discutir algo	etw. ^{acus} durchdiskutieren diskutierte durch, durchdiskutiert
discutir algo	etw. ^{acus} erörtern erörtere, erörtert
discutir algo	etw. ^{dat} widersprechen widersprach, widersprochen
discutir algo con alguien	etw. mit jmdm. diskutieren diskutierte, diskutiert
discutir algo - <i>alegar razones en contra</i>	etw. ^{acus} bestreiten bestritt, bestritten
discutir algo con alguien - <i>un tema, etc.</i>	etw. ^{acus} mit jmdm. besprechen besprach, besprochen - <i>Thema etc.</i>
discutir con alguien - <i>debatir</i>	sich. ^{acus} mit jmdm. auseinandersetzen setzte auseinander, auseinandergesetzt - <i>diskutieren</i>
discutirse con alguien	sich. ^{acus} mit jmdm. zoffen zoffte, gezoft
discutir por algo	über etw. ^{acus} streiten stritt, gestritten
discutir sobre algo	über etw. ^{acus} debattieren debattierte, debattiert
discutir sobre algo/alguien	gegen jmdn./etw. streiten stritt, gestritten
discutir algo a fondo	etw. ^{acus} ausdiskutieren diskutierte aus, ausdiskutiert
discutir sobre algo - <i>debatir</i>	sich. ^{acus} über etw. ^{acus} auseinandersetzen setzte auseinander, auseinandergesetzt
discutir sobre algo - <i>condiciones, etc.</i>	(über) etw. ^{acus} verhandeln verhandelte, verhandelt - <i>Bedingungen etc.</i>
discutir algo punto por punto	etw. ^{acus} durchsprechen sprach durch, durchgesprochen

(ii) Der Weg zum einsprachigen DaF-Lernerwörterbuch führt uns z.B. zu *DaF-Pons-online*. Obwohl dieser Wörterbuchtupus speziell für den DaF-Lernkontext konzipiert wurde (Runte 2015), zeigen Benutzerstudien auf, dass sie aus verschiedenen Gründen keine besonders hohe Beliebtheit genießen (Meliss et al. 2018). Der Eintrag zu *diskutieren* bietet neben den klassischen lexikografischen Angabeklassen auch Information zum Gebrauch über redaktionell überprüfte und nicht überprüfte Beispiele an. Für stilistische Fragen zur Auswahl aus einem bedeutungsähnlichen Repertoire werden die Benutzenden zu Synonymen, die aus *OpenThesaurus* stammen, geführt. Zusätzlich wird über eine Linkliste der Zugriff auf andere Ressourcen — auch zu unterschiedlichen Korpora — ermöglicht (Abb. 6).

Abbildung 6: Textausschnitt und Komposition des Informationsangebotes zu *diskutieren* aus *DaF-Pons-online* (Ausschnitt) mit Hervorhebungen durch die Autorin dieses Beitrages



Auch wenn die nötige lexikografische Kompetenz existiert, um das Informationsangebot zu finden und zu verstehen, stellt sich die Frage, ob es für den hier vorliegenden Schreibkontext (akademisches Schreiben: Zusammenfassung) wirklich nützlich ist. In Verbindung mit der besagten Aufgabenstellung wollen die Sprachlernenden z.B. wissen, welche anderen Verben für *diskutieren* verwendet werden können und sie dann auch korrekt gebrauchen. Dazu wäre konkrete und strukturierte Information bezüglich möglicher syntaktischer und stilistischer Distributionsbeschränkungen z.B. bei den Verben *diskutieren*, *erörtern*, *besprechen* und *debattieren* besonders relevant.

(iii) Der Weg führt auch zu einem einsprachigen Universalwörterbuch mit freiem Online-Zugriff. Hierfür wurde exemplarisch das *DWDS* ausgewählt. Der Eintrag zu *diskutieren* (Abb. 7) bietet neben den klassischen lexikografischen Angaben auch Information zum Gebrauch sowohl über redaktionell überprüfte als auch nicht überprüfte, maschinell ausgesuchte Beispiele an. Für stilistische Fragen zur Auswahl aus einem bedeutungsähnlichen Repertoire werden die Benutzenden zu Synonymen, die ebenfalls aus *OpenThesaurus* stammen (vgl. Strategie 4 (ii)), geführt.

Der vorher bemängelten fehlenden Informationen zu Distributionsbeschränkungen — besonders im kombinatorischen Bereich — kommt *DWDS* (teilweise) entgegen, da einige Korpora in die Ressource integriert und abrufbar sind und damit authentische Verwendungsbeispiele angeboten werden. Das *DWDS*-Wortprofil bietet automatisch generierte Information zu dem Kombinationspotenzial eines Lemmas und auch im Vergleich zu bedeutungsähnlichen, wie hier *diskutieren* und *erörtern* (Abb. 8).

Abbildung 7: Textausschnitt und Komposition des Informationsangebotes zu *diskutieren* aus DWDS (Ausschnitt) Hervorhebungen durch die Autorin dieses Beitrages

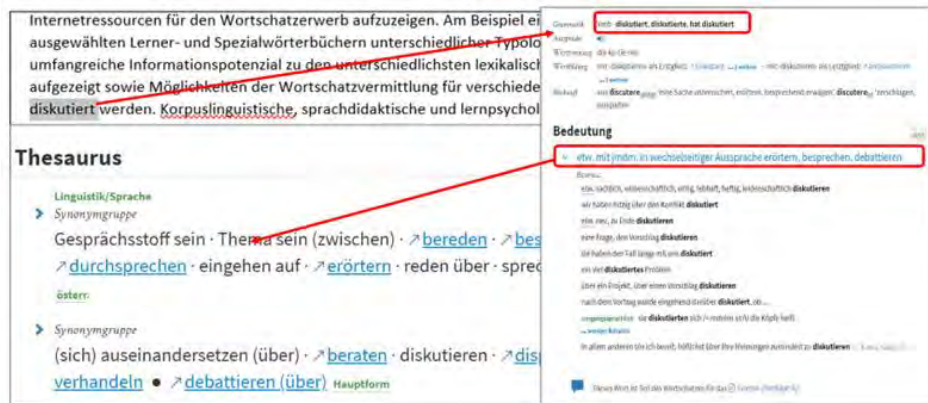
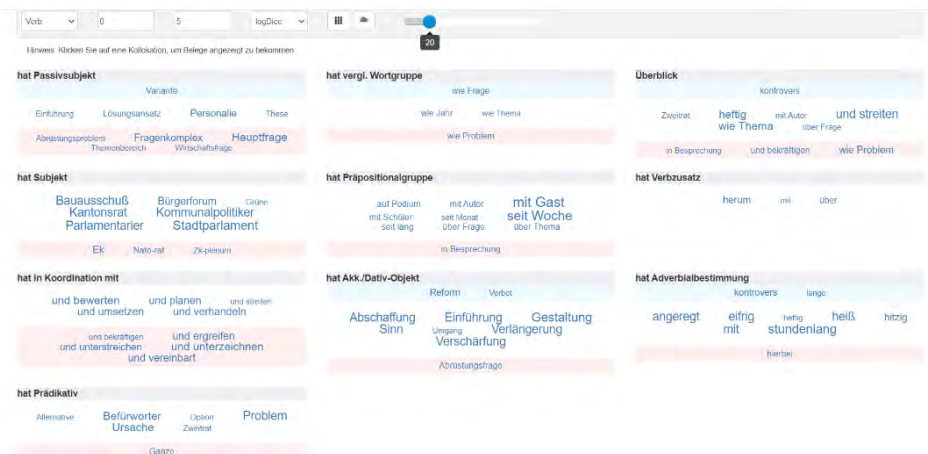


Abbildung 8: Wörterbucheintrag; aus DWDS-Wortprofil: *diskutieren* und *erörtern* im Vergleich



Es ist aber davon auszugehen, dass nur geschulte Linguisten und Linguistinnen und Germanistikstudierende diese Möglichkeiten kennen und dieses komplexe korpusbasierte Informationsangebot auch situationsadäquat nutzen können (Wallner 2013).

Strategie 5: Nutzen von Korpora im Schreibprozess

Uns stehen inzwischen viele Sprachkorpora unterschiedlichster Ausprägungen frei zur Verfügung. Nach Gouws handelt es sich um eine Korpusrevolution, die „[...] should be seen as a dynamic and ongoing process with new trends in speech technology, computer and information science constantly presenting new possibilities and innovative changes“ (2021: 5). Inkorporierte Tools ermöglichen in vielen Fällen einen intuitiven Umgang mit ihnen zu den unterschiedlichsten Zwecken (Kilgarriff et al. 2014) und erleichtern das Auffinden von Information, die der Lexikograf oder die Lexikografin eventuell nicht zur Aufnahme in das Wörterbuch berücksichtigt hat (Gouws 2021: 16). Einsprachige — oft sehr große Referenzkorpora — aber auch Lernerkorpora stehen mehrsprachigen Korpora, die kontrastiv oder parallel sein können, gegenüber. Schriftsprachliche aber auch immer mehr gesprochensprachliche und multimediale Korpora stehen ebenfalls bereit und können für ganz unterschiedliche Ziele genutzt werden. Neben konkreten Fragen bezüglich der sinnvollen Nutzung von Korpora im Schreibprozess (Conrad 2004; Yoon und Hirvela 2004; Zareva 2017; Varley 2009; Frankenberg-Garcia 2012; Granger 2015; Chen und Flowerdew 2018; Lee et al. 2019) stellen sich im DaF-Lehr- und Lernumfeld diesbezüglich auch folgende allgemeine Fragen (i) Wie verschaffe ich mir einen Überblick über die vorhandenen Korpora? (ii) Wo findet man sie? (iii) Welche Korpora sind wofür und für wen geeignet? Zwei Beispiele sollen die Problematik illustrieren, die die Korpusnutzung für den fremdsprachigen Schreibprozess impliziert.

(i) Das Portal von *Wortschatz Uni Leipzig* bietet neben korpusbasierten quantitativen Informationen zu Kookkurrenzen, Synonymen und Korpusbelegen auch über die Bedeutungsgruppen nach Dornseiff (Dornseiff und Quasthoff 2020) einen onomasiologischen Zugriff auf den Wortschatz an (Abb. 9). Gerade dieser Zugriff scheint für die Textproduktion und die Suche nach bedeutungsähnlichen Wörtern interessant. Er ermöglicht v.a. das Auffinden von Wörtern, die man noch nicht kennt — eine Möglichkeit, die bei einem semasiologischen Ansatz ausgeschlossen ist (González Ribao und Meliss 2015; Meliss und González Ribao 2016). Aber wer besitzt die sprachlichen und digitalen Kompetenzen aus der Fülle dieser Informationen konkrete anwendungsorientierte Daten herauszufiltern, die bei dem Verfassen einer Zusammenfassung dienlich sein könnten (Poole 2022)? Auch die Nutzung anderer Korpora wie die des Deutschen Referenzkorpus *DEREKO* über Werkzeuge wie *COSMAS II* oder *KorAP* stellt keine Alternative für die hier vorliegenden Fragestellungen dar. Der Zugriff auf die Daten erfolgt über eine Anmeldung und ist zu komplex für die hier relevante konkrete handlungsorientierte Situation.

Abbildung 9: Wörterbucheintrag zu *diskutieren* aus *Wortschatz Uni Leipzig* (Ausschnitt)

(ii) Benutzerfreundlicher scheint die Nutzung von Parallelkorpora zu sein. So ermöglicht z.B. das Portal *Linguee* die Konsultation von Übersetzungsentsprechungen durch Korpusbelege (Abb. 10). Gebrauchsinformation kann auf diese Weise intuitiv abgeleitet werden. Verfügen die Nutzenden über das entsprechende intuitive und kontrastive Sprachwissen, um dieses bei Schreibprozessen nutzbringend anzuwenden? Entsprechende aktuelle Benutzungsstudien wären hier sehr aufschlussreich, sind aber in jedem Fall direkt verbunden mit der vorhandenen linguistischen Kompetenz.

Abbildung 10: Eintrag zu *discutir* aus *Linguee* Spanisch→Deutsch (Ausschnitt)

[...] un producto nuevo, analizar sus perspectivas únicas, y discutir sus inquietudes en cuanto a la estrategia usada para fijar [...]	[...] treffen können, die sich in starken Verkaufszahlen und Produktannahme niederschlagen werden, wenn sie die verschiedenen Interessenvertreter [...]
↳ globalregulatorypress.com	↳ globalregulatorypress.com
2. En su país, ¿con qué frecuencia suelen los jóvenes discutir sus pensamientos o ideas con sus padres?	2. Wie häufig sprechen Jugendliche in Deutschland mit ihren Eltern über ihre Gedanken und neuen Ideen?
↳ zpjd-psychologie.de	↳ zpjd-psychologie.de
Pedir ponerse en contacto con su empresa para discutir el tema y la línea a seguir.	Bitte um Kontaktaufnahme mit Ihrem Unternehmen, um das Problem und die einzuschlagende Linie zu besprechen .
↳ riester.eu	↳ riester.eu
[...] la difusión de material racista y antisemita en Internet y discutir medidas concebibles para contrarrestar este fenómeno	[...] und antisemitischen Materials über das Internet zu beschaffen und vorstellbare Maßnahmen gegen dieses Phänomen zu erörtern
↳ oscepa.org	↳ oscepa.org
[...] desde Alemania, Gran Bretaña y Canadá, tuvo como objetivo, discutir con autoridades políticas acerca de las posibilidades que [...]	[...] Ziel, mit politischen Autoritäten die Möglichkeiten zu diskutieren, welche das Abkommen von Kyoto für die Region eröffnet.
↳ giz-cepal.cl	↳ giz-cepal.cl

3.2 Zwischenfazit und zwei weitere Strategien

Durch die Darstellung der oben aufgeführten Strategien zur Aufgabenbewälti-

gung konnte aufgezeigt werden, dass eine Vielzahl von Kompetenzen, Kenntnissen und Fähigkeiten für die Recherche und Informationsselektion vorhanden sein muss und dass in Worten von Bothma und Gouws die „[...] traditional notion of a dictionary culture has to be substituted by a more general information literacy culture which includes dictionary literacy“ (2020: 53). Auf der Grundlage der oben aufgezeigten Beobachtungen der fünf genutzten Strategien lassen sich zwei Hauptprobleme nennen, die erstens mit der Kenntnis der Ressourcen an sich und zweitens mit dem notwendigen Sprachbewusstsein und dem damit verbunden Nutzwert in Verbindung stehen.

(i) Wer verfolgt die aufgezeigten Strategien aus Kenntnis der vorgestellten Ressourcen und hat die erforderlichen sprachlichen, linguistischen und digitalen Kompetenzen, um das Informationsangebot gewinnbringend zu nutzen? Sind es nicht hauptsächlich Informationsangebote, die linguistischen Interessen dienen aber nicht dem konkreten Sprachgebrauch im DaF-Bereich? Viele DaF-Lernende aus dem universitären Bereich sind eher technischen Studiengängen zuzuordnen und nicht linguistisch interessierte Germanistikstudierende. Die meisten der Ressourcen sind ihnen gar nicht bekannt und/oder sie verfügen nicht über die linguistischen Kompetenzen und das erforderliche Sprachbewusstsein (Nied Curcio 2020) zur adäquaten Nutzung. Auch die DaF-Lehrenden sind diesbezüglich nicht immer ausreichend ausgebildet (Meliss et al. 2018).

(ii) Sind diese genannten Ressourcen für die Textproduktion wirklich adäquat? Trotz möglicher Kenntnis und Ausprägung der nötigen digitalen und linguistischen Kompetenz werden sie scheinbar für den Schreibprozess kaum bis gar nicht konsultiert, denn dafür sind u.a. zu viele Suchpfade notwendig. Außerdem muss oft die Arbeitsumgebung verlassen werden und dazu kommt auch, dass das Informationsangebot viel zu komplex ist und u.a. linguistisches Wissen voraussetzt. Ausgehend von diesen Feststellungen und den Ergebnissen erster Erhebungen bezüglich des Nutzens von Online-Ressourcen im Schreibprozess (Wolfer et al. 2018; Wolfer und Müller-Spitzer 2022; Nied Curcio et al. 2020) ist es sinnvoll, Zugriffe auf andere Ressourcen zu untersuchen, die vielleicht attraktiver und erfolgreicher für den fremdsprachigen Schreibprozess im virtuellen Raum sind. Kommen wir daher zurück zu den unterschiedlichen Apps der neusten KI-Generation (Pokrivcakova 2019; Strasser 2020; Renz et al. 2020; Hartmann 2021) und kommentieren die Strategien 6 und 7.

Strategie 6: Nutzen der Übersetzungsprogramme der neusten KI-Generation

Schreiben die Studierenden vielleicht die Zusammenfassung der oben gestellten Aufgabe in ihrer Muttersprache und lassen den Text dann von den inzwischen schon sehr guten Übersetzungsprogrammen übersetzen? Welche Kompetenzen werden gebraucht, um aus dem automatisch generierten übersetzten Text einen kontextadäquaten Text zu erstellen?⁹ Inzwischen existieren automatische Übersetzungsmaschinen, die — je nach Ausgangs- und Zielsprache — erstaun-

lich gute und korrekte Übersetzungen erstellen. Ihre Verwendung ist aus dem Lernkontext nicht mehr wegzudenken und es macht aus der Perspektive der Lehrenden keinen Sinn, ihre Existenz zu ignorieren (Lee 2021; Delorme Benites und Lehr 2021; Delorme Benites et al. 2021; Klimova et al. 2022). Zur Erstellung eines kontextadäquaten Texts in deutscher Sprache lässt die/der Lernende z.B. seinen in der Muttersprache formulierten Text von *DeepL* ins Deutsche als Zielsprache übersetzen. Voraussetzung dafür ist nur, dass die Ausgangssprache eine der Sprachen ist, die das Tool anbietet. Für den Zieltext gibt es für jedes lexikalische Element Vorschläge, die u.a. eine stilistische Anpassung an den anvisierten Kontext ermöglichen. Dies ist ein attraktives immer häufiger genutztes Vorgehen, durch das schnell und einfach Ergebnisse erzielt werden. Absolut erforderlich ist es aber, der Generation Z das Werkzeug für einen sinnvollen Umgang an die Hand zu geben, um die entsprechenden sprachlichen Kompetenzen zu erlangen, die die angeleitete stilistische Textkorrektur und das Redigieren eines automatisch übersetzten Texts betreffen. So erhält man z.B. mit *DeepL* für den Satz „El texto presenta y discute diferentes enfoques“ im Deutschen folgendes Übersetzungsangebot: „Im Text werden verschiedene Ansätze vorgestellt und diskutiert“. Für das Verb „diskutieren“ erhält man Alternativen (u.a. *erörtern*, *besprechen*, *erläutern*, *behandeln*, *kommentieren*) aus denen nur ausgewählt werden kann, wenn die entsprechende Sprachkompetenz vorhanden ist.

Strategie 7: Nutzen von Tools zur automatischen Textgenerierung

Welche Auswirkungen hat die automatische Textgenerierung auf den Spracherwerb und das Handeln mit Sprache in naher Zukunft? Es gibt zur Erstellung von Zusammenfassungen bereits verschiedene Apps, die diese Aufgabe (fast schon) automatisch erledigen. Warum sollten unsere DaF-Lernenden nicht auch diese Bequemlichkeiten — zumindest in naher Zukunft — wenn sie dann auch für das Deutsche gut funktionieren, nutzen? Die erforderlichen linguistischen Kompetenzen liegen dann u.a. auch wieder im Bereich der Stilistik und des Registers: Auswahl von kontextadäquater Lexik, die der Textsorte und dem Thema gerecht wird. Ebenso wie bei der automatischen/maschinellen Übersetzung (vgl. Strategie 6), sind für diese Strategie Kompetenzen im Bereich der Textüberarbeitung, d.h. des Redigierens notwendig. Die Diskussion um neue Herausforderungen und pädagogische Umorientierung im akademischen Umfeld steckt diesbezüglich noch in den Kinderschuhen (Salden und Weißels 2021; Limburg et al. 2022; Wilder et al. 2022; Blogeintrag von Henning 2022).

4. Fazit und Ausblick

In Verbindung mit dem Wortschatz als integrativer Komponente der Sprache existieren im virtuellen Raum verschiedene methodologische Zugänge für dessen Erwerb und handlungsorientierter Anwendung. In jedem Fall sind

unterschiedliche sprachliche und außersprachliche Kompetenzen im Umgang mit Wortschatz im virtuellen Raum erforderlich. Dabei konnte eine Diskrepanz zwischen Informationsangebot von Online-Ressourcen und Apps und ihrem realen Nutzen in zielorientierten Anwendungssituationen sichtbar gemacht werden. Die digitale Transformation und neue Entwicklungen im Bereich der KI-Sprachtechnologie erfordern eine Auseinandersetzung mit aktuellen Herausforderungen im Sprachenerlernen und Sprachenlehren¹⁰ im virtuellen Raum und die Überprüfung bzw. Neuformulierung der Kompetenzen für den handlungsorientierten Umgang mit Sprache im Allgemeinen und mit Wortschatz im Besonderen.

Es konnte ebenfalls auf die Tendenz aufmerksam gemacht werden, dass das Schreiben von wissenschaftlichen Texten im akademischen Bereich heutzutage immer mehr unter Einbezug von computerbasierten und digitalen Technologien stattfindet, die teilweise schon auf künstlicher Intelligenz (KI) basieren — wie intelligentes Tutoring (IT), automatisierte Schreibbewertung (AWE) und maschinelle Übersetzung (MT). Lexikografische Online-Ressourcen bilden nur einen Bruchteil des Angebotes, welches von den besagten Akteuren im virtuellen Raum als Strategie zur Aufgabenbewältigung aufgesucht wird. Diese Tatsache stellt für die Hochschulausbildung neue Chancen, Perspektiven aber auch Risiken dar. Für den Bereich des Fremdsprachenerwerbs besteht u.a. die Herausforderung darin, die Vorteile von KI-bezogenen Technologien im Sprachenlehren und -lernen und konkret beim akademischen Schreiben zielführend zu nutzen und gleichzeitig damit verbundene Probleme wie ineffiziente und missverständliche Kommunikation und Missbrauch (z.B. Plagiate) zu beseitigen. Der Aufbau von *awareness* und der Erlernen kritische Umgang sind hier die angestrebten Ziele.

Es erweist sich als unabdingbar, KI-bezogene Technologien in die Hochschullehre systematisch und sinnvoll einzuführen und somit die Hochschulen in Worten von Weßels aus dem „Dornröschenschlaf“ zu erwecken (Weßels 2020). Voraussetzung dafür ist allerdings die Förderung eines kompetenten Umgangs der Dozierenden mit diesen Technologien und die Befähigung der Studierenden dazu, mithilfe intelligenter Maschinen ihr Schreiben zu verbessern. In diesem Sinne kann z.B. das schweizer Projekt *DigLit* genannt werden, welches anstrebt, die Anwendbarkeit verschiedener Arten digitaler mehrsprachiger Schreibunterstützung an Schweizer Hochschulen zu evaluieren und das Bewusstsein aktueller und zukünftiger NutzerInnen für die damit verbundenen Chancen und Risiken zu schärfen.

Folgende Desiderata lassen sich konkret für den Wortschatzerwerb beim Schreibprozess im virtuellen Raum als digitalem Handlungsbereich zusammenfassend ableiten:

- Inter- und multidisziplinäre Forschung zur technologischen und linguistischen Verbesserung und Weiterentwicklung von Schreibtools für das Deutsche mit benutzerfreundlicher und benutzeradaptierter Inkorporation von lexikalischer Information als integrativer Komponente unserer

Sprache (Bothma und Gouws 2020)¹¹. Aus lexikografischer Perspektive wäre eine direkte Integrierung der vielschichtigen lexikalischen Information, so wie sie jetzt in den einzelnen lexikografischen Ressourcen isoliert vorliegen, in die entsprechenden Schreibtools wünschenswert.

- Entwicklung einer aktuellen Wortschatzdidaktik unter Berücksichtigung des virtuellen Raums und den diversen sprachtechnologischen Angeboten. In Verbindung damit stehen die Aktualisierung und Anpassung des GER und die Verstärkung der pragmatisch-stilistischen Aspekte (Textsorten, (mediale) Varianz etc.) (Fandrych und Thurmair 2011; Steinhoff 2011)¹².
- Entwicklung adäquater adaptiver Lehr- und Lernmaterialien zur Nutzung der Sprachressourcen und der Sprachtechnologie im virtuellen Raum
- Erweiterung der entsprechenden Kompetenzen der Lehrenden und Lernenden durch gezielte Schulungen, den Einsatz von geeigneten Lehr- und Lernmaterialien sowie Aktualisierung des bestehenden universitären Lehr- und Studienangebotes.

Aus lexikografischer Perspektive gibt es trotz der genannten Verbesserungs- und Aktualisierungswünsche nach Meinung von Tarp und Gouws genug Gründe zum Optimismus: „Modern-day lexicographers are in a position to make some of the unfulfilled dreams of the past a reality. The challenge of the future is to make the impossible possible. We have work to do“ (2019: 266).

Danksagung

An dieser Stelle möchte ich den anonymen Gutachtern/Gutachterinnen für ihre sehr wertvollen Anmerkungen danken.

Endnoten

1. Siehe dazu auch den Plenarvortrag von M. Glabionat „Digitale Transformation — Chancen, Herausforderungen und Grenzen des digitalen Sprachenlernens, -lehrens und -beurteilens“. 16.08.2022 IDT Wien.
2. Siehe dazu *Multiliteracies* Plenarvortrag von K. Biebighäuser „Kritische Mediennutzung und Sprachenlernen“, 16.08.2022 IDT Wien.
3. Der Beitrag steht in direkter Verbindung mit dem Forschungsprojekt COMBIDIGILEX. Eins der Projektziele fokussiert die Anwendung von unterschiedlichen Forschungsergebnissen, die aus kontrastiven korpusbasierten Studien zu verbalen Argumentstrukturmustern resultieren. Diese Ergebnisse münden in die Entwicklung der Ressource CombiDigiLex, einem onomasiologisch ausgerichteten, digitalen, multilingualen lexiko-grammatischen Informationssystem für L2-Produktionssituationen, welches sich z.Z. noch im Aufbau befindet (abrufbar unter: <https://combidigilex.usc.gal/index.php#>) (vgl. Fernández Méndez et al. 2022).
4. Siehe dazu auch den Plenarvortrag von L. Konstantinidou „Schriftbasiertes Handeln: Förderung der Schreibkompetenz in der Fremd- und Zweitsprache Deutsch“, 19.08.2022 IDT Wien.

5. Dieser Artikel ist eine erweiterte und aktualisierte Version des Vortrages: "Das Zusammenspiel von Wortschatzerwerb und lexikografischer Kompetenz im virtuellen Raum: anwendungsorientierte Beispiele für den DaF-Bereich", den die Autorin auf der XVII. *Internationalen Tagung der Deutschlehrerinnen und Deutschlehrer* (IDT) in Wien (15. – 20.08.2022) in der Sektion F.2 „Konzepte der Wortschatzvermittlung und mentales Lexikon“ virtuell vorgetragen hat.
6. Siehe dazu auch den Plenarvortrag von A. Ender „Wortschatz: Was lexikalische Assoziationen über Erwerb, Organisation und Vermittlung aussagen können?“, 16.08.2022 IDT Wien.
7. Die redaktionelle Arbeit des Eintrages zu *erörtern* ist nicht abgeschlossen. Für die hier verfolgte Argumentationslinie ist die vorliegende Information durchaus ausreichend.
8. Für die zwei- und mehrsprachigen Ressourcen wird als Ausgangssprache das Spanische gewählt, da sich der Forschungs- und Lehrkontext der Autorin des Beitrages im Deutsch-Spanischen Kontext bewegt.
9. Siehe dazu auch die aktuellen Kongressbeiträge von Ehrensberger-Dow (2022) und Delorme Benites et al. (2022).
10. Siehe dazu auch den Plenarvortrag von M. Glabionat (2022) „Digitale Transformation — Chancen, Herausforderungen und Grenzen des digitalen Sprachenlernens, -lehrens und -beurteilens“, 16.08.2022 IDT Wien.
11. Siehe dazu auch Podium Plus: „Künstliche Intelligenz und DaF/DaZ-Lehren und Lernen: Wo stehen wir und wohin geht der Weg?“ unter der Leitung von N. Würfel und K. Han, 16.08.2022 IDT Wien.
12. Siehe dazu auch den Plenarvortrag von Chr. Fandrych „Konzepte der Grammatikvermittlung (auch) im Kontext digitaler Kommunikationsformate“, 18.08.2022 IDT Wien.

(Elektronische) Wörterbücher und andere Ressourcen

(letzter Zugriff: 10.04.2023)

DaF-Pons-online: <https://de.pons.com/>

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

DigLit: Digital Literacy im Hochschulkontext: Projektleiterin: Liana Konstantinidou.

<https://www.zhaw.ch/de/linguistik/digital-literacy-im-hochschulkontext-diglit/#c165852>

Dornseiff, F. und U. Quasthoff (Hrsg.). 2020. *Der deutsche Wortschatz nach Sachgruppen*. (9. überarbeitete und erweiterte Auflage). Berlin/Boston: De Gruyter.

Duden-Mentor: <https://mentor.duden.de/>

DWDS: Digitales Wörterbuch der Deutschen Sprache:

<https://www.dwds.de/>

E-VALBU: Elektronisches Valenzwörterbuch deutscher Verben:

<https://grammis.ids-mannheim.de/verbvalenz>

Language Tool: <https://languagetool.org/de>

LEO: <https://dict.leo.org/alem%C3%A1n-espa%C3%B1ol/discutir?side=left>

Linguee: <https://www.linguee.de/deutsch-spanisch>

OpenThesaurus: <https://www.openththesaurus.de/>

Wortschatz Uni Leipzig: <https://wortschatz.uni-leipzig.de/de>

Woxikon: https://synonyme.woxikon.de/?_ga=2.212549590.393933964.1669721238-2016084285.1669721238

Literaturhinweise

(letzter Zugriff auf die elektronischen Materialien: 10.04.2023)

- Aguado, K. und D. Warneke.** 2021. Aufbau und Vernetzung lexikalischer und grammatischer Sprachkompetenzen: Aktuelle Entwicklungen in der Fachdiskussion und Schlussfolgerungen für die Unterrichtspraxis. *Fremdsprache Deutsch* 64: Wortschatz: 3-12.
DOI: 10.37307/j.2194-1823.2021.64.03
- Altmayer, C., K. Biebighäuser, St. Haberzettl und A. Heine (Hrsg.).** 2021. *Handbuch Deutsch als Fremd- und Zweitsprache. Kontexte — Themen — Methoden*. Stuttgart: Metzler.
DOI: 10.1007/978-3-476-04858-5
- Andresen, M. und D. Knorr.** 2017. KoLaS — Ein Lernendenkorpus in der Schreibberatungsbildung einsetzen. *Zeitschrift Schreiben*: 10-17.
http://fox.leuphana.de/portal/files/14261826/andresen_knorr_lernendenkorpus_kolas2_1.pdf
- Biebighäuser, K. (Hrsg.).** 2022. Sondernummer zum Thema "Virtueller DaF-Unterricht". *German as a Foreign Language* 1: 1-17.
<http://www.gfl-journal.de/>
- Biebighäuser, K. und D. Feick (Hrsg.).** 2020. *Digitale Medien in Deutsch als Fremd- und Zweitsprache* [Studien Deutsch als Fremd- und Zweitsprache 8]. Berlin: Erich Schmidt Verlag.
- Biebighäuser, K., S. Falk, D. Feick und M. Schart (Hrsg.).** 2021. DaF-Unterricht im virtuellen Unterrichtsraum — Zur Einführung in das Themenheft. *InfoDaF* 48(5): 443-463.
<https://www.degruyter.com/journal/key/infodaf/48/5/html>
- Bothma, T.J.D. und R.H. Gouws.** 2020. e-Dictionaries in a Network of Information Tools in the e-Environment. *Lexikos* 30: 29-56
DOI: 10.5788/30-1-1588
- Braun, A. und P. Klimaszyk (Hrsg.).** 2022. Aktivierung und Interaktion durch digitale Medien. *Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschunterrichts* 66: 3-10.
DOI: 10.37307/j.2194-1823.2022.66
- Buhofer, H., M. Dräger, St. Meier und T. Roth.** 2014. *Feste Wortverbindungen des Deutschen. Kollokationswörterbuch für den Alltag*. Tübingen: Francke.
<https://kollokationenwoerterbuch.ch/web/>
- Chen, M. und J. Flowerdew.** 2018. A Critical Review of Research and Practice in Data-driven Learning (DDL) in the Academic Writing Classroom. *International Journal of Corpus Linguistics* 23(3): 335-369.
DOI: 10.1075/ijcl.16130.che
- Conrad, S.** 2004. Corpus Linguistics, Language Variation, and Language Teaching. Sinclair, J.M. (Hrsg.). 2004. *How to Use Corpora in Language Teaching*: 67-86. Amsterdam: John Benjamins.
- Delorme Benites, A. und C. Lehr.** 2021. Neural Machine Translation and Language Teaching: Possible Implications for the CEFR. *Bulletin suisse de linguistique appliquée* 114: 47-66.
- Delorme Benites, A., S. Cotelli Kureth, C. Lehr und E. Steele.** 2021. Machine Translation Literacy: A Panorama of Practices at Swiss Universities and Implications for Language Teaching. Zoghلامي, N., C. Brudermann, C. Sarré, M. Grosbois, L. Bradley und S. Thouësny (Hrsg.). 2021. *CALL and Professionalisation: Short Papers from EUROCALL 2021*: 80-87. Reserche-Publishing.net.
DOI: 10.14705/rpnet.2021.54.1313

- Delorme Benites, A., S. Cotelli Kureth, C. Lehr, E. Steele and L. Franciello.** 2022. Kongressbeitrag "I looked it up in DeepL: Online Dictionaries and Online Machine Translation". *Tralogy3: Tralogy: Human Translation and Natural Language Processing: Forging a New Consensus?* Paris, France, 7–8 April 2022.
<https://tralogy3.sciencesconf.org/>
- Dengscherz, S.E.** 2018. Wie schreibt man anspruchsvolle Texte in einer L2? Strategien und Routinen von erfolgreichen Schreiber*innen. *ÖDaF-Mitteilungen* 34(2): 9-21.
DOI: 10.14220/odaf.2018.34.2.9
- Dengscherz, S.E.** 2020. Professionelles Schreiben in mehreren Sprachen. Das PROSIMS-Schreibprozessmodell. *ZIF Zeitschrift für interkulturellen Fremdsprachenunterricht* 25(1): 397-422.
<https://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/view/1023/1020>
- Ehlich, K.** 2000. Deutsch als Wissenschaftssprache für das 21. Jahrhundert. *German as a Foreign Language (GFL)* 1: 47-63.
- Ehlich, K.** 2003. Universitäre Textarten, universitäre Struktur. Ehlich, K. und A. Steets (Hrsg.). 2003. *Wissenschaftlich schreiben — lehren und lernen*: 13-28. Berlin: De Gruyter.
- Ehlich, K.** 2018. Wissenschaftlich schreiben lernen — von diskursiver Mündlichkeit zu textueller Schriftlichkeit. Schmölzer-Eibinger, S., B. Bushati, Ch. Ebner und L. Niederdorfer (Hrsg.). 2018. *Wissenschaftliches Schreiben lehren und lernen. Diagnose und Förderung wissenschaftlicher Textkompetenz in Schule und Universität*: 15-32. Münster/New York: Waxmann.
- Ehrensberger-Dow, M.** 2022. Kongressbeitrag "Exploring the Role of MT in Science News Flows". *Abstract Booklet, Panel 17 (Interlingual and Intralingual Translation), 10th Congress of the European Society for Translation Studies (EST 2022, Oslo, Norway, 22.-25. Juni 2022)*: 245-246.
<https://www.hf.uio.no/ilos/english/research/news-and-events/events/conferences/2022/est22/program/est22-oslo-abstract-booklet/est22-abstract-booklet.pdf>
- Fandrych, Ch. und M. Thurmair.** 2011. Plädoyer für eine textsortenbezogene Sprachdidaktik. *Deutsch als Fremdsprache* 48(2): 84-93.
- Fernández Méndez, M., I. Mas Álvarez und M. Meliss.** 2022. Herausforderungen bei der Erstellung der multilingualen, korpusbasierten lexikografischen Ressource CombiDigiLex. *TEISEL: Tecnologías para la investigación en segundas lenguas* 1: 1-26.
DOI: 10.1344/teisel.v1.36590
- Frankenberg-Garcia, A.** 2012. Raising Teachers' Awareness of Corpora. *Language Teaching* 45(4): 475-489.
DOI: 10.1017/S0261444810000480
- GER: Deutsche Online-Version des GER:
<https://www.goethe.de/z/50/commeuro/i3.htm>
- Glaboniat, M., M. Müller, P. Rusch, H. Schmitz und L. Wertenschlag.** 2002. *Profile deutsch. Gemeinsamer europäischer Referenzrahmen*. Berlin/München: Langenscheidt.
- González Ribao, V. und M. Meliss.** 2015. Theoretische und methodologische Grundlagen eines onomasiologisch-konzeptuell orientierten Produktionswörterbuches für den zweisprachigen Lernerkontext: Deutsch–Spanisch. Calañas Continente, J.-A. und F. Robles i Sabater (Hrsg.). 2015. *Die Wörterbücher des Deutschen: Entwicklungen und neue Perspektiven* [Spanische Akzente. Studien zur Linguistik des Deutschen 2]: 109-136. Frankfurt: Peter Lang Edition.

- Gouws, R.H.** 2021. Expanding the Use of Corpora in the Lexicographic Process of Online Dictionaries. Piosik, M., J. Taborek und M. Woźnicka (Hrsg.). 2021. *Korpora in der Lexikographie und Phraseologie: Stand und Perspektiven*: 1-20. Berlin/Boston: De Gruyter.
DOI: 10.1515/9783110716955-001
- Graefen, G.** 2014. Schwer zu begreifen, was ein Begriff ist. Die „Alltägliche Wissenschaftssprache“ im Studienverlauf. Boócz-Barna, K., P. Kertes, B. Palotás, G. Perge und A. Reder (Hrsg.). 2014. *Aktuelle Fragen der Fremdsprachendidaktik. Festschrift für Ilona Feld-Knapp. Sonderheft der Zeitschrift Deutschunterricht für Ungarn* 26: 61-73.
- Graefen, G. und M. Moll.** 2011. *Wissenschaftssprache Deutsch: lesen — verstehen — schreiben. Ein Lehr- und Arbeitsbuch*. Frankfurt am Main: Peter Lang.
- Granger, S.** 2015. The Contribution of Learner Corpora to Reference and Instructional Materials Design. Granger, S., G. Gilquin und F. Meunier (Hrsg.). 2015. *The Cambridge Handbook of Learner Corpus Research*: 486-510. Cambridge: Cambridge University Press.
- Hartmann, D.** 2021. Künstliche Intelligenz im DaF-Unterricht? Disruptive Technologien als Herausforderung und Chance. *InfoDaF Informationen Deutsch als Fremdsprache* 48(6): 683-696.
DOI: <https://doi.org/10.1515/infodaf-2021-0078>
- Henning, U.** 2022. Blogbeitrag „Herausforderung KI-Text-Tools“ (01. Mai, 2022).
<https://web2-unterricht.ch/deutschunterricht/herausforderung-ki-text-tools/>
- Henriksen, B.** 1999. Three Dimensions of Vocabulary Development. *Studies in Second Language Acquisition* 21(2). Special Issue: “Incidental L2 Vocabulary Acquisition: Theory, Current Research, and Instructional Implications”: 303-317.
- Jaworska, S.** 2011. Der Wissenschaftlichkeit auf der Spur. Zum Einsatz von Korpora in der Vermittlung des Deutschen als (fremder) Wissenschaftssprache. *Deutsch als Fremdsprache* 4: 233-242.
- Kilgariff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý and V. Suchomel.** 2014. The Sketch Engine: Ten Years On. *Lexicography* 1: 7-36.
- Klimova, B., M. Pikhart, A. Delorme Benites, C. Lehr und Ch. Sanchez-Stockhammer.** 2022. Neural Machine Translation in Foreign Language Teaching and Learning: A Systematic Review. *Education and Information Technologies* 28(1): 663-682.
DOI: 10.1007/s10639-022-11194-2
- Kretzenbacher, H.L.** 1998. Fachsprache als Wissenschaftssprache. Hoffmann, L., H. Kalverkämper und H.E. Wiegand (Hrsg.). 1998. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Vol. 1*: 133-142. Berlin/New York: De Gruyter.
- Kühn, P.** 2007. Rezeptive und produktive Wortschatzkompetenzen. Willenberg, H. (Hrsg.). 2007. *Kompetenzhandbuch für den Deutschunterricht: auf der empirischen Basis des DESI-Projekts*: 159-167. Baltmannsweiler: Hohengehren.
- Küster, L. (Koord.).** 2014. Themenschwerpunkt: Multilateralität. *FLUIL — Fremdsprachen Lehren und Lernen* 43(2).
- Lee, H., M. Warschauer and J. Ho Lee.** 2019. The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis. *Applied Linguistics* 40(5): 721-753.
DOI 10.1093/applin/amy012
- Lee, S.M.** 2021. The Effectiveness of Machine Translation in Foreign Language Education: A Systematic Review and Meta-analysis. *Computer Assisted Language Learning*: 1-23.
DOI: 10.1080/09588221.2021.1901745

- Leimbrink, K.** 2015. Wortschatzerwerb. Haß, U. und P. Storjohann (Hrsg.). 2015. *Handbuch Wort und Wortschatz* (Handbücher Sprachwissen 3): 27-52. Berlin/München/Boston: De Gruyter.
DOI: 10.1515/9783110296013-002
- Limburg, A., M. Lucht, M. Mundorf, P. Salden und D. Weßels.** 2022. Künstliche Intelligenz in Schreibdidaktik und -beratung: Quo vadis? *JoSch — Journal für Schreibwissenschaft* 23(1): 53-67.
DOI: 10.3278/JOS2201W
- Lotze, N.** 2018. Künstliche Intelligenz fürs Sprachenlernen. *Magazin Sprache*. Goethe-Institut.
<https://www.goethe.de/ins/in/de/spr/mag/21290629.htm> [zuletzt abgerufen am 15.01.2023]
- Lüdeling, A. und M. Walter.** 2009. Korpuslinguistik für Deutsch als Fremdsprache. *Sprachvermittlung und Spracherwerbsforschung*: 1-37.
<https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiterinnen/anke/pdf/LuedelingWalterDaF.pdf>
- Meissner, C.** 2021. Digitale linguistische Ressourcen für die mehrsprachigkeitsorientierte Vermittlung von Bildungs- und Wissenschaftssprache nutzen. Möglichkeiten und Herausforderungen. *Bulletin suisse de linguistique appliquée* 1. No. spécial "La linguistique appliquée à l'ère digitale": 149-167.
https://www.vals-asla.ch/fileadmin/user_upload/Journal/Special_2021_vol_1_def_A4.pdf
- Meliss, M. und V. González Ribao.** 2016. Wortprofile, Kookkurrenzen und Konkordanzen in Forschung und Lehre im hispanophonen DaF-Umfeld: Bestandsaufnahme und -beschreibung für kontrastive Aufgabenstellungen. Robles i Sabater, F., D. Reimann und R. Sánchez Prieto (Hrsg.). 2016. *Sprachdidaktik Spanisch-Deutsch. Forschungen an der Schnittstelle von Linguistik und Fremdsprachendidaktik* [Romanistische Fremdsprachenforschung und Unterrichtsentwicklung 6]: 93-112. Tübingen: Narr.
- Meliss, M., M. Egidio Vicente und M. Fernández Méndez.** 2018. Plädoyer für die Entwicklung einer digital-lexikografischen Kompetenz im Fremdsprachenunterricht. *Lexicographica* 34: 123-156.
DOI: 10.1515/lex-2018-340107
- Müller-Spitzer, C., A. Koplenig und S. Wolfer.** 2018. Dictionary Usage Research in the Internet Era. Fuertes-Olivera, P.A. (Hrsg.). 2018. *The Routledge Handbook of Lexicography*: 715-734. London/New York: Routledge.
- Nardi, A. und C. Farroni.** 2022. Wissenschaftssprachliche Kompetenz beim Schreiben in Deutsch als Fremde Wissenschaftssprache. Eine korpusbasierte Untersuchung. *KorDaF — Korpora Deutsch als Fremdsprache* 2(1): 61-80.
DOI: 10.48694/kordaf-3488
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
DOI: 10.1017/CBO9781139524759
- Nied Curcio, M.** 2014. Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht. Abel, A., Ch. Vettori und N. Ralli (Hrsg.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus*, 15–19 July 2014, Bozano/Bozen, Italy: 263-280. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202014/euralex_2014_018_p_263.pdf
- Nied Curcio, M.** 2020. Sprachbewusstheit als wichtige Voraussetzung bei der Recherche in mehrsprachigen Online-Ressourcen. Hepp, M. und K. Salzmann (Hrsg.). 2020. *Sprachvergleich in der mehrsprachig orientierten DaF-Didaktik. Theorie und Praxis*: 85-109. Rome: Istituto Italiano di Studi Germanici.

- Nied Curcio, M.** 2022. Dictionaries, Foreign Language Learners and Teachers. New Challenges in the Digital Era. Klosa-Kückelhaus, A., St. Engelberg, Ch. Möhrs und P. Storjohann (Hrsg.). 2022. *Dictionaries and Society. Proceedings of the XX Euralex International Congress*, 12–16 July 2022, Mannheim, Germany: 71-84. Mannheim: IDS Verlag.
https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202022/EURALEX2022_Pr_p71-84_Nied-Curcio.pdf
- Nied Curcio, M., C. Müller-Spitzer, I.M. Silva Días, M.J. Domínguez Vázquez und S. Wolfer.** 2020. Welche Online-Ressourcen nutzen romanischsprachige DaF-Lernende bei der Verbesserung deutscher Texte? Wulff, N., S. Steinmetz, D. Strömsdörfer und M. Willmann (Hrsg.). 2020. *Deutsch weltweit — Grenzüberschreitende Perspektiven auf die Schnittstellen von Forschung und Vermittlung. 45. Jahrestagung des Fachverbandes Deutsch als Fremd- und Zweitsprache an der Universität Mannheim 2018*. (Materialien Deutsch als Fremdsprache 104): 185-206. Göttingen: Universitätsverlag.
DOI: 10.17875/gup2020-1359
- Perkuhn, R.** 2021. Korpusfrequenzen und andere Metriken zur Strukturierung von DaF-Lehrmaterial. *KorDaF — Korpora Deutsch als Fremdsprache* 1(2): 116-136.
DOI: 10.48694/tujournals-86
- Pokrivcakova, S.** 2019. Preparing Teachers for the Application of AI-powered Technologies in Foreign Language Education. *Journal of Language and Cultural Education* 7(3): 135-153.
DOI: 10.2478/jolace-2019-0025
- Poole, R.** 2022. Corpus Can be Tricky: Revisiting Teacher Attitudes towards Corpus-aided Language Learning and Teaching. *Computer Assisted Language Learning* 35(7): 1620-1641.
DOI: 10.1080/09588221.2020.1825095
- Pressemitteilung ARD/ZDF-Onlinestudie 2020: Zahl der Internetnutzer wächst um 3,5 Millionen / Deutlicher Zuwachs für Mediatheken von ARD und ZDF.*
https://www.ard-zdf-onlinestudie.de/files/2020/Pressemitteilung_ARD_ZDF_Onlinestudie_2020.pdf
- Renz, A., S. Krishnaraja and E. Gronau.** 2020. Demystification of Artificial Intelligence in Education. How much AI is Really in the Educational Technology? *International Journal of Learning Analytics and Artificial Intelligence for Education* 2(1): 14-30.
DOI: 10.3991/ijai.v2i1.12675
- Runte, M.** 2015. *Lernerlexikographie und Wortschatzerwerb*. Berlin/München/Boston: De Gruyter.
DOI: 10.1515/9783110428476
- Salden, P. und D. Weßels.** 2021. Künstlich intelligentes Schreiben. *DUIZ — Magazin — Wissenschaft & Management* 11: 45-48.
<https://www.duz.de/ausgabe/!/id/562>
- Schippan, Th.** 1992. *Lexikologie der deutschen Gegenwartssprache: Studienbuch*. Tübingen: Niemeyer
- Schmidt, T. und Th. Strasser.** 2022. Artificial Intelligence in Foreign Language Learning and Teaching: A CALL for Intelligent Practice. *Anglistik. International Journal of English Studies* 33(1): 165-184.
DOI: 10.33675/ANGL/2022/1/14
- Steinhoff, T.** 2011. Unterrichtsideen zur textorientierten Wortschatzarbeit: Aneignungs- und Gebrauchskontexte lexikalischer Mittel. Pohl, I. und W. Ulrich (Hrsg.). 2011. *Wortschatzarbeit*: 577-591. Baltmannsweiler: Hohengehren.

- Strank, W.** 2020. *Da fehlen mir die Worte. Systematischer Wortschatzerwerb für fortgeschrittene Lerner in Deutsch als Fremdsprache*. 2. Aufl. Leipzig: Schubert Verlag.
- Strasser, Th.** 2020. Künstliche Intelligenz im Sprachunterricht: Ein Überblick. *Revista Lengua y Cultura* 1: 1-6.
DOI: 10.29057/lc.v1i2.5533
- Tarp, S., K. Fisker und P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
DOI: 10.5788/27-1-1412
- Tarp, S. und R. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.
DOI: 10.5788/29-1-152.
- Thielmann, W.** 2017. Genuin wissenschaftssprachliche Strukturen. *InfoDaF* 44(5): 546-569.
DOI: 10.1515/infodaf-2017-0091 / [zuletzt abgerufen am 15.01.2023]
- Tschirner, E.** 2010. Wortschatz. Krumm, H.-J. et al. (Hrsg.). 2010. *Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch*: 236-245. Berlin/New York: De Gruyter Mouton.
DOI: 10.1515/9783110240245
- Varley, S.** 2009. I'll Just Look that up in the Concordancer: Integrating Corpus Consultation into the Language Learning Environment. *Computer Assisted Language Learning* 22(2): 133-152.
DOI: 10.1080/09588220902778294
- Vuorikari, R., St. Kluzer und Y. Punie.** 2022. *DigComp 2.2, The Digital Competence Framework for Citizens: With New Examples of Knowledge, Skills and Attitudes*. Luxemburg: Publications Office of the European Union.
DOI: 10.2760/115376
- Wallner, F.** 2013. Korpora im Daf-Unterricht? Potentiale und Perspektiven am Beispiel des DWDS. *Revista Nebrija de Lingüística Aplicada* 13.
<https://www.nebrija.com/revista-linguistica/korpora-im-daf-unterricht-potentiale-und-perspektiven-am-beispiel-des-dwds.html>
- Weßels, D.** 2020. Digitale Disruption und Künstliche Intelligenz — Hochschulen im Dornröschenschlaf? Friedrich-Ebert-Stiftung (Hrsg.). 2020. *Netzwerk Wissenschaft*. Berlin: Friedrich-Ebert-Stiftung, Abt. Studienförderung.
<https://www.fes.de/themenportal-bildung-arbeit-digitalisierung/artikelseite/digitale-disruption-und-kuenstliche-intelligenz-hochschulen-im-dornroeschenschlaf>
- Wilder, N., D. Weßels, J. Gröpler, A. Klein, A. und M. Mundorf.** 2022. Forschungsintegrität und Künstliche Intelligenz mit Fokus auf den wissenschaftlichen Schreibprozess — Traditionelle Werte auf dem Prüfstand für eine neue Ära. Miller, K., M. Valeva und J. Prieß-Buchheit (Hrsg.). 2022. *Verlässliche Wissenschaft — Bedingungen, Analyse, Reflexion*: 203-223. Darmstadt: wbg Academic.
https://files.wbg-wissenverbindet.de/Files/Article/ARTK_ZOA_1025976_0002.pdf
- Wolfer, S. und C. Müller-Spitzer.** 2022. Sind Wörterbücher wirklich nützliche Werkzeuge beim Überarbeiten von Texten? Ein experimenteller Zugang. Beißwenger, M., L. Lemnitzer und C. Müller-Spitzer (Hrsg.). 2022. *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*: 162-178. Paderborn: Brill/Fink.

- Wolfer, S., Th. Bartz, T. Weber, A. Abel, Ch. M. Meyer, C. Müller-Spitzer und A. Storrer.** 2018. The Effectiveness of Lexicographic Tools for Optimising Written L1-texts. *International Journal of Lexicography* 31(1): 1-28.
DOI: 10.1093/ijl/ecw038
- Wolfer, S., F. Michaelis und C. Müller-Spitzer.** 2021. Nutzungsstatistiken von linguistischen Online-Ressourcen: Potentiale und Limitationen. *Zeitschrift für Angewandte Linguistik* 75: 395-430.
DOI: 10.1515/zfal-2021-2070
- Yoon, H. und A. Hirvela.** 2004. ESL Student Attitudes toward Corpus Use in L2 Writing. *Journal of Second Language Writing* 13(4): 257-283.
DOI: 10.1016/j.jslw.2004.06.002
- Zareva, A.** 2017. Incorporating Corpus Literacy Skills into TESOL Teacher Training. *ELT Journal* 71(1): 69-79.
DOI: 10.1093/elt/ccw045

Lexical Analysis of Nautical and Marine Engineering Corpora: Similar or Different Lexicographic Results

Zorica Đurović, *Faculty of Maritime Studies,
University of Montenegro, Kotor, Montenegro*
(zoricag@ucg.ac.me)

and

Milena Dževerdanović-Pejović, *Faculty of Maritime Studies,
University of Montenegro, Kotor, Montenegro*
(milenadz@ucg.ac.me)

Abstract: English for Maritime Purposes comprises a variety of different registers. However, most of the dictionaries and glossaries are dedicated either to General Maritime English or Nautical English, while other aspects of the maritime industry are poorly covered by specialized lexicographic material, especially as regards formal dictionaries. Considering that the main classification onboard ship in terms of the crew and operational systems is related to the Deck and the Engineering Department, we sought to explore the similarities and differences between the technical lexes of the two areas, in order to see whether separated dictionaries and glossaries might be required, or a common maritime one would suffice. To test and measure the tackled technical vocabularies, we utilized contemporary corpus linguistics methods and software. The results show significant differences in the key vocabulary of the compared corpora. The findings clearly and once more call for special attention and focus when it comes to the interpretation of maritime lexis, as well as for establishing a clear distinction between English for Nautical and for Marine Engineering purposes.

Keywords: MARITIME ENGLISH, NAUTICAL ENGLISH, MARINE ENGINEERING, CORPUS, VOCABULARY, LEXIS, FREQUENCY, KEYWORDS, WORD LIST, TECHNICAL DICTIONARY

Opsomming: Die leksikale analise van korpora vir seevaart en skeepsvaart-ingenieurswese: Ooreenkomstige of uiteenlopende leksikografiese resultate.

Engels vir Maritieme Doeleindes sluit 'n verskeidenheid uiteenlopende registers in. Die meeste van die woordeboeke en glossariums word egter toegespits op óf Algemene Maritieme Engels óf Seevaartengels, terwyl ander aspekte van die maritieme bedryf onvoldoende deur gespesialiseerde leksikografiese materiaal gedek word, veral deur formele woordeboeke. Aangesien die hoofklassifikasie rakende die bemanning en operasionele stelsels aan boord 'n skip gekoppel is aan die Dek- en Ingenieursdepartement, het ons gepoog om die ooreenkomste en verskille tussen die tegniese

leksikons van die twee gebiede te verken om te bepaal of aparte woordeboeke en glossariums dalk benodig word, en of 'n algemene maritieme woordeboek voldoende sal wees. Om die tegniese leksikons waarop gefokus is, te toets en te vergelyk, het ons hedendaagse korpuslinguistiese metodes en -sagteware gebruik. Die resultate toon beduidende verskille in die kernwoordeskat van die korpora wat vergelyk is. Die bevindings dui ongetwyfeld en weer eens daarop dat spesiale aandag en fokus vereis word wanneer die maritieme leksis geïnterpreteer word, en dui ook daarop dat 'n duidelike onderskeid tussen Engels vir Seevaart- en Engels vir Skeepsvaartingenieurswesedoeleindes getref moet word.

Sleutelwoorde: MARITIEME ENGELS, SEEVAARTENGELS, SKEEPSVAARTINGENIEURSWESE, KORPUS, WOORDESKAT, LEKSIS, FREKWENSIE, SLEUTELWOORDE, WOORDELYS, TEGNIESE WOORDEBOEK

1. Introduction

Having extensive experience in the education and training of seafarers, as well as expertise in English for Maritime Purposes, we have dealt with a huge range of professional genres, literature, dictionaries and glossaries. Considering the multicultural and multinational character of a global human activity such as seafaring, English language skills remain a major issue and a basic requirement for every non-native speaker of this global language and, at the same time, *lingua franca* of the shipping industry. Bearing in mind that specific, technical vocabulary is considered the main predictor of reading comprehension (Đurović et al. 2021), technical dictionaries and glossaries are always likely to be part of the basic kit of every seafarer.

Based on the two main departments that make a ship a specific working environment, this article aims to investigate the similarities and differences between two major aspects of English for Maritime Purposes. The main idea behind this research is to point to the existence and development of two related, but still separate and very specific, branches of English for Maritime Purposes: English for Nautical Purposes and English for Marine Engineering Purposes, and the reasons behind the necessity of separate lexicographical treatment in each case.

In addition, the scarcity of specialized technical lexicographic aids for marine engineers compared to nautical ones has become more than evident. Considering embarking on an endeavour as extensive as compiling a Marine Engineering Bilingual Dictionary, we firstly wanted to test and prove the necessity of and justification for such a project. Aiming to prove the distinction between the two lexes, the article focuses on the analysis of the lexical aspect of both the nautical and engineering professional books. In addition, we observed that some vocabulary items found in both corpora have different meanings, which demonstrates the polysemy and ambiguity of the two discourses, especially when taken out of context (Bisht et al. 2011).

2. Theoretical background

2.1 English for Maritime Purposes

The term Maritime English embraces a range of distinctive registers utilised in onboard and international communications (Cole et al. 2007). As a specialised branch of English for Specific Purposes (ESP), its specific communicative features, genres and vocabulary used by the vast maritime discourse community have been explored by many Maritime English (ME) researchers. According to Čulić-Viskota and Kalebota (2013) Maritime English is characterized by its own particular jargon with specific grammatical forms. Bocanegra-Valle (2013) states that English for Navigation and English for Marine Engineering are among the five subcategories of ME, emphasizing that English for Marine Engineering stands out as the most technical of all the varieties. ME vocabulary covers the vital language of written documents dealing with ship construction, engineering, surveillance, maintenance and the operation of ships. The proper interpretation of such a specialised vocabulary requires extra-linguistic knowledge or, more simply, knowledge of the field (Borucinsky and Kegelj 2019). With this in mind, Đurović (2021) extracted and analysed a marine engineering frequency word list from the corpus of marine engineering instruction books and manuals. This article aims to do the same for the nautical database and present comparative results regarding the nautical vs. marine engineering corpora in terms of their respective lexis content.

2.2 Maritime dictionaries and corpus-based headword selection for technical (sub-)areas

The first lexicographic activities regarding seafaring, one of the oldest professions that have connected people all around the world, can be traced back to the thirteen century in the forms of nautical glossaries and sea-related nomenclature (cf. Pritchard 2013). *The Seaman's Dictionary* by Henry Mainwaring (1644) represents one of the oldest English dictionaries of Maritime English (ibid.). The first dictionaries were written for the sake of naming basic concepts regarding ship's construction and concepts related to the sea (ibid.). Over the centuries, maritime dictionaries and subject-specific glossaries appeared under a number of titles such as: a sailor's dictionary, a dictionary of sea terms, a nautical dictionary, a dictionary of seafaring, or a marine/maritime dictionary (ibid.). With the rise of maritime activities and maritime nations, maritime monolingual and bilingual dictionaries have expanded and embraced a variety of maritime-related concepts regarding technology, trade, law, ship's business, and so on. As was, and still is generally the case in specialized lexicography, a variety of methods have been applied to gathering and presenting specialized vocabulary and terms, with corpus linguistics approach being slower than in

general lexicography (Bowker 2010).

Corpus linguistics methods are not a novelty in lexicography, emerging first with the *Dictionary of the English Language* by Samuel Johnson from the 18th century, followed by the *American Webster Dictionary* in the nineteenth, and then the first *Oxford English Dictionary*, all the way up to modern paper and online dictionaries (Cambridge, MacMillan, Oxford, Longman and others). The first and most salient contributions of corpus tools in lexicography were made in the context of monolingual dictionaries such as the *Collins COBUILD English Language Dictionary* (Sinclair 1992; Abdelzaher 2022).

Today, this way of compiling corpora is considered conventional and very helpful for creating macrostructures of lexicographical materials. In the present study, we partly deal with the methodology that is most conveniently used and evaluated with numerous word lists tailored for LSP/ESP courses and vocabulary materials. However, bearing in mind that maritime communications are generally prone to restrictiveness and the practical needs of concise message transmissions, we consider the methodology generally applicable in building any lexicographical material, where anything from reference vocabulary lists to comprehensive dictionaries can be acknowledged as a lexicographic attempt to isolate a distinctive register, as Opitz (cf. Čulić-Viskota and Rummel 2022) advocates in terms of segmental dictionaries. Therefore, the aim is to bring the basic word-list building methodology to an extended, more compound and higher level of application in designing technical (segmental) lexicographical products.

In addition, corpus selection also provides for retrieving illustrative contextual examples and cross-referencing among sub-corpora. The metalanguage built this way is very convenient for technical dictionaries, given their dense and often complicated nominal groups, which can often be ambiguous when they are not placed in a wider context (Borucinsky and Kegelj 2019; Čulić-Viskota and Rummel 2022).

2.3 Frequency word lists and keywords

For practical reasons, especially for a quick and practical response to the professional needs of ESP learners, a new software-based methodology was developed in the second half of the previous century, providing the statistically-justified extraction of frequency-based headwords. The first frequency word lists, ranging from simple ones to composite lexicographical publications, were those created for General English (GE). They provide (English) language learners with access to the most frequent English vocabulary found in various types of texts. Following this psycholinguistic perspective in building a native-speaker mental lexicon (Abdelzaher 2022: 168), the extraction of the most frequent General English words has been applied to a variety of ESPs. Contemporary software solutions such as RANGE (Nation and Heatley 1994) and its upgraded version

AntWordProfiler (<https://www.laurenceanthony.net/software/antwordprofiler/>) enable the elimination of the most frequent GE vocabulary from further analysis, thus focusing directly on the most frequent technical vocabulary. In this way, more and more ESP word lists have been generated, offering the core technical vocabulary of specific ESP texts and genres.

Although today it is generally considered outdated, one of the most influential GE word lists in regard to the recorded lexical analyses is still West's General Service List (GSL), dating back to 1953. It comprises 2,000 word families¹ and was obtained from an English corpus of 5 million words. It is often used together with the Academic Word List (AWL) of Coxhead (2000) which built upon the GSL (to avoid overlapping) and which consists of 570 word families extracted from various academic texts. These lists were (and still are) usually used together to test a target corpus and measure the GE and academic vocabulary load, providing the opportunity to compare this with other scholarly findings grounded in the same methodology.

A notable contribution to the GE word lists was provided by Nation (2012), who elicited 25 GE word lists from the textual compilation of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). In this way, 25 BNC/COCA lists of 1,000 GE word families were offered, with an additional four lists of the most frequent proper nouns, abbreviations, marginal words and transparent compounds, respectively.² These lists have been used to measure the coverage of GE vocabulary in various corpora.

In addition to the statistically most frequent vocabulary, it has become evident that certain words are very specific to a distinctive type of text, and the knowledge of these words facilitates understanding of the texts (Baker 2004; Al-Rawi 2017). These words of 'special status' (Stubbs 2010: 21) are considered key vocabulary. In statistical lexical analysis, they are obtained according to their frequency compared to a reference corpus, which is most often a GE one.

Considering the specific nature of the maritime lexicon, for both nautical and marine engineering, we will use both criteria, i.e. frequency and the key nature of the vocabulary, to analyse the vocabulary types and loads of both target corpora (see also Đurović 2023). In addition, we will analyse and compare them in order to provide solid and measurable answers to our research questions.

3. Research questions

The three research questions posed in this study are:

1. How many high-frequency, general-purpose and academic tokens are detected in nautical and marine engineering corpora?

2. How many headwords need to be mastered for an adequate reading comprehension of nautical and marine engineering corpora?
3. What does the detection of keywords and metadata reveal in terms of lexicographic similarities and differences of the two corpora referring to two maritime subdomains?

4. Corpus and research methodology

This section of the article will provide more specific detail about the corpus and the methodology used to answer the research questions given above.

4.1 Corpus details

The number of maritime communication spheres is becoming increasingly diverse, given the continuing globalisation and internalisation of the navigating profession. Maritime and marine-related literature comprises a myriad of academic, professional and technical publications covering different branches of the seafaring profession. For this study, we have deliberately chosen to analyse corpora comprised of nautical and marine engineering professional books.

Aware of the dynamic changes characteristic of the shipping industry and emerging concepts in the maritime profession, our choice was to cover the chronological range of books dating from 1990 to 2021. In the selection of the texts, we started with various conventional concepts in navigation (good seamanship, manoeuvring, vessel position and maintenance) and marine engineering (marine propulsion, engine operation and auxiliaries). Nevertheless, we also included some novel maritime- and marine-related concepts linked to smart applications, automation, propulsion systems, electronic navigation and 'ultimate concepts' in shipboard operations. In addition, we took into consideration the professional genres and the narrative character of the publications from the two professional areas. For certain other types of publications, such as marine engineering instruction books, it would be difficult to find a comparable counterpart in nautical publications.

Finally, we identified the eight selected books, available in electronic form, from the field of nautical studies, and our eight marine engineering books (Table 1 and Table 2). Totalling 768,135 and 813,429 running words (tokens) for the respective tables, the corpora can be considered both sufficient and relevant in size, as well as convenient for further comparative analysis.

Table 1: Corpus of Nautical Books (CONB)

No.	Book title	No. of types	No. of tokens
1	Laugier, C. and R. Chatila (Eds.). 2007. <i>Autonomous Navigation in Dynamic Environments</i> . STAR Springer Tracts in Advanced Robotics 35. Berlin: Springer.	5,194	57,656
2	Manley, P. 2008. <i>Practical Navigation for the Modern Boat Owner</i> . San Francisco: John Wiley & Sons.	3,203	33,052
3	Weintrit, A. and T. Neumann (Eds.). 2013. <i>Marine Navigation and Safety of Sea Transportation</i> . London: Taylor & Francis.	9,641	126,451
4	Dixon, C. and J.K. Spencer. 2021. <i>The Ocean: The Ultimate Handbook of Nautical Knowledge</i> . San Francisco: Chronicle Books.	10,861	90,743
5	Touche, F. 2005. <i>Wilderness Navigation Handbook</i> . Canada: Friesens.	4,344	56,649
6	Tetley, L. and D. Calcutt. 2001. <i>Electronic Navigation Systems</i> . 3rd edition. London: Routledge.	6,567	127,842
7	Grewal, M.S., L.R. Weill and A.P. Andrews. 2007. <i>Global Positioning Systems, Inertial Navigation, and Integration</i> . 2nd edition. Hoboken, New Jersey: Wiley-Interscience.	8,586	160,249
8	Barrass, C.B. and D.R. Derrett. 2012. <i>Ship Stability for Masters and Mates</i> . 7th edition. Oxford: Butterworth-Heinemann.	5,055	115,493
	Total	53,451	768,135

Table 2: Corpus of Marine Engineering Books (COMEB)

No.	Book title	No. of types	No. of tokens
1	Watson, G.O. 1990. <i>Marine Electrical Practice</i> . 6th edition. London: Butterworth-Heinemann.	6,266	108,200
2	Taylor, D.A. 1996. <i>Introduction to Marine Engineering</i> . 2nd edition. London: Butterworth-Heinemann.	5,680	92,124
3	Jackson, L. 2001. <i>Reed's General Engineering Knowledge for Marine Engineers</i> . 4th edition. London: Thomas Reed.	7,187	114,036
4	Tsinker, G.P. 1995. <i>Marine Structures Engineering: Specialized Applications</i> . Dordrecht: Springer Science+Business Media.	9,721	153,542
5	Martelli, M. 2014. <i>Marine Propulsion Simulation</i> . Warsaw/Berlin: De Gruyter.	3,061	48,759

6	Kantharia, R. 2013. <i>Marine Engineer's Handbook — A Resource Guide to Marine Engineering</i> . (e-Book) Marine Insight official website.	1,417	6,860
7	Hobart, H.M. 1911. <i>The Electric Propulsion of Ships</i> . London/New York: Harper and Brothers.	4,149	54,895
8	Woodyard, D. 2004. <i>Pounder's Marine Diesel Engines and Gas Turbines</i> . 8th edition. Oxford: Butterworth-Heinemann.	8,537	235,013
	Total	46,018	813,429

4.2 Research methodology

The initial method applied is known as Lexical Frequency Profiling (Laufer and Nation 1995), used for measuring vocabulary types and levels in a corpus. For this purpose, we used the latest AntWordProfiler software version 2.0.1, which is largely an upgrade of the previously used RANGE programme (Nation and Heatley 1994), developed for the lexical analysis of texts. The word lists used for testing coverage with general and academic vocabulary in our target corpora were the General Service List (West 1953) and the Academic Word List (Coxhead 2000), as they are generally used together in this kind of analysis, and for the purpose of comparison.

The same software was used to test the corpus coverage by General English word lists. Following the same reason of comparability, for measuring the coverage level of General English vocabulary, we used the Nation's (2012) BNC/COCA word lists developed from the British National Corpus and Corpus of Contemporary American English.

For keyness analysis, version 4.1.0 of the AntConc software (<https://www.laurenceanthony.net/software.html>) was used. As a referent GE corpus for keyness analysis, we used the Freiburg version of the Lancaster-Oslo-Bergen (FLOB) corpus. The rationale behind the selected referent GE corpus was its size, as it is close to the target corpora, as well as the fact that it was developed as a British counterpart to the Brown GE corpus of American English.

For preparing the corpora for further software analysis, we used AntFileConverter 2.0.2 by the same author (<https://www.laurenceanthony.net/software/antfileconverter/>), since the referent files need to be uploaded in plain text format.

In addition, the lists are generally available as headwords (word families), and thus were expanded to all family members using the Familizer + Lemmatizer programme (Cobb 2018). Special attention should here be paid to the headword entries, such as lemmas or word families, whereas the final selection should always be dictated by both user needs and expert advice (Atkins 2008; Đurović 2021).

In choosing the most appropriate methods, we were led by the most practical and illustrative results that can be obtained and that were tested by utilising various software options and settings.

5. Findings

5.1 Research Question 1

Aiming to test the vocabulary load and types in the target corpora, we used West's General Service List (West 1953) and the Academic Word List of Coxhead (2000) as mentioned in Section 5.2. We tested the corpora (CONB and COMEB) individually and the comparative results are presented in Table 3.

Table 3: Coverage of GSL and AWL in the CONB and COMEB

Word lists	Tokens CONB	Tokens COMEB	Coverage CONB (%)	Coverage COMEB (%)
GSL	548,287	600,580	71.38	73.83
AWL	69,468	63,289	9.04	7.78
Not in the lists	150,380	149,560	19.58	18.39
Total	768,135	813,429	100.00	100.00

The coverage of the Nautical corpus by GE vocabulary is 71.38%, which is similar to the GSL coverage measured in Marine Engineering instruction books (Đurović et al. 2021), but, as can be seen from Table 3, this is still somewhat higher in Marine Engineering general books (73.83%). As expected, considering the technical and specific nature of maritime publications, the GE coverage (GSL) is below the general coverage of 78–98% expected to be found in various types of written text (Nation and Waring 1997). Interestingly, it is slightly higher than in various academic texts, where it reached the level of 70–71.9% (Coxhead 2000).

On the other hand, the coverage of the Academic Word List solely does not reach the average of about 10% which is measured in research articles and textbooks (Coxhead 2000) or in Medical texts (Chen and Ge 2007), or the values of 11.17% in Applied Linguistics (Vongpumivitch et al. 2009), 9.96% for Chemistry texts (Valipouri and Nassaji 2013) and 9.47% for Pharmacy related material (Fraser 2007). However, the values reached are still higher than the 8.07% found in Marine Engineering instruction books (Đurović et al. 2021). Bearing all the figures in mind, as well as the genre in use, the given corpora can be considered an academic type of text to some extent, but more evidently it fits the genre of technical literature. The presence of the academic discourse can be explained by the fact that the professional books analysed in here contain an academic narrative for educational and pedagogical purposes as a means by

which to make the teaching of distinctive subjects to seafarers and officers in the classroom easier (cf. Franceschi 2014).

If we take into account the cumulative coverage of GSL and AWL (80.42% in the CONB and 81.61% in the COMEB) we observe that it is below the average of 86.1% found in academic texts (Nation 2000: 27). This clearly points to the demanding nature of maritime publications in terms of technical vocabulary load. Considering that, and knowing the first 2,000 GE words and most common academic vocabulary, we are left with nearly 20% of the specific vocabulary being unknown, or every fourth to fifth word of the narrative, which would hinder proper comprehension of these publications. This fact speaks in favour of the importance of maritime technical vocabulary, in the view of both Nautical Sciences and Marine Engineering.

From this specific research, it is evident that Nautical books are more demanding in terms of vocabulary load than Marine Engineering books. This again can be associated with the nature of the work on deck, including the larger scope of activities on the ship's bridge, including external communications with other ships and shore stations. In the case of marine engineers, communication rests on intra-ship communicative activities and deck-engine speech activities (De Castro 2020). However, further caution is required here since different findings have emerged from previous research conducted on Marine Engineering publications and vocabulary (Bocanegra-Valle 2013; Hsu 2014; Đurović et al. 2021), which all point to Marine Engineering technical publications being the most demanding ones. For example, in Marine Engineering instruction books, over a fifth (20.5%) of the vocabulary was "unknown", i.e. not covered by the GSL and the AWL (Đurović et al. 2021). These results point to the complexity and multidisciplinary nature of the maritime subareas, as well as the necessity of a detailed justification of the corpus and methodology applied, including their limitations.

5.2 Research Question 2

Considering the previous analysis, we raise the question regarding the amount of vocabulary needed for an adequate reading comprehension of Nautical and Marine Engineering publications. This evaluation method is one of the main advantages of word lists compared to dictionaries, which are generally much richer in entries, but at the same time, can be considered much more robust, as well.

For testing the required comprehension level of our corpora, we examined the coverage of the BNC/COCA GE word lists (Nation 2012), respectively. As usual in this type of research, we also tried to exclude the most frequent abbreviations, transparent compounds, marginal words and proper nouns (the four additional lists).

The comparative findings of the analysis conducted using the AntWord-Profiler are given in Table 4.

Table 4: Coverage of the BNC/COCA lists in CONB and COMEB

Word lists	Coverage in CONB (%)	Coverage in COMEB (%)
2,000 + 4 additional lists	78.65	78.68
3,000 + 4 additional lists	87.74	87.12
4,000 + 4 additional lists	90.66	90.01
5,000 + 4 additional lists	92.05	92.52
6,000 + 4 additional lists	98.98	93.74
7,000 + 4 additional lists	93.51	94.66
8,000 + 4 additional lists	93.96	95.25
9,000 + 4 additional lists	94.23	95.83
10,000 + 4 additional lists	94.51	96.15
11,000 + 4 additional lists	94.69	96.33
12,000 + 4 additional lists	94.76	96.50
13,000 + 4 additional lists	94.90	96.61
14,000 + 4 additional lists	95.05	96.71
25,000 + 4 additional lists	95.60	97.32

From Table 4 it is apparent that the corpus of Marine Engineering books had better coverage in GE vocabulary. Hence, we note that the desired level of 95% required for adequate reading comprehension (Laufer 1989) is reached at the level of 8,000 GE words, which corresponds with the results obtained by Hsu (2014) for Marine Engineering textbooks in comparison with other areas of engineering. On the other hand, in Marine Engineering instruction books, that level is reached only after 12,000 GE words. Interestingly again, the corpus of Nautical books shows even lower coverage by GE vocabulary, since 95% coverage is reached at the level of 14,000 GE words. We can assume that this is the case because Marine Engineering is a branch of engineering that shares many discursive aspects with general engineering discourse (Borucinsky and Kegalj 2019; Bocanegra-Valle 2013), which makes it more multidisciplinary and more familiar to a wider range of English speakers. Furthermore, Nautical English can be considered the most technical in terms of the Maritime sciences. The only other domain sharing lexical registers with navigation is aviation. To the best of our knowledge, no similar research has been conducted on aviation publications or (text-)books, which is a potential area for further comparative research.

In addition, the ideal coverage of 98% (Hu and Nation 2000) is not reached in either corpus even with all the available 25,000 GE words, which definitely confirms that English for Maritime Purposes is a very technical ESP, the vocabulary of which requires special attention or teaching and learning efforts.

5.3 Research Question 3

A further distinction, and certain similarities, were analysed by testing the keyness of the two corpora.

Firstly, we tested the target corpora against each other, to obtain the key nautical terms tested against the ME corpus, and then the other way around: we analysed the ME keywords against the referent corpus of Nautical books (CONB). For this purpose, we used AntConc software (Anthony 2022). For practical reasons, we present here only the first 19 results extracted from the programme tables.

	A	B	C	D	E	F	G
1	type	freq_tar	freq_ref	range_tar	range_ref	likelihood	effect
2	you	2508	67	8	5	3094.486	0.007
3	your	1999	17	8	3	2716.386	0.005
4	gps	1368	0	7	0	1977.204	0.004
5	navigation	1602	112	8	7	1636.774	0.004
6	code	970	22	8	5	1219.845	0.003
7	error	1114	60	8	6	1215.87	0.003
8	satellite	830	2	5	2	1173.86	0.002
9	signal	1480	213	8	6	1140.987	0.004
10	draft	917	29	6	6	1104.389	0.002
11	maritime	866	32	6	6	1017.63	0.002
12	receiver	1119	121	6	6	984.936	0.003
13	compass	654	2	6	1	920.421	0.002
14	data	1569	353	8	7	903.514	0.004
15	tonnes	835	55	3	5	866.875	0.002
16	errors	723	34	8	5	812.387	0.002
17	chart	682	31	5	6	771.563	0.002
18	robot	525	1	2	1	745.299	0.001
19	map	578	16	7	2	709.135	0.002
20	center	606	24	7	2	703.533	0.002

Figure 1: Keyness of CONB vocabulary compared against COMEB

As can be seen from the columns in this programme setting, we have a list of the word types ranked by the Effect Size Measure + Threshold — which is the result of the keyness strength calculation and points to the possible cut-off points that we can choose as the thresholds for our table size, i.e. the length of the word list. Since we did not opt for a word list here but for comparative analyses of the keywords of the two corpora, we did not have to choose the cut-off point but, rather, a convenient presentation of the results. The same is true for the Likelihood Measure + Threshold column.

Regarding the list contents, most of the words belong to the nautical/navigational register (*GPS*, *navigation*, *maritime*, *compass*, and so on). On the other hand, we were surprised by the fact that the personal pronoun *you* was detected as one with a very distinctive frequency compared to the COMEB. Intrigued by the finding, and thanks to various other options offered by the

AntConc software, we explored the context of the word *you* in the nautical corpus further. The majority of the widespread use of *you* in instructions and explanations is found in the form of conditionals (Figure 2), which is obviously not the case with the COMEB, where, generally, passive forms prevail in describing systems and operations.

PADI instructors are held to the same exact standards whether	you'	re in the Caribbean, Germany, or Japan. It's
if you're in the eastern hemisphere and west if	you'	re in the western hemisphere. For example, if you'
if you're in the western hemisphere. For example, if	you'	re in the eastern hemisphere and the official time
you a precisely interpolated but uncorrected bearing of $225^\circ + 107/4 = 252^\circ$.	You'	re in the eastern hemisphere and the official time
you a precisely interpolated but uncorrected bearing of $45^\circ - 72/4 = 27^\circ$.	You'	re in the western hemisphere and the official time
should avoid them all. [Pufferfish are extremely poisonous; avoid unless	you'	re in a restaurant being served by a chef
the rushing water. Do not stay in a car. If	you'	re in a boat at sea, avoid shoals, shallow
the surf zone for the lull between wave sets. If	you'	re in a life raft or boat, see "Be
risks. Kidnapping on land—all pirates can do that. If	you'	re in a pirate gang, and you decide kidnapping
use an acoustic device that sounds like a bird if	you'	re in a bird sanctuary. Do not depend on

Figure 2: The word "you" used in the CONB context

These variations concerning the second-person pronoun *you* can be explained by the differences in the two discourses: nautical narrative storytelling, since the books are aimed at Nautical students and seafarers and cover countless "if" navigational situations that a navigator may encounter at sea (piracy, determining bearing and heading, navigation in shallow waters, etc.). Conversely, the discourse prevalent in Marine Engineering narratives is instructional, and focuses on lexical words (mainly nouns and verbs) which carry their meaning in collocations (e.g. *run the engine*) or are passive in the written texts (*the valve is opened*).

Keeping in mind the mentioned specificities of the navigational/engineering context, in Figure 3, specifically in the Range columns, we show that some nautical terms rarely appear in ME books, such as *GPS* (Global Positioning System) or *compass*, whilst some are present in both corpora, but much more frequently in Nautical books (*navigation, error, maritime, receiver*). Moreover, if we count the frequencies and keyness cumulatively for word families, some words such as *error(s)* would be even higher-ranked than they are in what is presented here.

Unlike the results given in Figure 1, Figure 3 presents typical Marine Engineering vocabulary that rarely appears in Nautical books. This includes words such as *valve* and *turbine*, but also words that appear in both corpora in the same range (or number of publications, respectively), such as *air* and *temperature*, but with relatively high frequency in Marine Engineering books. This makes them prevalent lexemes in this kind of maritime publication.

	A	B	C	D	E	F	G
1	type	freq_tar	freq_ref	range_tar	range_ref	likelihood	effect
2	engine	3569	136	8	6	3783.747	0.009
3	fuel	2701	81	8	6	2980.409	0.007
4	cylinder	2138	9	7	4	2742.362	0.005
5	engines	2147	14	7	5	2709.058	0.005
6	pressure	2513	239	7	7	2065.445	0.006
7	oil	2419	218	8	6	2030.218	0.006
8	valve	1618	17	7	2	1988.664	0.004
9	gas	1599	72	8	6	1638.106	0.004
10	piston	1161	0	6	0	1544.728	0.003
11	diesel	1209	9	8	2	1515.341	0.003
12	injection	1051	0	5	0	1398.303	0.003
13	exhaust	1057	8	6	3	1323.633	0.003
14	steam	997	6	7	3	1261.693	0.002
15	air	1874	272	8	8	1254.837	0.005
16	pump	1120	40	8	4	1199.882	0.003
17	turbine	911	3	7	1	1175.988	0.002
18	temperature	1327	131	7	7	1073.553	0.003
19	turbines	794	0	7	0	1056.255	0.002
20	propeller	932	25	7	4	1044.406	0.002

Figure 3: Keyness of COMEB vocabulary compared against CONB

Finally, we wanted to obtain the keyword lists for both areas/types of publications with reference to GE English, and compare the results. Considering the software requirements, best practice and the size of the corpora, we used, as mentioned, the FLOB corpus as the referent GE corpus.

Through separate analysis procedures, we obtained keyword lists from the CONB and COMEB, represented by the first 50 keywords for each of the vocabulary types (Table 5). For practical reasons, in this table, we put family members together (e.g. *signal* and *signals*, *user* and *use*, and so on)

Table 5: Keywords of the CONB and COMEB against the FLOB

No.	Nautical keywords	Marine engineering keywords
1	ship(s)	engine(s)
2	Navigation	fuel
3	GPS	speed
4	signal(s)	cylinder
5	Data	oil
6	receiver(s)	system
7	Figure	pressure
8	error(s)	figure
9	system, systems	water
10	Position	ship

11	Maritime	ice
12	Draft	valve
13	Code	marine
14	Tonnes	load
15	satellite, satellites	gas
16	Vessel	air
17	Using	diesel
18	Frequency	piston
19	Compass	pump
20	Water	injection
21	Angle	exhaust
22	Speed	control
23	Stability	temperature
24	Centre	propeller
25	Chart	low, lower
26	Matrix	steam
27	Velocity	turbine
28	meter(s)	current
29	Safety	shaft
30	Map	propulsion

Here it is clear that the key/technical vocabulary is different not only when directly compared against each other, but also when tested against GE. Considering that the two maritime areas share the language for general maritime purposes (IMO 2015), we can say that they are particularly distinct in terms of their lexical registers.

In addition, we have to pay special attention even to seemingly common technical vocabulary. For example, we have *bearing* in both lists, but in Nautical English it points to a position (e.g. 'Our vessel is bearing 215 degrees from you'), while in Marine Engineering it refers to a machinery component ('We must replace the main bearing'). To make it even more complicated, we have words from GE that have a completely different meaning in Maritime English. Another example of an ambiguous word detected in both corpora is the term *pressure*, since in Marine Engineering it refers to the physical force in the elements, while it is connected with weather systems in Nautical English.

The stated difference in the keywords in both corpora confirms the "polarity" of these two discourse communities and illustrates the problematic process of denotations in technical dictionaries. This brings us to the standing "conflict" among linguists and lexicographers with regard to the lexical vs. contextual meaning of entries (cf. Abdelzaher 2022). Taking into consideration the pronounced technical nature of the target corpus, as examined above, an eclectic approach is required in building technical dictionaries, still inclining towards contextual considerations, or even specific models such as collocate-to-sense mapping, as proposed by, for example, Kilgarrieff (2005).

Another distinction worth mentioning in the specific case is the distinction between marine and maritime, as the two terms are often used interchangeably although they are related to different aspects of shipping. The word *maritime* occurs in the Nautical corpus but not in the Marine Engineering one, suggesting that the adjective *maritime* refers to activities associated with the navigational actions happening on deck. On the other hand, the word *marine* refers to operations relating to machinery, ship engines, environmental protection, or other activities under the ship's hull or in the sea (e.g. *marine engine*, *marine propulsion*, *marine fuels*, *marine environment*, *marine species*) (Dževerdanović-Pejović 2020a). In addition, the most frequent noun in the Nautical corpus, the word *ship*, has the most generic meaning in the Nautical corpus, as it refers to sailing, type of vessel, legal framework, conventions, and so on. Finally, the term *water* prevails in the Marine Engineering corpus as it is associated with the proper operation of the vital elements in the engine room, as in *the level of water in boilers*, *fresh water*, *bilge water*, *sea water*, and other similar phrases, whereas in the Nautical corpus it is mostly connected with navigational conditions (*calm water*, *safe waters*, *busy waters*, *high water*).

6. Limitations of the study

With using the above statistical methods, we should bear in mind not to strictly abide by the computational counts only. A further investigation into the obtained results and statistics, including as they relate to technical expertise and experience, is indispensable in the process.

As we can infer from the above analysis and discussion, there are numerous polysemous or cryptotechnical words (Fraser 2009) found in maritime publications that limit the accuracy of the results. For example, the noun *list*, found in the most frequent GE words in Maritime English, may refer to the inclination of the ship to one side or the other. In addition, some GE words in collocations with others refer to specific marine systems, such as *jacket water cooling system* or *guide shoe*, again referring to specific engine components. The same is found with nautical vocabulary such as *draft*, *stability* and *code*.

In addition, as was also found by Bocanegra-Valle (2013), seafaring- and ship-related words, especially nouns, are prone to being merged into compounds. For example, in Table 5 we find *shaft* as one of the most frequent keywords. If we were to consider the fact that it is frequently used in, for example, *crankshaft* and *camshaft*, by grouping these together with *shaft*, it would rank even higher on the list. The same criteria, generally not recognised by the lexical analysis software, would bring additional words beyond the decided cut-off point for frequency word lists. One additional detail should be borne in mind. The definition of *word* should be defined in advance, as it is (differently) the case in different programmes and their settings (e.g. headword/word family, lemma, word type or token/running word).

In addition, the specific results are always limited and conditioned by the selected corpus, as with the example of the results obtained from our research using Marine Engineering books and the results obtained from Marine technical manuals (Đurović et al. 2021). This means that even within a unique area of ESP we can obtain a variety of results. Therefore some authors (e.g. Gizatova 2018) only partly rely on corpus linguistics in building lexical macrostructures, such as in checking the frequency of otherwise selected terms, collocations or idioms. Conversely, our intention was to show the further prospective utility of technical corpora and their lexical analysis.

Although frequency is generally considered a solid base for headword selection, Rundell and Kilgariff (2011) rightly mention that it is not an adequate criterion when, for example, extracting multiword items. In addition, there is always the dilemma on the right cut-off point, which, again, suggests that frequency on its own cannot guarantee that all the relevant words will be selected (Nielsen 2018; Vuković-Stamatović and Živković 2022). That is why, for these analyses, a detailed presentation of the methodology and the interpretation of the derived specific vocabulary in a specific context are required and recommended. In order to avoid the above-mentioned limitations, quantitative methods should be combined with qualitative expertise. In that way, a comprehensive study of the elicited figures regarding frequency, coverage and keywords should include both the data and a 'knowledge of the world' interface (Van Dijk 2014: 5).

7. Practical implications of the research

Regardless of the limitations of the methodology, the study still provides a very useful insight and measurable results in terms of the types of vocabulary, vocabulary load and specificity that this ESP may feature in comparison with similar research. In addition, lexicographers are provided with solid and justifiable methods for headword extraction, while language instructors and learners are offered corpus-based technical vocabulary and a methodology to focus on.

Our research aimed primarily to compare the discourses of two specific maritime communicative domains in terms of their technical vocabulary. However, the methodology and software used could be employed to derive a number of further lexicographic benefits. For example, the AntConc software can detect the most frequent collocations and *n*-grams, which enabled us to reveal the words in a particular context and isolate their contextual instances, including their specific and often different semantic aspects and connotations.

AntWordProfiler, on the other hand, could be used to develop frequency lists, such as the one generated from Marine Engineering instruction books (Đurović et al. 2021). The purpose of those lists (and their evaluation, at the same time) is to reach the level of 95% coverage in a target text sooner than

with GE word lists, and these lists are usually built upon the first 2,000–3,000 GE words, which are considered the most commonly known among English language learners. On the other hand, corpus-based frequency is a solid criterion for any word list, including glossaries and more complex lexicographic endeavours such as dictionaries. This has been the case with GE dictionaries, such as, for example the Collins COBUILD English Language Dictionary, Macmillan's Dictionary or many others produced in the meantime (Abdelzaher 2022; Đurović 2021). Having said that, the utilization of this criterion and generally corpus-linguistics methods has been rather slow with ESP dictionaries (Bowker 2010). One of the most widely cited reasons for this is the less abundant corpora compared to the one for GE. Thus, the frequency-based methodology of vocabulary detection has not been commonly used for technical dictionaries and glossaries, except for the numerous specialized word lists tailored to meet the specific needs of the language learners or professionals.

Frequency and keyword lists might have still further lexicographical implications. In addition to building frequency-based technical glossaries or dictionaries, both criteria can be combined for the purpose of comprehensiveness. In addition, knowing that deck and engine officers are in favour of patterns or schemes in the acquisition and the production of restricted verbal and written genres (Dževerdanović-Pejović 2020b), frequency and keyword lists, as well as certain other options offered by modern lexical software, can also serve to establish the co-occurrence and mapping of a specific genre, contributing to the accurate and reliable metalanguage of a technical dictionary or similar segmental lexicographical product.

In addition, the methodology (or methodologies) can be combined and utilized not only for segmental lexicography, but also for building hybrid dictionaries (cf. Bowker 2010). For example, a hybrid ME dictionary could include the most frequent nautical and/or marine engineering words, but also the most frequent GE ones as extracted from the same corpus, with additional semantic notions in the given corporal context due to the polysemous character of some GE and "cryptotechnical" words. As such, it would cover the full English vocabulary load of the specific genre(s).

Finally, the possibility of using modern software tools in the analysis of large corpora and thus of tackling a huge amount of data enables us not only to dig into a particular professional communicative domain, but also to engage in comparative and contrastive research on specific lexes and to explore the worlds in which their semantic matters are imprinted. In this way, the specific lexicographic metalanguage would reflect the peculiarities of distinctive technical subareas, such as those shown in the nautical vs. marine engineering context.

8. Conclusion

Intrigued and inspired by the distinctiveness of the lexical registers characteristic of the nautical and marine engineering genres, which at the same time share

the General Maritime vocabulary, we were looking for adequate corpus linguistics and statistical lexical methods that could provide us with measurable results in terms of the differences between the two types of vocabulary. For that purpose, we sought to provide answers to three research questions related to the compiled corpora of Nautical and Marine Engineering books. The aim of the three-step research was to provide a solid foundation for the separate treatment of specialized maritime lexicons dedicated to the professionals in those specific fields.

Firstly, both genres proved to be very challenging in terms of the technical vocabulary load, since the coverage of the corpora with GE vocabulary together with academic vocabulary was lower than in other types of texts. This was even more the case for Nautical books.

Similar results were found in measuring the coverage of the corpora by GE vocabulary only (BNC/COCA word lists). This part of the lexical analysis showed that adequate reading comprehension (expected at a level of 95% coverage with GE vocabulary) would be achieved with no fewer than 14,000 GE words in the case of Nautical books and with 8,000 words in the case of Marine Engineering books. Again, Nautical books proved to be more technical vocabulary-wise, whilst the ideal threshold of 98% for ideal reading comprehension was not reachable even with all the available word lists covering GE vocabulary.

In addition, we extracted the keyword lists from both corpora (in comparison with a referent GE corpus) and examined them for potential similarities and differences. This provided us with additional evident differences in the two registers, genres and discourses. Some common terms, such as *pressure*, *bearing* and so on, hold different meanings in the two respective corpora, pointing to the highly represented phenomenon of polysemy in maritime lexis.

Considering the decisions made by throughout the process, we followed the general recommendations for the methodologies applied, but also stated the limitations of the study, primarily those related to the selected genres i.e., the specific content of the corpora. It was also noted that the entire process of analysis cannot rest on statistical results only, but also requires expert knowledge with regard to the two maritime areas and their respective registers. In addition, the research findings point to the great challenge imposed on Maritime English lexicographers and the special attention required when dealing with these demanding and intriguing areas of English for Specific Purposes.

Finally, having at hand very meticulous methodologies for technical vocabulary extraction, a more comprehensive project could be conducted to comprise various genres and combined methodologies for a more complex lexicographical product such as, for example, a (mono-, bi- or multi-lingual) Marine Engineering dictionary.

Endnotes

1. A word family comprises the head word with all its inflected and derived forms.
2. Transparent compounds are compounds where the meaning can be understood from the separate meanings of their constituents.

References

- Abdelzaher, E.M.** 2022. An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years. *Lexikos* 32(1): 162-179.
- Al-Rawi, M.K.S.** 2017. Using AntConc: A Corpus-based Tool to Investigate and Analyse the Keywords in Dickens' Novel 'A Tale Of Two Cities'. *International Journal of Advanced Research* 5(2): 366-372.
- 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.
- Baker, P.** 2004. Querying Keywords: Questions of Difference, Frequency and Sense in Keywords Analysis. *Journal of English Linguistics* 32(4): 346-359.
- Bisht, R.K., Dhami, H.S. and N. Tiwari.** 2011. An Evaluation of Different Statistical Techniques of Collocation Extraction Using a Probability Measure to Word Combinations. *Journal of Quantitative Linguistics* 13(2-3): 161-175.
- Bocanegra-Valle, A.** 2013. Maritime English. Chapelle, C.A. (Ed.). 2013. *The Encyclopedia of Applied Linguistics*: 3570-3583. Hoboken: Blackwell Publishing.
- Borucinsky, M. and J. Kegalj.** 2019. Syntactic Ambiguity of (Complex) Nominal Groups in Technical English. *International Journal of English Studies* 19(2): 83-102.
- 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.
- 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 (RAs). *English for Specific Purposes* 26: 502-514.
- Cobb, T.** 2018. From Corpus to CALL: The Use of Technology in Teaching and Learning Formulaic Language. Siyanova-Chanturia, A. and A. Pellicer-Sanchez (Eds.). 2018. *Understanding Formulaic Language: A Second Language Acquisition Perspective*: 192-211. New York: Taylor & Francis.
- Cole, C., Pritchard, B. and P. Trenkner.** 2007. Maritime English Instruction — Ensuring Instructors' Competence. *Ibérica* 14: 123-148.
- Coxhead, A.** 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Čulić-Viskota, A. and S. Kalebota.** 2013. Maritime English — What Does It Communicate? *Transactions on Maritime Science* 2(2): 109-114.
- Čulić-Viskota, A. and H. Rummel.** 2022. Tribute to Kurt Opitz (1931–2021) — German Contributor to Maritime English — What Does It Communicate? *Transactions on Maritime Science* 11(2). <https://doi.org/10.7225/toms.v11.n02.019>
- De Castro, G.** 2020. A Needs Analysis Study of Engine-Deck Communication: Towards Improving Syllabus Design. *Taiwan International ESP Journal* 11(2): 35-53.

- Dževerdanović-Pejović, M.** 2020a. Discourse Analysis of the Research Articles About Marine Environment Relating to the Adriatic Coast. Joksimović, D., M. Đurović, I.S. Zonn, A.G. Kostianoy and A.V. Semenov (Eds.). 2020. *The Montenegrin Adriatic Coast. The Handbook of Environmental Chemistry* 110: 175-189. Cham: Springer.
- Dževerdanović-Pejović, M.** 2020b. Learning Technical Genres — A Blended Learning Approach. *Pomorstvo: Scientific Journal of Maritime Research* 34(2): 212-222.
- 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.
<http://dx.doi.org/10.5788/31-1-1647>
- Đurović, Z.** 2023. Frequency or Keyness? *Lexikos* 33: 184-206.
<https://doi.org/10.5788/33-1-1807>
- Đurović, Z., M. Vuković Stamatović and M. Vukičević.** 2021. How Much and What Kind of Vocabulary do Marine Engineers Need for Adequate Comprehension of Ship Instruction Books and Manuals? *Círculo de lingüística aplicada a la comunicación* 88: 123-133.
<https://doi.org/10.5209/clac.78300>
- Franceschi, D.** 2014. The Features of Maritime English Discourse. *International Journal of English Linguistics* 4(2): 78-87.
- 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.
- Fraser, S.** 2009. Breaking Down the Divisions between General, Academic, and Technical Vocabulary: The Establishment of a Single, Discipline-based Word List for ESP Learners. *Hiroshima Studies in Language and Language Education* 12: 151-167.
- Gizatova, G.** 2018. A Corpus-based Approach to Lexicography: A New English–Russian Phraseological Dictionary. *International Journal of English Linguistics* 8(3): 357-363.
- Hsu, W.** 2014. Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates. *English for Specific Purposes* 33: 54-65.
- Hu, M. and I.S.P. Nation.** 2000. Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language* 13: 403-430.
- IMO Model Course 3.17, Maritime English.** 2015. London: International Maritime Organization.
- Kilgariff, A.** 2005. Putting the Corpus into the Dictionary. *Proceedings of the Second MEANING Workshop*. Trento, Italy, 3–4 February 2005.
- Laufer, B.** 1989. What Percentage of Text-lexis is Essential for Comprehension? Lauren, C. and M. Nordman (Eds.). 1989. *Special Language: From Humans Thinking to Thinking Machines*: 316-323. Clevedon: Multilingual Matters.
- Laufer, B. and I.S.P. Nation.** 1995. Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16(3): 307-322.
- Nation, I.S.P.** 2000. Review of *What's in a Word? Vocabulary Development in Multilingual Classrooms* by N. McWilliam. 2012. *Studies in Second Language Acquisition* 22(1): 126-127.
- Nation, I.S.P.** 2012. *The BNC/COCA Word Family Lists*. Available at:
<http://www.victoria.ac.nz/lals/about/staff/paul-nation>.
- Nation, I.S.P. and A. Heatley.** 1994. *Range: A Program for the Analysis of Vocabulary in Texts (Computer Software)*. Retrieved from:
<http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>

- Nation, I.S.P. and R. Waring.** 1997. Vocabulary Size, Text Coverage, and Word Lists. Schmitt, N. and M. McCarthy (Eds.). 1997. *Vocabulary: Description, Acquisition and Pedagogy*: 6-19. Cambridge: Cambridge University Press.
- 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.
- Pritchard, B.** 2013. The English Element in the Development of Croatian Maritime Terminology. *Pomorstvo: Scientific Journal of Maritime Research* 27(1): 247-259.
- Rundell, M. and A. Kilgariff.** 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.
- Sinclair, J.** 1992. *Collins COBUILD English Language Dictionary*. London: HarperCollins.
- Stubbs, M.** 2010. Three Concepts of Keywords. Bondi, M. and M. Scott (Eds.). 2010. *Keyness in Texts: Corpus Linguistic Investigations*: 21-42. Amsterdam: John Benjamins.
- 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.
- Van Dijk, T.A.** 2014. *Discourse and Knowledge: A Sociocognitive Approach*. Cambridge/New York: Cambridge University Press.
- 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.
- Vuković-Stamatović, M. and B. Živković.** 2022. Corpus-based Headword Selection Procedures for LSP Word Lists and LSP Dictionaries. *Lexikos* 32(1): 141-161.
- West, M.** 1953. *A General Service List of English Words*. London: Longman, Green & Co.

The Operative Function in Spanish Lexicography Exemplified through Sport Dictionaries and Other Reference Works

Ángel Huete-García, *Faculty of Medieval and Modern Languages,
University of Oxford, Oxford, Great Britain*
(angel.huete-garcia@mod-langs.ox.ac.uk)

Abstract: Over time, dictionaries have been adapted to meet users' consultation needs. As a result, other types of data such as instructions or specialised explanations have been included in these works in order to fulfil their purpose. It is in this context that operative lexicographical tools are produced, although such tools have traditionally received very little attention from a metalexicographic point of view. Through the analyses of 5 existing dictionaries and other reference works from the 18th to the 20th centuries regarding equestrianism, football and fencing, this article shows that many lexicographical information tools in Spanish were indeed produced to satisfy the need for practical knowledge through instructions (operative need). Based on this finding, this article argues that the operative function of lexicography deserves a place in lexicographical theory and lexicographical academic literature. An interesting finding that emerges from these analyses is that almost all the dictionaries analysed are polyfunctional, and they share a list of characteristics based on their typology, media and historical moment that can help to make clear the importance of the operative function in the Spanish lexicographical tradition. Furthermore, on the basis of the above analyses, several answers are given to guide future research on operative lexicographical products.

Keywords: FUNCTION THEORY, SPANISH LEXICOGRAPHY, SPORT DICTIONARIES, POLYFUNCTIONAL TOOLS, OPERATIVE FUNCTION, SPECIALISED DICTIONARIES

Opsomming: Die operatiewe funksie in die Spaanse leksikografie soos geïlustreer met voorbeelde uit sportwoordeboeke en ander naslaanbronne. Woordeboeke is mettertyd aangepas om aan die naslaanbehoefes van gebruikers te voldoen. As gevolg hiervan is ander tipe data soos aanwysings of gespesialiseerde verklarings in hierdie werke ingesluit om aan hul doelwitte te voldoen. Dit is binne hierdie konteks dat operatiewe leksikografiese hulpmiddels geskep word, alhoewel hierdie hulpmiddels gewoonlik baie min aandag vanuit 'n metaleksikografiese oogpunt geniet het. Deur die ontleding van 5 bestaande woordeboeke en ander naslaanwerke vanuit die 18de tot 20ste eeu betreffende ruiterskuns, voetbal en skermkuns, word daar in hierdie artikel aangetoon dat baie leksikografiese inligtingshulpmiddels in Spaans inderdaad geskep is om in die behoefte aan praktiese kennis deur middel van aanwysings (die operatiewe behoefte) te voorsien. Gebaseer op hierdie bevinding, word daar in hierdie artikel aangevoer dat die operatiewe funksie van die leksikografie 'n plek in die leksikografiese teorie en leksikografiese akademiese literatuur verdien. 'n Interessante bevinding wat uit hierdie ontleding na

vore kom, is dat byna al die woordeboeke wat geanaliseer is, polifunksioneel is, en dat hulle 'n lys kenmerke rakende hul tipologie, medium en geskiedkundige tydperk deel wat van hulp kan wees om die belangrikheid van die operatiewe funksie in die Spaanse leksikografiese tradisie te verstaan. Gegronde op bogenoemde ontledings, word ook verskeie oplossings gebied om toekomstige navorsing oor operatiewe leksikografiese produkte te rig.

Sleutelwoorde: FUNKSIETEORIE, SPAANSE LEKSIKOGRAFIE, SPORTWOORDEBOEKE, POLIFUNKSIONELE HULPMIDDELS, OPERATIEWE FUNKSIE, GESPECIALISEERDE WOORDEBOEKE

1. Introduction

Lexicography, as a historical discipline, is characterised by its adaptability to the needs of users at any given moment in history. One of the most successful examples of this idea is the form of the traditional dictionary as we know it today: in alphabetical order and with the aim of including as many meanings as possible. This success was partly due to the ease with which the product could be accessed through its simplified content. For this reason and in order to make the knowledge of trade and commerce accessible to all users, as well as to strengthen the British Empire's knowledge thereof, Postlethwayt (1749: 2) decided to create an alphabetically structured dictionary:

A subject of this extensive nature therefore being reduced to the form of a Dictionary, for alphabetical reference, seems the most naturally adapted to answer.

Modifying or adapting the structure of lexicographic products has always been one of the main motivations of lexicographers in order to achieve the desired user-retrieved knowledge with each dictionary. Even more so in the case of specialised lexicography.

In this regard, the Theory of Lexicographic Functions which was first introduced in Bergenholtz and Tarp (2003) developed in the last decades a classification of four main lexicographical functions into which lexicographical products can be divided according to their purpose(s): communicative, cognitive, operative and interpretative. The main focus of this study is on the operative function, which together with the interpretative function is concerned with practical situations (Tarp 2010: 52). More specifically, the operative function is concerned with providing an answer to an individual who needs to solve a problem in which he or she has to act in a physical or mental way and, to do so, needs instructions on how to perform that specific act (Agerbo 2017: 363).

The first reference to the operative function in the specialist literature is in Tarp (2007: 177), where he points out that up to that time there had been many dictionaries and other reference products such as encyclopedias or manuals that were not designed to help in cognitive or communicative situations, but were works that provided instructions on how to act in very specific situations where the user was looking for instructions on how to perform some kind of

physical or mental act (e.g. how to use a machine, how to assemble a cabinet, or how to speak well in public). Tarp (ibid.) proposed that these situations should be called 'operative situations', as they had not been taken into account by any theory of lexicography up to that time, including function theory. There is no doubt that dictionaries and other reference products that include instructions do so as a mechanism of adaptation to their users and to the nature of the discipline covered. For this reason, these products share a common basis with traditional dictionaries:

They all have something fundamental in common with traditional lexicographic products, i.e. they are tools conceived to be consulted by specific types of users in order to satisfy specific types of information needs in specific types of social situations. In this respect, they are expected to provide quick and easy access to the relevant data from which the needed information can be retrieved (Tarp 2007: 177).

Tarp's (2008: 129) functional definition of a lexicographical tool is also transferred to operative tools, and he describes operative user situations as the third leg of the lexicographic field. Lexicographic tools that cover these situations, and thus have an operative function, are intended for people who need advice, guidance or instructions on how to perform a certain type of physical or mental act (Agerbo 2019: 84). It was only after Bothma and Tarp's (2013) study of the eighteenth-century dictionary of industry and commerce (*The Universal Dictionary of Trade and Commerce*) that the research about the existence of more dictionaries with an operative function was opened (Agerbo 2017: 362) — cf. Leroyer (2013), Bergenholtz et al. (2015), Agerbo (2015), although a thorough discussion about the nature of this function did not come to fruition. According to Agerbo (2019: 84), her survey of the lexicographic literature published between 2007 and 2016 found very little evidence of citations and definitions of the operative function through an internet search for the words *operation*, *operative* and *operational*. It should be noted that if this is a search in English, the concepts are much less noticeable in Spanish, since it is enough to google *lexicografía operativa* to see that there are no results, or *función operativa lexicografía* (as on 13 June 2020) to see that, among the 200 or so results, only four or five have to do with the discipline of lexicography from the perspective of function theory.

While operative function has not been extensively covered in lexicographic literature, certain authors have introduced or tried to analyse this function, thereby creating a novel and unexplored research avenue within meta- and practical lexicography. Only few authors have discussed the operative function in metalexicographical literature, and there are several examples of: (1) lexicographers who argue that operative needs (though they do not call them operative needs) do not belong in "dictionaries" — such as Sviķe (2018); and (2) lexicographers who have accommodated operative needs in the production of their tools, but who have not thought of these as operative needs and therefore do not refer to their tools as (particularly) operative tools. Examples of recent research in operative function can be found in Rodríguez Gallardo (2013, 2014) in the Spanish field and Agerbo (2019) in the English and Danish field, which point to

users' growing need for information that involves instructions (for example, how to use plants as medicinal remedies). However, this growing need is not always taken into account by the producers of lexicographic products, as in the case of Sviķe (2018: 238), who associates this type of need with a type of work that does not fall within the remit of lexicography: "these are not the characteristics of a multilingual dictionary, but of a practical manual." However, as Agerbo (2018, 2019) shows on a more global level, and as will be shown here from the perspective of Spanish lexicography, there are indeed dictionaries with an operative function.

2. The operative function in the Spanish context¹

The history of lexicography with its operative function cannot be understood without taking into account the need to disseminate specialised knowledge, which arose especially between the 18th and 19th centuries and was closely linked to the development of practical sciences in the world, such as electricity, agriculture and chemistry. This social need for specialised and practical knowledge led to the production of dictionaries and other reference works:

The dictionary, and in particular the encyclopaedic dictionary, proved to be the ideal instrument for cataloguing this knowledge: on the one hand, the alphabetical order made it possible to present it in a different way, more in keeping with the objectivity of scientific practice; on the other hand, its practical and synthetic aspect made it an attractive and very accessible reference work for the general public, especially for the uninitiated (Translated from Spanish from Moreno Villanueva and Madrona Cao 2004).

Lexicography is constantly changing and adapting to satisfy new (user) needs in new situations of practical use. This justifies not only versatility of dictionaries, but also blurring of the line that apparently separates them from manuals and handbooks, and also assimilation of manuals to dictionaries. An example of this trend is analysed by Moreno Villanueva (2017: 654) in his study of the Gallach collection of manuals (1899, 1915), copies of which from 1915 onwards included a vocabulary to compensate for the lack of technical terms. This study is an example of an attempt to revalue the manuals as a lexicographical reference tool. It is important to note that in the Spanish lexicographic tradition, reference works such as the *Manuales Soler* were not conceived as lexicographic tools, although they did have a lexicographic function:

From then on, the inclusion of these vocabularies in the form of small dictionaries became a new argument for promoting the collection and claiming its usefulness in comparison with the dictionaries in use. This was a line that had already been drawn when it was published as *Manuales Soler*. "Why is the acquisition of a MANUALES-SOLER collection more convenient than that of a dictionary?" could be read in the advertising of that first phase (Translated from Spanish from Moreno Villanueva 2017: 654).

When dealing with practical knowledge, it is inevitable to point out that knowl-

edge and skills are not completely isolated concepts; rather, there seems to be a transitional phase between the two. But as Fuertes-Olivera and Tarp (2014: 52) point out, not all types of knowledge can be transformed into skills:

[...] knowledge can be transformed into skills, while a systematic observation and study of the latter can lead to knowledge. The information acquired through consultation of dictionaries can be stored as knowledge but it cannot be directly transformed into skills. This information can assist users in performing specific mental, physical and linguistic tasks related to interpretive, operative or communicative situations and, in this way, it may be gradually internalised and reappear in the form of skills.

Knowledge is relevant to solve problems in operative situations, but it is important to distinguish between two types of knowledge (Agerbo 2017: 364): theoretical knowledge and practical knowledge. Of these two types, practical knowledge is the one that is related to the development of skills, which is the intended purpose of the operative function.

Thus, the Spanish dictionaries and other reference works analysed in this article are examples of the priority given to the operative function of lexicography, although this does not mean that this function appears in isolation from other lexicographic purposes.

3. Existing sport dictionaries and other reference works with an operative function

This section gives examples of existing dictionaries and other reference products with an operative function in Spanish. These analyses clearly show how the selected dictionaries are designed to help their users in operative situations. The dictionaries are presented in chronological order. For this purpose, Agerbo's (2017 and 2019) studies, in which she describes twelve dictionaries with an operative function in Danish, English and German, were used for guidance. Operative function-related lexicographic works follow a different architecture than standard dictionaries usually have, so that

[...] the analysis of their construction is more complex than that of products generated from communicative and cognitive situations. Operative products, such as instruction manuals or user guides, are organised on the basis of an information architecture centred on instructions and indications. They have an explanatory and applicative purpose (Translated from Spanish from Rodríguez Gallardo 2014: 320).

In order to carry out the analysis of the different selected works, three basic criteria which were established by Tarp (2018: 244) in his analysis of the criteria for defining the term *dictionary* were taken into account. These three criteria are

- (1) The objective (genuine purpose or function): it can be used to obtain concrete information about language or the non-linguistic world from specific lexicographical data.

- (2) Content: more specific than the data, it refers to the nature of the data included in the work, such as illustrations, symbols, etc.
- (3) Form: this refers both to the lemma of the work and to the access and form of the entries (definitions, explanations, instructions, etc.).

In Spanish lexicography, as in English and other languages, it has not been possible to distinguish dictionaries with an operative function from dictionaries with other functions by their titles, since they are all known as *dictionaries*. However, the nature of the data contained in these dictionaries shows that lexicographical works, since their inception, have included the data that users need to solve their lexicographical problems whether linguistic or not. The same applies to other reference products, such as handbooks or manuals, whose internal architecture is alphabetical and whose data access distribution is very similar to that of traditional dictionaries.

In this section three dictionaries and two technical manuals from the fields of equestrianism, fencing and football are analysed. They were chosen because they relate to practical fields, which means they meet the challenge of covering areas where practical knowledge is relevant and prevalent. The fact that they deal with subjects whose extra-lexicographic situations are practical will allow us to better understand the characteristics of the operative function. Several examples are offered for each reference lexicographic product. These were especially selected in order to identify and demonstrate the use or manifestation of operative function in the different reference products. The examples given were chosen arbitrarily to show that they could be identified relatively easily by simply opening each of these works. The result shows that instructions rather than definitions can be found in all the works analysed. A list of the works analysed in chronological order is shown in Figure 1.

Dictionaries and other reference products		
Year	Title	Author
1701–1800?	<i>Diccionario ecuestre</i> (Equestrian Dictionary)	Anonymous
1881	<i>Diccionario hípico y del sport</i> (Equestrian and Sport Dictionary)	Federico Huesca
1998	<i>Diccionario de fútbol</i> (Dictionary of Football)	Wolfgang Koch
1900	<i>Teoría y práctica de la esgrima</i> (Fencing Theory and Practice)	Pedro Carbonel
1913	<i>Novísimo tratado de Foot-Ball</i> (New Foot-Ball treatise)	Georges Graham

Figure 1: Sport dictionaries and other reference products analysed

3.1 *Diccionario ecuestre* (1701–1800?)

The 18th-century *Diccionario ecuestre* by an anonymous author is a 291-page manuscript that, according to the National Spanish Library, dates from 1701 to 1800. It is the first documented equestrian dictionary in Spanish and is entirely handwritten. The only description of its intended users is a brief introduction which reads: *Diccionario ecuestre o exacta definición de los términos del jinete, del albéitar y del herrador para completa instrucción del caballero y de todo hombre a caballo* ("Equestrian dictionary or exact definition of the terms of the rider, the groom and the blacksmith for the complete instruction of the rider and every man on horseback"). The words *definition* and *instruction* stand out in this introduction, which might lead us to believe that the dictionary could be polyfunctional. However, the function of the dictionary is not clearly stated. Since it is an 18th-century dictionary, from the age of Enlightenment, it would not be surprising if its function was to compile the meaning of these words, with no other pretension than that of being a compendium of this specialist field, in order to valorise and extend it over time.

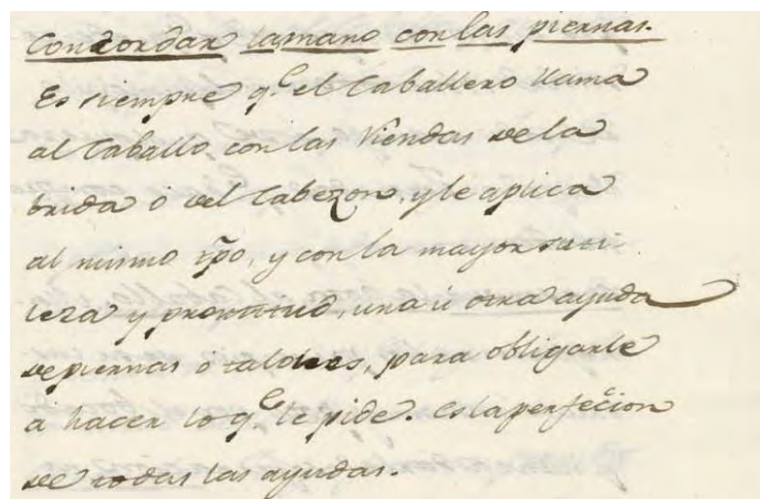


Figure 2: Example of 'Concordar la mano con las piernas' (How to match the hand to the legs) in the *Diccionario ecuestre*

One of the noticeable features offered by this dictionary is the form of the lemmas, which are compound lemmas (clauses that are often equivalent to actions), such as the example in Figure 2 (*How to match the hand to the legs*). Since the function that can be deduced from the examples analysed is communicative, the purpose of the dictionary does not seem to go beyond the intellectual demands of the time. In this respect, the absence of linguistic data shows that specialised

lexicography has been present in different contexts in which there was no apparent need for it. However, it can be deduced that there is indeed a deliberate aim to provide instructions for users at the entry of *How to bridle a horse*:

Poner la brida al potro. Entiéndese en término de amarre por la primera vez que se pone la silla al caballo debe hacerse estas operaciones con mucho tiempo y paciencia para no ostigarle porque si en los principios se le exapera puede quedarse [exaltado] para siempre: Por eso combiene que la primera vez que se pone la silla al potro sea en la caballeriza o en la cuadra ... (Diccionario ecuestre, original transcription: 323-324]

How to bridle a horse: As far as harnessing is concerned, it is clear that the first time the saddle is put on the horse, these operations must be carried out with a lot of time and patience so as not to damage the horse, because if it is exaggerated at the beginning, it can remain [exaggerated] forever: For this reason, it is advisable that the first time the saddle is put on the foal, it should be in the stable or in the stall ... (English translation from Spanish)

First of all, the compound lemmas take the form of a clause in addition to the fact that they are practical actions. In this case, the article has a clear operative approach since it does not simply define what is meant by putting the saddle on the foal, which would correspond to the first line, but goes on to give an explanation based on advice and instructions (two types of needs that occur in essential operative situations). Examples such as these are repeated throughout the work and show that the *Diccionario ecuestre* (1701–1800?) is probably one of the first Spanish dictionaries to incorporate an operative function of lexicography in addition to the communicative one.

3.2 *Diccionario hípico y del sport* by Federico Huesca (1881)

The *Diccionario hípico y del sport* by Federico Huesca (1881) is one of the first sport dictionaries (as we understand the term today) in Spanish and consists of 759 pages. From the very beginning, in the "Preface", the author of the dictionary confesses that: *No hemos tenido la pretensión de hacer un libro de enseñanza, sino simplemente de exposición de doctrinas, de descripción de ejercicios, de compilación de técnica de datos estadísticos y algunos bibliográficos* ("We did not have the pretension of making a textbook, but simply of exposing doctrines, describing exercises, compiling statistical data and some bibliographical data") (1881: 5). There is a direct reference to a *descripción de ejercicios* (description of exercises), although it is not clear whether this is a list of instructions. The function of the dictionary seems to be cognitive, which in function theory means that it is focused on describing non-linguistic data, in other words, giving specialised explanations. The reason for this is that a dictionary with a cognitive function transmits knowledge (about something) to its users when they want to acquire knowledge, in the same way as dictionaries with communicative functions provide help in ongoing or planned communicative situations. One clear example of

this can be found in the entry of 'Lagrange (conde de)', where we find an extensive article with explanations about the life of an important horseman:

[...] *Toda su vida se ha ocupado de carreras y de caballos; pero hasta 1857 no se hizo propietario de una caballeriza importante. En esta época compró á Mr. Alejandro Aumont todos los caballos que tenia en el entrainement, entre los cuales figuraba Monarca. [...]* (*Diccionario ecuestre*, original transcription: 410)

[...] All his life he had been occupied with racing and horses; but it was not until 1857 that he became the owner of an important stable. At that time he bought from Mr. Alexandre Aumont all the horses he had in entrainement, among which was Monarch. [...] (English translation from Spanish)

While the entries that deal with actions are concerned with offering specialised knowledge, which means that entries offer explanations rather than linguistic definitions or instructions, the dictionary also offers instructions and recommendations which, together with specialised explanations, cover an operative need that is not very well defined in the initial 'warning' prologue. This is the case of the entry on *alfalfa* (lucerne), which not only defines it as an important fodder plant for horses, but also explains in detail the conditions for its satisfactory cultivation by the user. Thus, we find an extensive lexicographical article with paragraphs very well divided by subject, and its opening shows us its instructive and proscriptive intention, *Esta planta es de terrenos frescos [...]* ("This plant is a plant for cool soils [...]"), *Exige terreno indispensablemente calizo [...]* ("It requires a soil that must be chalky [...]"), *La alfalfa se siembra á boleó [...]* ("Lucerne is sown on the ground [...]"), *Si la siembra se verifica al principio del otoño [...]* ("If it is sown at the beginning of autumn [...]"), *Puede consumirse la alfalfa ó en el prado ó en el establo [...]* ("Lucerne can be eaten either in the meadow or in the barn [...]"), *La alfalfa debe regarse siempre en flor [...]* ("Lucerne should always be watered when in flower [...]") (1881: 31).

As we can see, the characteristics of this type of instruction do not correspond to the usual model for expressing instructions through the imperative or infinitive form in Spanish, but the dictionary uses a more indirect structure, despite the use of the impersonal pronoun 'se', which is undoubtedly used to give instructions in Spanish too. The same happens with the entry for *forrajes* (fodder). The first part is rather non-linguistic, with explanations that offer specialised information on the benefits of this type of feed for young and even sick horses: *Los forrajes á pico son muy beneficiosos, porque adelanta el pelecho del ganado y son purgantes* ("The fodder is very beneficial, because it helps the cattle's fodder to grow and is a purgative"). However, as the definition progresses, the content begins to focus more on 'how to do it', completely abandoning the definition itself. It begins to explain where the user should feed the horse in order to obtain the aforementioned benefits, how and when to feed it, and finally, as if it were a recipe, the exact amount per horse is given: *El forraje debe darse picado para que no se desperdicie, y conducirlo en espuertas á los artesones. Las pasturas será de*

poca cantidad [...] ("The feed must be chopped so that it is not wasted and carried to the troughs in baskets. The amount of pasture should be small [...]).

Other examples of the operative function being employed can be found in the entries *aciones* (belts) and *adelgazar* (slimming). In the case of *aciones*, what is observed is not a sequential component of instructions, but a full-blown proscription. After a very brief definition of the "what", recommendations are made on how to carry out the actions and their immediate consequences. To this end, evaluative expressions are used such as *si se rompe una* [...] ("if a [...] breaks"), expressions of possibility with *puede suceder* ("it may happen"), with *se procura* [...] ("we will try to [...]") or even with circumstantial cause *para que* [...] ("so that [...]"), interspersed with the impersonal pronoun "se". All this shows that the recommendations are based on real cases experienced by the specialists themselves. In addition, the entry ends with another example of proscription, but this time to solve specific problems that steeplechase riders may have.

To end the analysis, it is necessary to look at the entry of *adelgazar* (slimming), which refers to an action that does not affect the horses but the riders. In this case it is used to give instructions. The entry states that *Para evitar el engordar en todo tiempo, pero principalmente en el invierno, se ideó un sistema para adelgazar, parecido al que se emplea en la preparación de los caballos* [...] ("In order to avoid getting fat in all weather, but mainly in winter, a system was devised for slimming, similar to that used in the preparation of horses") (1881: 12). The operative process being referred to is always shrouded in specialised explanations, as it usually happens in all the entries regarding riders and its direct actions. Even so, there is a very clear intention to offer some indications that, if the user is a horseman, he will find very useful in order not to put on weight. In addition to the instructions and recommendations offered, other data explanations are also provided that confirm the implementation of this method in England and France.

In conclusion, the dictionary has an operative function, although this function does not manifest itself clearly in all entries, but only in those entries which, by their practical nature, require instructions to the user. From the examples given, it is clear that this is not a reference book for learning about horsemanship in an operative sense. However, you can learn how to prepare food for horses and how riders should look after themselves. It is a good example of how instructional data can be interwoven with the dictionary definition to satisfy different kinds of lexicographical needs. However, this combination only occurs in lemmas that have no direct connection with horses or the use of horses in the development of the sport, leaving the operative function focused on giving answers to equestrian practice.

3.3 *Diccionario de fútbol* by Wolfgang Koch (1998)

Koch's dictionary is a translation of the German *Fussball Von A–Z, Begriffe, Fakten, Regeln* by Wolfgang Simon. This dictionary is unique for several reasons. It is

one of the first dictionaries to focus on the Spanish football lexicon (even though it is a translation) and it is also a dictionary that, because of the nature of the lexicon it contains, uses strategies that are relevant to this research.

The dictionary consists of a short introduction called *Instrucciones para el lector* ("Instructions for the reader"), which gives information about the lemmas, symbols or other conventions for using the dictionary. There is no explanation of the target audience or the function(s) of the dictionary. After the brief instructions for the reader, there is an index of signs, the football dictionary itself, bibliographical references (in German), an appendix with a series of questions on the rules of the game, and finally an anthology of the development of the game and the rules, with dates. This dictionary contains parts that were unusual for a dictionary published at a time when there were not so many works that deviated from the usual lexicographical canon (dominated by linguistic theory). Firstly, the dictionary contains lemmas composed of both simple and compound units. For example, *táctica de cobertura* (hedging tactics) is lemma-tised as a unit and not under the same entry as *táctica* (tactics) (although this is probably due to the direct translation from German). Furthermore, this dictionary prefers *explicación* (explanation) to *definición* (definition), as stated in the introduction:

En los párrafos explicativos se ha prescindido expresamente de la máxima precisión científica, prevaleciendo la explicación por delante de una definición con el fin de conseguir la máxima comprensión (In the explanatory paragraphs, maximum scientific precision has been deliberately avoided, preferring to explain rather than define in order to achieve maximum understanding) (Koch 1998: 5).

This means that the dictionary transcends the established limits based on linguistic background and has a certain specialised character, as also highlighted by Nomdedeu Rull (2004: 192), who comments on the same dictionary and its definition of *árbitro* (referee), which is not only extensive, but also intertwined with the rules of football. Another example of this type of specialised definition is the one that appears in the *calentamiento* (warm-up) entry, where the field of sports medicine takes on special importance. As Nomdedeu Rull (2004: 195-196) points out, we are dealing with a dictionary which, apart from its specialised nature, sometimes gives the reader the impression that he/she is reading a football manual, because the operative function can be deduced from the type of definitions. This feature blurs the barrier that has always existed between dictionary-encyclopaedia-manual.

An example of this is the *barrera* (barrier) entry, where recommendations are made after the definitions, at a time when instructions are given on how to act: *Si el adversario emplea variantes del tiro libre, es preciso reaccionar rápidamente de manera que se bloquee un tiro a puerta (salir)* ("If the opponent uses variants of the free kick, it is necessary to react quickly in order to block a shot on goal (exit)").

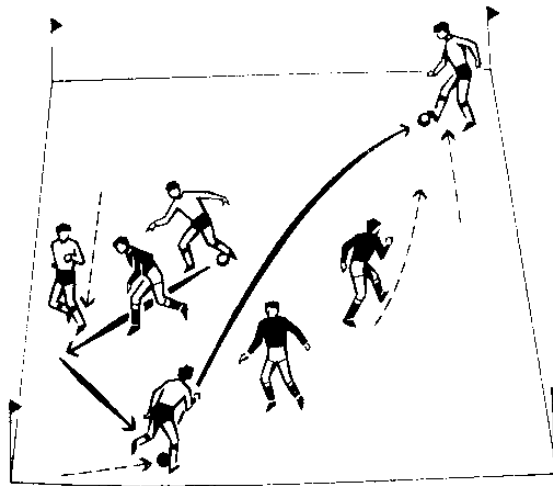


Figure 3: Picture of *ejercicios de marcaje-desmarque* (marking-unmarking exercises), *Diccionario de fútbol* by W. Koch (1998: 73)

This is not the only example in which Koch's football dictionary goes beyond the strictly lexicographical (according to the canon), since the use of instructions and explanations is generally manifested in two ways: (1) through illustrations that complement the explanations of the lemmas, as in the case of the *ejercicios de marcaje-desmarque* (marking-unmarking exercises) (see Figure 3); or (2) the definition of ball reception is given in a way that is typical of instructional texts, where resources such as numbers are used to sequence different types of ball reception, and illustrations are used to complete the explanation (see Figure 4).

The illustrations undoubtedly have a clear instructive purpose, especially since, in the case of football, an explanation like 3. *Recoger un balón que viene rodando, botando o volando a nivel del suelo o en el aire* ("3. Catching a ball that is rolling, bouncing or flying at ground level or in the air") may be clear to an expert or even semi-expert on the subject without the need for illustrations, but it will not be clear to someone who has no knowledge of this sport. In this case, there is not even an explanation of how to do it, only the type of ball reception is highlighted, while the illustrations show how to act in each case.

Recepción del balón: Forma importante de obtener el control del balón. Sirve para asegurar balones que provienen del compañero o del adversario y se caracteriza por la acción de ceder (amortiguar) que efectúa la parte del cuerpo que recibe el balón. Al ceder, le sigue a ser posible un movimiento contrario. La técnica de la R. se diferencia de la siguiente manera: 1. Detener un balón que viene rodando. 2. Parada del balón a nivel del suelo o en el aire. 3. Recoger un balón que viene rodando, botando o volando a nivel del suelo o en el aire. La → recogida del balón adquiere cada vez mayor importancia ("técnica rápida", ganar espacio/tiempo). Véase también → *parada*.

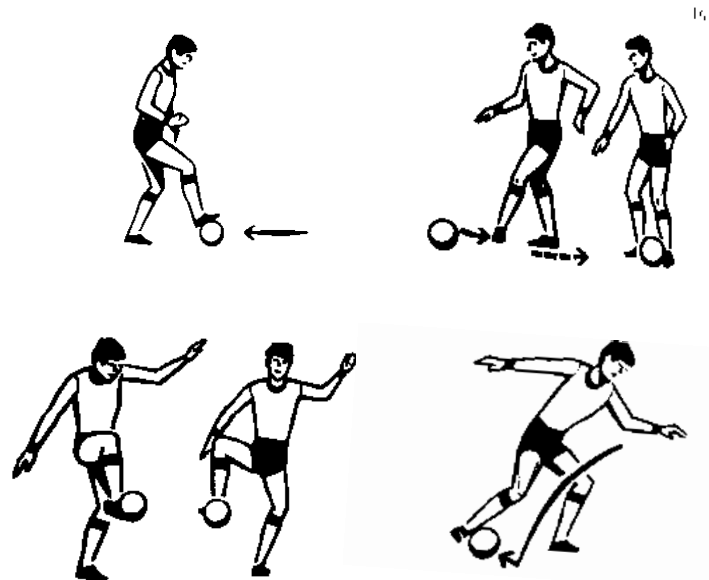


Figure 4: Pictures in the *recepción del balón* (ball reception) entry, *Diccionario de fútbol* by W. Koch (1998: 160)

3.4 *Teoría y práctica de la esgrima* by Pedro Carbonel (1900)

Teoría y práctica de la esgrima (1900) consists of two different editions (850 standard copies and 150 deluxe copies) produced by King Alfonso XIII's Master of Arms, Pedro Carbonel. The book analysed is dedicated to the Marquis of Heredia, whose reply to the dedication is also included. Thanks to this introduction, we know that the work contains useful advice, the function of which, as the Marquis states, is educational: [...] *donde encontrarán los fuertes tiradores gran número de avisos y consejos útiles, y las naturalezas más rebeldes la seguridad de vencer las dificultades y llegar, por medio de tu hábil enseñanza, á conseguir lo que les parecía inaccesible* (1900: X) ("[...] where strong marksmen will find a great number of useful warnings and advice, and the most rebellious natures will find the security to overcome difficulties and, by means of your skilful teaching, to achieve what seemed inaccessible to them"). In the prologue, the author remarks on the importance of practice and theory, whose function is *facilitar con la teoría la práctica de este arte, que, si bien como sport es uno de los más convenientes, considerado como lucha reviste mayor importancia ...* ("to facilitate with theory the practice of this art, which, although as a sport it is one of the most convenient, considered as a fight is of greater importance ..."). (Prologue, 1900: XII).

The manual is divided into four parts. The first two parts begin with a technical glossary whose macro-structure is that of a traditional dictionary, as it contains the essential vocabulary in alphabetical order, but without linguistic

information (Figure 5). Parts three and four deal with fencing movements and techniques, not in alphabetical order, although in this case they are arranged in order of difficulty (from the easiest to the most difficult) and follow a very heterogeneous structure. Each entry consists of a motto based on an action, and the data are instructions interwoven with some advice and recommendations (proscription) and specialised explanations. In addition, these instructions are supported by images that help the user understand how to proceed (Figure 6).

CAPÍTULO PRIMERO

VOCABULARIO TÉCNICO

A

Abrir la guardia.—Es tener el pie derecho más á la derecha de la línea en que debe estar cuando se tiene dicha posición. Es lo contrario de *cruzar la guardia*.

Abrir ó romper la distancia.—Es, estando en la posición de guardia, separarse del adversario por medio de uno ó más pasos para ponerse fuera de su alcance.

Abrirse.—Se emplea este término cuando al caer en guardia ó partir á fondo se hace exageradamente.

Acostarse.—Es inclinar el cuerpo exagerada-

Figure 5: Glossary in *Teoría y práctica de la esgrima* (1900)

Manera de empuñar el sable.

Éste debe sujetarse bien en la mano, extendiendo el pulgar en la parte superior del puño; este dedo, que podemos considerar como el timón del sable por ser con el que le damos *di-*



reción, hay que procurar que no toque la guardia, antes, por el contrario, separado unos dos centímetros; los otros cuatro dedos, unidos y abarcando bien el puño.

Figure 6: Entry of *Manera de empuñar el sable* (Way of wielding the blade) (1900: 118)

The third part focuses on pedagogical dimension of teaching fencing and is clearly aimed at instructors or teachers. It is therefore a clear example of "how to teach", going beyond the "what to teach" of the first parts. The third part begins with a section on *Advertencias y preliminares de la enseñanza* ("Warnings and Preliminaries of Teaching"), with extensive advice on class management, followed by various lessons (e.g. the first foil lesson or finger exercises), preceded by a lemma and pedagogical instructions ranging from the way the room is divided to the different subjects to be taught. The fourth part focuses on the protocols and rules to be followed during the attacks, as well as advice that can help the correct development of the activity. It is interesting to note the specificity of some of the situations dealt with in this last part, such as *De los zurdos y manera de contrarrestar su juego* ("Of the left-handed and how to counteract their game") (1900: 253), which demonstrate the accuracy and variety of the situations covered.

In conclusion, *Teoría y práctica de la esgrima* by Pedro Carbonel is a polyfunctional manual that covers different types of situations on different levels: it gives precise instructions on how to fence, and it also gives instructions on how to teach it; it incorporates quite frequently and in very different ways the resource of proscription (recommendations). The manual is an example of a hybrid work that aims to cover very different needs in very different situations, and it also follows the structure of a traditional dictionary (it has alphabetical structure with easy access for consultation).

3.5 *Novísimo tratado de Foot-Ball* by Georges Graham (1913)

Georges Graham's *Novísimo tratado de Foot-Ball* is one of the first documented works on football in Spanish. The title page states that it is a *Método práctico para jugar al FOOT-BALL y apreciar la licitud y oportunidad de las jugadas* ("Practical method for playing football and appreciating the legality and opportunity of the plays"), so this subtitle seems to anticipate the function of the handbook: *ilustrado con numerosos grabados* ("illustrated with many engravings or illustrations"). The book, which is only 48 pages long, is divided into two parts: the first, in five chapters, introduces the reader to the sport and its benefits; the second, also in five chapters, deals with the technique of the game.

The first part of the guide has a didactic approach and is divided into chapters with explanations aimed at satisfying the needs for knowledge (in function theory, this would be called a *cognitive* lexicographical tool). This first part could be classified as a football encyclopaedia, with the exception of *CAPÍTULO CUARTO. Entrenamiento del jugador* ("Chapter four. Player training"), in which the operative function is introduced as a natural development of the cognitive progress of the previous chapters. In this chapter we can identify what Rodríguez Gallardo (2014: 323) calls *degrees of operativeness* in relation to an operative situation, which in this case deals with two situations: (1) *Ejercicios de gimnasia y desarrollo de la fuerza muscular* ("gymnastic exercises and development of mus-

cular strength"), and (2) *Ejercicios técnicos* ("technical exercises"). These degrees classify the operative situations to be solved according to their specificity which means that each concrete situation requires specific instructional data. They include an introduction by definition or explanation, followed by instructions of varying difficulty, adapted to the way in which the information is accessed: written and illustrated (Figure 7).

The written instructions are numbered and preceded by a short lemma, which in the case of Figure 7 is *Detener la pelota* ("Stop the ball"). In addition, instructions for different types of sub-situations can appear within the same entry, as on page 25: *estudiemos dos casos que pueden ofrecerse* ("let's study two possible cases"). As for the illustrations, they support what is written, as can be seen in Figure 7.

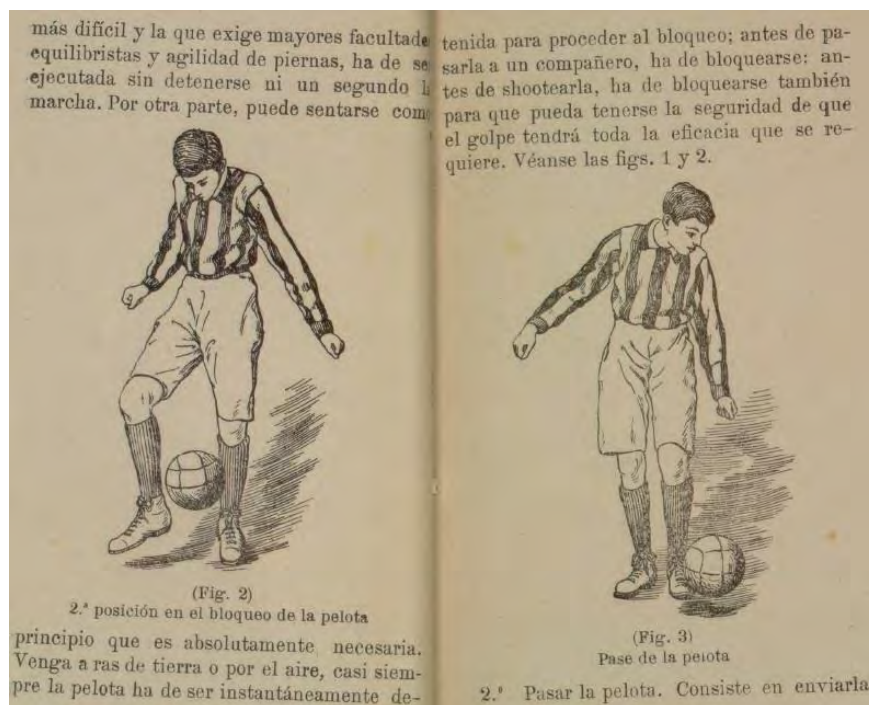


Figure 7: Illustrations in *Novísimo tratado de Foot-Ball* (1913: 28-29)

In the second part of the manual, the tool is designed to meet the user's operative needs. From the first chapter, which focuses on the rules of the game for the general public, instructions are used and a very clear lexicographical structure is followed: starting with a lemma and its lexicographical article with the applicable rule. The second chapter, on the other hand, has a different structure, avoiding the introduction of new lemmas in order to focus on the team

captain. However, unlike the explanatory chapters of the first part, in this case the whole chapter has a clear operative function, but applied to a new type of user, need and situation, which mediates between knowledge and the teaching of an action. The complexity of its typology of users and situations increases in the third chapter, which introduces the figure of the referee and the need to resolve "some dubious plays" such as *offside*, with up to ten explanations of how to proceed and illustrated examples (see Figure 8).

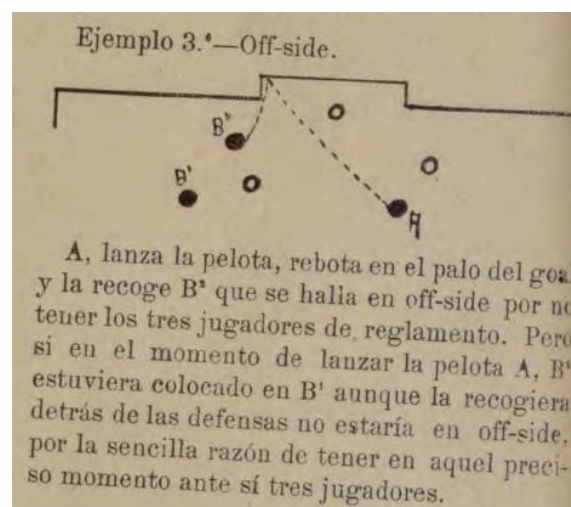


Figure 8: Example of explanation of *offside* (1913: 64)

Chapter four continues the tendency to follow a complex lexicographical structure by again introducing lexicographical entries preceded by a lemma, this time to introduce the profile of the players of the team and what are the *cualidades que debe poseer cada uno* ("qualities that each one must possess") (1913: 69). The last chapter of the second part is formulated as a kind of commandment on how to act according to the different sub-situations. This takes the form of an article, without entries or lemmas, but in which recommendations predominate.

To summarise, it can be said without a doubt that Georges Graham's handbook has an operative function, although it is a polyfunctional work, as is usual in this specialised context, since it also has a cognitive function. However, what is truly characteristic of this work is its desire to bring together many different types of users and situations — and sub-situations — in a single short volume. At least four different types of users can be identified: the general public (most of Part I and Chapter I of Part II), the team captain (Part II, Chapter 2), the referee (Part II, Chapter 3) and the players (Part II, Chapters 4 and 5). The weakness of the handbook is precisely the lack of organisation to ensure that the different types of users can easily access the data.

4. Main characteristics of the Spanish sport works with an operative function

The lexicographical products analysed, which all have an operative function, share different characteristics:

- (1) They are all polyfunctional.
- (2) They usually have a longer lemma (or 'action'), such as *acortar los estribos* (how to shorten the stirrups).
- (3) They include different types of data, such as instructions, explanations and definitions, in order to meet lexicographic needs arising from operative situations.
- (4) They have a complex lexicographical data structure, sometimes not following a coherent pattern, in order to facilitate user access.
- (5) They are aimed at solving reference needs that are practical and therefore require the acquisition of skills that can only be performed on the move.

In addition, this study shows that there is a number of similarities between works of a different typology (but similar subject matter — sport) that have an operative lexicographic function. One remarkable similarity is the access to data, which is organised assuming a lexicographic structure. This means that in most of them data access distribution is very similar to that of traditional dictionaries (alphabetically structured). After the analysis, it can be concluded that reference tools with an operative function are consultation tools for specific (social) situations that focus on the need to develop a practical skill.

It is important to highlight that in sport reference tools with an operative function access to and distribution of the data is more complex and not governed by homogeneous criteria, which often makes it difficult for users to obtain information. Both dictionaries and reference works analysed, have in common that they have more than one lexicographic function, even if this is not made explicit at first sight. The fact that a single work could have different functions (in many cases operative and cognitive) shows that even in the past there was an interest, unconscious or not, in achieving a greater degree of individualisation.

The reference works from the early days of the sport (dictionaries and other reference works) gave priority to the cognitive function because the need for information at that time was focused on understanding the newly emerging sport (theoretical knowledge). Their real aim was to explain the basics of the sport before teaching how to play it. Depending on the model, works/tools with an operative function seek different types of individualisation, which is determined according to the characteristics of the users (needs and situations) they are designed for. In fact, some of them cover not only different consultation needs and situations, but also different types of users (experts and laypeople). An example of this can be found in Carbonel's fencing reference work, where there is a pedagogical need that is closely linked to the didactic component of

the task — "how to teach". This pedagogical-operative need belongs to the coach user, which means that different types of user can be covered by a single reference work.

Finally, sport operative products analysed usually cover a procedural need for "how to act" (through practical knowledge), which is not so common for an expert user, and also the need for understanding the different types of situations and conditions in each sport (through theoretical knowledge). This means that data is presented in different ways depending on the function/s of each product. These explanations are constructed to respond to the specific needs of users in specific situations: definitions, specialised explanations or instructions can be intercalated to fulfil the function of each product type (poly-functional or pluri-monofunctional).

5. Discussion

Taking into account the limitations of the tools analysed, a roadmap can be established for the development of future tools with an operative function in today's world. Thus, it is important to emphasise that the needs of the users of these tools are linked to practical knowledge that leads them to develop practical skills. In this respect, it will be important to carry out a prior study of the potential user's profile, whose needs are linked to the practical situations in which they will need to use the tool. The data to be included in new tools of this type should therefore correspond to a functional approach, which means that the type of explanation to include is adapted to the specific needs of the user in the relevant lexicographic situations. This also means that the type of article — be it a definition, a specialised explanation or instructions — is chosen on the basis of the function(s) it is intended to cover. One of these functions will undoubtedly be operative, but this should not be understood as incompatible with sharing space with other functions, such as cognitive functions, as demonstrated in previous analyses. For example, the basic football instructions for an inexperienced young learner will not be the same as those for a football coach seeking to improve; although the situation is operative and the sub-situations that arise are also likely to coincide, the profile means that the type of data must necessarily change in order to fulfil each function. In order to address this, a pedagogical operative function ("how to teach") must have a clear didactic character, and to fulfil this purpose, the data it contains must be adapted to the specific functional and retrieval characteristics of this operative situation. This process is reminiscent of Pétroff's (1984: 66-67) so-called "reformulation", because it involves the adaptation of scientific and technical discourse to the intended user. Reformulation consists of adapting data to a new situation, to a specific type of recipient and in accordance with a given action. According to this, all categories and units from a lexicographic article, both linguistic and graphic, are susceptible to reformulation.

Finally, to guarantee a good data structure in operative function products,

it is useful to include all the main categories and all the concepts significant to the topic and the individual user needs. The reference works analysed are generally not organised according to consistent criteria or their structure is not clear, which means that users often get lost in a maze of mixed-up data. One solution for new tools is to follow the principle of intuitive design for each of the categories so that they can be understood globally and rationally.

6. Conclusions

The purpose of this article has been to analyse a series of sport dictionaries and other reference works in Spanish that have an operative function without them explicitly stating it. These analyses have demonstrated the existence of the operative function in Spanish lexicography, even if the tools have not explicitly stated this as their purpose. However, this is only a small part of what is known about this practical dimension of lexicography, which is still insufficiently studied. Throughout the history of lexicography, the presence of instructions to meet operative needs is common in several languages, as demonstrated by Agerbo (2017) for English, German and Danish. Similarly, the long-standing lexicographic tradition of Spanish is not different. Agerbo (2017: 387) states that her work aims to "convince lexicographers that the unexplored but highly relevant operative function deserves their attention, both in their meta-lexicographical and practical work", and thanks to this, a completely new line of research has been opened for the historical lexicography of Spanish. Studies about the operative function will lay the foundation and catalogue an unexplored function of lexicography that is still relevant for the development of future lexicographic tools related to sport and other practical subjects.

Endnotes

1. The quotes from Spanish authors have been translated into English by me. The original source can be found in the references section.

References

A. Dictionaries and other reference works

- Anonymous.** 1701–1800? *Diccionario ecuestre*. Madrid. Available from the Biblioteca Digital Hispánica and Biblioteca Nacional de España:
<http://bdh-rd.bne.es/viewer.vm?id=0000062278&page=1>
- Carbonel, P.** 1900. *Teoría y práctica de la esgrima*. Madrid: Suc. de Rivadeneyra. Available from the Biblioteca Digital Hispánica and Biblioteca Nacional de España:
<http://bdh-rd.bne.es/viewer.vm?id=0000084352&page=1>

- Graham, G.** 1913. *Novísimo tratado de Foot-ball. Método práctico para jugar y apreciar la licitud y oportunidad de las jugadas*. Barcelona: Ciencias y Letras.
- Huesca, F.** 1881. *Diccionario hípico y del sport*. Madrid: J. M. Pérez. Available from the Biblioteca Digital Hispánica and Biblioteca Nacional de España:
<http://bdh-rd.bne.es/viewer.vm?id=0000084341&page=1>
- Koch, W.** 1998. *Diccionario de fútbol* (Spanish translation from German *Fussball Von A–Z, Begriffe, Fakten, Regeln* a cargo de Wolfgang Simon). Barcelona: Editorial Paidotribo.

B. Other literature

- Agerbo, H.** 2015. How to Describe Sports Terms in Information Tools for Laymen. *Estudios de Lexicografía* 4: 137-160.
- Agerbo, H.** 2017. Monofunctional and Polyfunctional Information Tools with an Operative Function. *Lexicographica* 33: 361-390.
- Agerbo, H.** 2018. Explaining Meaning in Lexicographical Information Tools. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 59-77. London: Routledge.
- Agerbo, H.** 2019. *A Theoretical Description of the Operative Function and its Implementation in the Conceptualisation of Danish Lexicographical Tools in Physical Training*. Unpublished Doctoral Dissertation. Aarhus: Aarhus University.
- Bergenholtz, H., T.J.D. Bothma and R.H. Gouws.** 2015. Phases and Steps in the Access to Data in Information Tools. *Lexikos* 25: 1-30.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Language and Communication in Business* 31: 171-196.
- Bothma, T.J.D. and S. Tarp.** 2013. An Alternative Approach to Enlightenment Age Lexicography: The Universal Dictionary of Trade and Commerce. *Lexicographica — International Annual for Lexicography* 29(1): 222-284.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Lexicographica Series Maior 146. Berlin/New York: De Gruyter.
- Leroyer, P.** 2013. New Proposals for the Design of Integrated Online Wine Industry Dictionaries. *Lexikos* 23: 209-227.
- Moreno Villanueva, J.A.** 2017. Los Manuales Gallach: materiales para la historia de la lexicografía especializada. Garriga Escribano, C., I. Sariego López and J. Gutiérrez Cuadrado (Eds.). 2017. *El diccionario en la encrucijada: de la sintaxis y la cultura al desafío digital: actas del VII Congreso Internacional de Lexicografía, Santander, 27–29 de junio de 2016*: 647-664. Cantabria: Escuela Universitaria de Turismo Altamira.
- Moreno Villanueva, J.A. and A. Madrona Cao.** 2004. Los primeros diccionarios de electricidad en español: el Diccionario de electricidad y magnetismo (1893) de Lefèvre y el Diccionario práctico de electricidad (1898) de O'Connor Sloane. Battaner, P. and J. DeCesaris (Eds.). 2004. *Actas del I Symposium internacional de lexicografía, Barcelona, 16–18 de maig de 2002*: 1000-1013. Barcelona: Universitat Pompeu Fabra, Documenta Universitaria.
- Nomdedeu Rull, A.** 2004. *Terminología del fútbol y diccionarios: elaboración de un diccionario de especialidad para el gran público*. Unpublished Doctoral Dissertation. Barcelona: Universitat Autònoma de Barcelona.

- Pétroff, A.J.** 1984. Sémiologie de la reformulation dans le discours scientifique et technique. *Langue française* 64: 53-67.
- Postlethwayt, M.** 1749. *A Dissertation on the Plan, Use, and Importance, of the Universal Dictionary of Trade and Commerce; Translated from the French of the Late Celebrated Mons. Savary, Inspector-General of the Manufactures of the Custom-House of Paris: With such Considerable Additions and Improvements, As will Appear at Large Throughout this Dissertation; And which More Particularly Accommodate the Fame to the Trade and Navigation of the British Empire, Addressed to the Nobility, Gentry, Merchants, and Traders of Great Britain.* London: John and Paul Knapton.
- Rodríguez Gallardo, Á.** 2013. The Function Theory and Its Application on Manuals of Economics. *Hermes, Journal of Language and Communication in Business* 50: 83-95.
- Rodríguez Gallardo, Á.** 2014. Situaciones sociales y lexicografía operacional. Garcés Gómez, M^a.P. (Ed.). 2014. *Lexicografía teórica y aplicada.* Anexos Revista de Lexicografía 26: 317-330. A Coruña: Servizo de Publicacións Universidade da Coruña.
- Sviķe, S.** 2018. A New Dictionary of Botanical Terms: Data Analysis of a Lexicographic Survey. *Economics World* 6(3): 228-241.
- Tarp, S.** 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Tarp, S.** 2008. The Third Leg of Two-legged Lexicography. *Hermes* 40: 117-131.
- Tarp, S.** 2010. Functions of Specialised Learner's Dictionaries. Fuertes-Olivera, P.A. (Ed.). 2010. *Specialised Dictionaries for Learners*: 39-53. Berlin: De Gruyter.
- Tarp, S.** 2018. The Concept of Dictionary. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 237-249. London: Routledge.

The Language of Ethnic Conflict in English Online Lexicography: Ethnophaulisms in "powered by Oxford" Lexico.com

Silvia Pettini, *Department of Foreign Languages, Literatures and Cultures, Roma Tre University, Rome, Italy* (silvia.pettini@uniroma3.it)

Abstract: This article aims to explore the relationship between the language of ethnic conflict (Allen 1983; Palmore 1962) and English online lexicography in the present cultural moment. Given the influence of the Internet on dictionary consulting (Béjoint 2016; Jackson 2017) and the alarming increase of racism and xenophobia, especially online, at the global level in this digital age (see Gagliardone et al. 2015), this article presents a pilot study examining the treatment of "ethnophaulisms" (Roback 1944), commonly referred to as ethnic or racial slurs, in the "powered by Oxford" dictionary content, which is licensed for use to technology giants like Google, Microsoft, and Apple by Oxford University Press (Ferrett and Dollinger 2021; Pettini 2021). In particular, the analysis focuses on the online edition of the *Oxford Dictionary of English*, as hosted on the "powered by Oxford" Lexico.com website. Preliminary findings show how this free online dictionary mirrors the taboo and discriminatory nature of ethnophaulisms and warns the Internet user against the derogatory and offensive power of these words.

Keywords: ENGLISH LEXICOGRAPHY, ONLINE DICTIONARIES, LINGUISTIC RACISM, XENOPHOBIA, HATE SPEECH, ETHNIC SLURS, ETHNOPHAULISMS, OXFORD DICTIONARY OF ENGLISH, LEXICO.COM

Opsomming: Die taal van etniese konflik in die Engelse aanlyn leksikografie: Etnofaulisme in die "Oxford-gedrewe" Lexico.com. In hierdie artikel word gepoog om die verhouding tussen die taal van etniese konflik (Allen 1983; Palmore 1962) en die Engelse aanlyn leksikografie soos op hierdie kulturele moment te ondersoek. Gegewe die invloed van die internet op die raadpleging van woordeboeke (Béjoint 2016; Jackson 2017) en die, veral aanlyn, kommerwekkende toename van rassistiese en xenofobiese gevalle op globale vlak in hierdie digitale era (sien Gagliardone et al. 2015), word daar in hierdie artikel 'n loodsstudie aangebied waarin die hantering van etnofaulisme (Roback 1944), waarna meestal verwys word as etniese of rassistiese beledigings, in die "Oxford-gedrewe" woordeboekinhoud, wat deur Oxford University Press gelisensieer is vir gebruik deur tegnologieuse soos Google, Microsoft, en Apple (Ferrett en Dollinger 2021; Pettini 2021), ondersoek word. Die ontleding fokus spesifiek op die aanlyn weergawe van die *Oxford Dictionary of English*, soos beskikbaar op die "Oxford-gedrewe" Lexico.com-webtuiste. Voorlopige resultate dui daarop dat hierdie gratis aanlyn woordeboek die taboe- en diskriminerende aard van etnofaulisme weerspieël en die internetgebruiker teen die neerhalende en aanstootlike effek van hierdie woorde waarsku.

Sleutelwoorde: ENGELSE LEKSIKOGRAFIE, AANLYN WOORDEBOEKE, LINGUISTIESE RASSISME, XENOFOBIE, HAATSPRAAK, ETNIESE BELEDIGINGS, ETNOFAULISME, *OXFORD DICTIONARY OF ENGLISH*, LEXICO.COM

1. Introduction

In order to investigate the language of ethnic conflict in English online lexicography, this article presents the preliminary findings of a wider ongoing study on the treatment of ethnophaulisms, commonly referred to as ethnic or racial slurs, in the so-called "powered by Oxford" dictionary content.

As Ferrett and Dollinger (2021) explain, "powered by Oxford" is the content Oxford University Press (OUP hereafter) license for use to technology giants like Google, Yahoo, and Bing, as regards search engines, and to the pre-installed dictionaries on dominant operating systems like Microsoft and Apple. This means, for example, that due to OUP's partnership with Google, search operators like 'define ...' or '... definition' or 'what does ... mean' in Google's search engine bring up and explicitly cite Oxford definitions first, because Google English dictionary is "powered by Oxford" (Oxford Languages 2023). Moreover, "powered by Oxford" is also the dictionary website scrutinized in this paper, namely Lexico.com. Previously known as Oxford Dictionaries Online, Lexico.com was OUP's domain for the free online version of the *Oxford Dictionary of English* from June 2019 to August 26, 2022, the day on which the Lexico.com website was closed and redirected to Dictionary.com, the original website operator (Wikipedia 2023a). However, as users can learn when reading the 'About' section of Dictionary.com (2023), the online edition of the *Oxford Dictionary of English* was not actually moved to this website, since "Dictionary.com's main, proprietary source is the Random House Unabridged Dictionary".

Going back to the present case study, it is easy to understand that due to their partnerships, OUP has a remarkable market advantage because "powered by Oxford" dictionary content is extremely widespread, and this cannot but influence the way Internet users deal with language issues in this information age. As an example, Google alone accounts for more than 90% of the search engine market share worldwide (Statcounter 2023) and it is the most visited website in the world (Semrush 2023).

The rationale behind this study lies exactly in the dominant market position of "powered by Oxford" content, in connection with two typical phenomena of this digital age.

The first one is the remarkable influence of the Internet on dictionary consulting. Indeed, as many authors have highlighted (cf. Béjoint 2016; Jackson 2017; Lew and De Schryver 2014; Lorentzen and Theilgaard 2012; Müller-Spitzer and Koplenig 2014), there is a clear and increasing tendency among Internet users to search for lexical information via general search engines, that is to say, most current users tend to google their language issues (Jackson 2017: 540).

The second phenomenon linked to this study rationale is the alarming in-

crease of xenophobia, racism, and intolerance around the world, which, linguistically, translates into a proliferation of cases of hate speech, especially online, the majority of which target individuals based on ethnicity and nationality (Gagliardone et al. 2015: 13). Although it is beyond the scope of this article, it seems worth briefly mentioning that hate speech is used here as an umbrella term and includes pejorative, offensive or discriminatory uses of language with reference to a person or a group based on identity factors like ethnicity, nationality, colour, and descent.

For these reasons, the free and almost ubiquitous "powered by Oxford" dictionary content represents a good case in point to analyze the treatment of ethnic slurs or verbal expressions of intercultural or ethnic conflict, that is "ethnophaulisms" (Roback 1944), in English online lexicography with a focus on the perspective of the general user of the Internet. In other words, this study aims to answer the following research question: what does a general user of the Internet learn about the offensiveness of ethnophaulisms when looking them up on major online platforms? Since the latter are "powered by Oxford", the objective is to examine whether and how this free and pervasive online dictionary content (a) reflects the taboo nature of ethnophaulisms, (b) warns Internet users against the discriminatory power of these words, and, thus, (c) signals their hurtful nature.

To address these questions, this article explores and contextualizes the concept of ethnophaulism in the light of the relation between language, dictionaries, and society, with special attention to the increasing taboo and politically incorrect status of these expressions, which has inevitably affected both the theory and the practice of English monolingual lexicography. This critical perspective indeed characterizes a long and productive tradition of studies on the topic, which, however, have paid little attention to online dictionaries. Accordingly, to contribute to the development of this area of research, which is of relevance at the present cultural moment, this article presents the analysis of the treatment of a representative sample of ethnophaulisms in the "powered by Oxford" Lexico.com.¹ Lastly, preliminary conclusions are drawn and future research avenues are presented.

2. Ethnophaulisms in language, dictionaries, and society

As Iamartino (2020: 36) observes, when dealing with the relationship between language, dictionaries, and society, of special interest are all those entries that belong to sensitive or taboo issues in a given culture and historical period: political and social ideas, religious faith, age, sex, gender, and ethnicity. Regarding the last-mentioned, since the development of political correctness in the late 20th century thanks to the civil rights movements (see Pinnavaia 2020), ethnic slurs have turned into a new social taboo (Zgusta 2000; Wachal 2002; Green 2005). They have become increasingly offensive and racial abuse has been evaluated as the most derogatory area of language, considered to be even more

severe than profanities (Allan and Burridge 2006: 105).

According to Hughes (2010: 11-12), among "inappropriate linguistic behavior", and especially "in the category of swearing, only ethnic slurs qualify unambiguously" as politically incorrect, while "religious swearing generally does not" and "sexual swearing is divided along gender lines". Since the degree of tolerance towards politically incorrect or taboo language varies across space and time, depending on the ever-changing values and belief systems of societies, the sociocultural dynamics behind "the evolving nature of taboo" are reflected by dictionaries and "revealed in changes to lexicographic conventions" (Allan and Burridge 2006: 105, 108). Consequently, notwithstanding the perennial "dilemma between inclusiveness and 'decency'" (Hughes 2006: ix) or omission, or between inclusiveness and "censorship" (Mackintosh 2006: 54), since the late 20th century, in response to social pressure, "dictionaries makers have been much more regulative in their policy" and started to "clearly explain, label and exemplify offensive senses and uses in the dictionary's metalanguage" (Allan and Burridge 2006: 108).

Considering the increasingly sensitive nature of this lexical field, it comes as no surprise to learn that in the long and productive tradition of studies that have examined the treatment of ethnophaulisms in English monolingual lexicography, academic attention from this critical perspective has intensified exactly since the late 20th century. Scholars have addressed this topic in particular, or as part of a wider phenomenon like offensive or bad language, taboo words, and sensitive terms, with different approaches and foci: they have investigated only one lemma or a group of ethnophaulisms, only one or more dictionaries, of the same or different type and size, designed for the same or different users, concentrating on the same or different aspects of dictionary entries (cf. Murphy 1991, Norri 2000, Pinnavaia 2014). As a result, the literature is extremely rich and heterogeneous. Furthermore, and more relevantly for this article, very few studies have observed general-purpose online dictionaries, at least to an extent and they have highlighted different aspects of the English language of ethnic conflict (cf. Henderson 2003, Nissinen 2015, Žugić and Vuković-Stamatović 2021).

Of studies including online dictionaries, Henderson (2003) focuses on the treatment of a group of ethnic slurs used for black and white Americans in five monolingual dictionaries, two of which are online sources, namely *Merriam-Webster's online Collegiate Dictionary* (MWOCOD 2001) and the historical *Oxford English Dictionary Online* (OEDO 2002). Henderson (2003: 56-57) shows that, although all five dictionaries include a higher number of slurs for black people compared to those for white people, the OEDO records the most slurs but does not present any consistent pattern in labeling them, as opposed to the MWOCOD, which tends to describe words applied to black people and words applied to white people as offensive and disparaging, respectively. The OEDO (2015) is also the only non-learner online dictionary, out of a total of 20 reference works of different types and sizes, British and American, examined by Nissinen (2015)

in her study on the treatment of 37 potentially offensive nationality words. According to the research, the OEDO proved to be the dictionary which records the highest number of ethnophaulisms (35 out of 37), whose offensiveness is indicated in 69% of instances through the use of labels, definitions, usage notes, or a combination of these sections (Nissinen 2015: 56-57). Of a different nature is the research carried out by Žugić and Vuković-Stamatović (2021), which concentrates on the qualitative analysis of the definition of one single lemma, namely the word for *Albanian*, in a group of 19 online and freely accessible monolingual dictionaries of multiple languages. As concerns English, as the authors explain (2021: 183-185), these mostly include learner's dictionaries and only three general-purpose works comprising the Merriam-Webster.com dictionary site (2020), *The Random House Unabridged Dictionary* hosted on the Dictionary.com site (2020) and *The American Heritage Dictionary of the English Language* (2020) hosted on Thefreedictionary.com site.

Going back to the great academic interest the topic has attracted in more general terms, a prime example of the heterogeneity of the literature is the wide range of labels scholars have used to term the phenomenon under scrutiny since the late 20th century. Such labels include "terms for racial abuse" (Burchfield 1980), "people names" or "ethnonyms" (Rader 1989), "words offensive to groups" (McCluskey 1989), "racial labels" (Murphy 1991, 1997, 1998), "the language of racism" or "racist language" (Hauptfleisch 1993; Krishnamurthy 1996), "derogatory words for nationality and for a racial or cultural group" (Norri 2000), "racial slurs" (Himma 2002), "taboo words" (Wachal 2002), "offensive language" or "offensive items" (Coffey 2010; Schutz 2002), "ethnocentrism" (Benson 2001), "ethnic slurs" or "ethnic epithets" (Croom 2015; Henderson 2003; Pullum 2018), "bad language words" (Pinnavaia 2014), "ethnocentricity" (Moon 2014), "insulting nationality words" (Nissinen 2015), and "ethnicity terms" (Žugić and Vuković-Stamatović 2021), among others. Nevertheless, as Filmer (2011: 21-25) argues, "whichever term we use to denote ethnophaulisms", they "are the linguistic manifestation of one culture's attitudes to the other", and, as such, they evidence the language of intercultural or ethnic conflict, which is the focus of this research.

"The language of ethnic conflict", as seminally defined by Allen (1983), is a long-standing universal phenomenon. As Palmore (1962: 442) has remarked, "it seems to be universal for racial and ethnic groups to coin derogatory terms and sayings to refer to other ethnic groups". In the words of Allan and Burridge (2006: 83), "all human groups, it seems, have available in their language a derogatory term for at least one other group with which they have contact". In sum, as Filmer (2011: 18) observes, a tendency to intolerance towards ethnic diversity has manifested itself linguistically since humans began traveling and encountering peoples from other cultures and religions. This offers a reasonable explanation why the first offensive ethnic epithets appeared in the English language in the Middle Ages (Hughes 2006: 147).

Thus, thousands of "ethnophaulisms" exist across languages. This term was first introduced by American psychologist Abraham Aaron Roback (1944)

in his *Dictionary of International Slurs (Ethnophaulisms)*. It refers to "a contemptuous expression for (a member of) a people or ethnic group; an expression containing a disparaging allusion to another people or ethnic group", as defined by the "powered by Oxford" Lexico.com.

According to Hughes (2006: 146), in the English-speaking world, this lexical field shows stages of growth and decline, stages of comparative stasis and marked expansion, linked to "periods of migration, religious conflict, war, territorial expansion, political and business rivalry, immigration, and colonialism". However, since Anglophone cultures have had contact with other ethnicities for centuries, and almost always from a dominant position, English has developed an extremely vast array of ethnophaulisms, especially if compared to other languages such as Italian (Filmer 2011, 2012).

Ethnophaulisms give voice to intercultural or ethnic conflict, they represent verbal expressions of ethnic stereotypes. More specifically, as Reid and Anderson explain (2010: 100), ethnophaulisms are usually "derisive ethnic slurs" which focus on highly distinctive yet concrete aspects of group practices or characteristics (food preference or habits, physical traits, personal and group names). These references allow high-status groups to maintain hierarchies by treating low-status groups as a whole, and the lower the perceived status of a group, the higher the number of and the more negative the nature of ethnophaulisms for that group.

Today, even if we live in a politically correct cultural climate, which "prescribes and proscribes public language for ethnicity, race, gender, sexual preference, appearance, religion, (dis)ability" (Allan and Burridge 2006: 105), we read or hear about hate speech and hate words almost every day (Faloppa 2020). Ethnophaulisms often hit the headlines as the media continue to report these instances of hate speech in the many news concerning verbal and physical attacks against individuals or groups of people of different ethnicities or nationalities. Ethnophaulisms are indeed a prime example of linguistic racism. According to Hughes (2006: 146), they are manifest forms of racial intolerance, "the most obvious linguistic manifestation of xenophobia and prejudice against outgroups [...] based on malicious, ironic, or humorous distortion of the target group's identity or 'otherness'". In Van Dijk's words (2004: 427), they are a form of "racist discourse", one of the major discriminatory practices reproducing racism "as a system of social domination and inequality", and "racist prejudices and ideologies" which "in turn are the basis of discriminatory practices (including discourse)".

3. Methodology

As concerns the working methodology used in this pilot study, two aspects deserve special attention to describe the way lexical items have been identified and analyzed as representative of the English language of ethnic conflict online.

First, as to the selection of ethnophaulisms, given the focus on online lexicography and on the perspective of the Internet general user, the lexemes examined in this pilot study were derived from Wikipedia, because it is the world's largest online encyclopedia and one of the top ten most visited websites in the world (Semrush 2023). More specifically, the terms were collected from the "List of ethnic slurs" (Wikipedia 2023b), which is the first site that appears in Google search results when a general user of the Internet googles 'ethnic slur'. As stated at the beginning of this Wikipedia entry (2023b), "an ethnic slur is a term designed to insult others on the basis of race, ethnicity, or nationality". At the time of this research (August 2022) the Wikipedia list included a total of 430 ethnic slurs, but the number of the terms examined is 285. Further selection has been made according to language-specific criteria and 145 terms were excluded because they belong to languages other than English, including, for example, Italian, Spanish, Arabic, Chinese, Indonesian, Romanian, Japanese, and Russian.

The second important aspect of the methodology used in this study regards the analysis of the entries. Only the lexicographic data which proved to be relevant to the research were scrutinized and they include (a) usage labels, (b) definitions, (c) usage notes, and (d) word origin. Although these data items will be described in-depth in the analysis presented in section 4, some features of the dictionary's entries are worth briefly mentioning here. In the online edition of the *Oxford Dictionary of English* hosted on "powered by Oxford" Lexico.com, usage labels are shown in italics and placed after grammatical information. Definitions are presented as a list of senses and subsenses, each displayed on a new line, usually numbered if the lemma is polysemous; as concerns polysemy, this study has examined only the ethnicity-related sense of each term. Usage notes and word origins are generally isolated and placed after the definitions. Other sections of the dictionary entries like audio pronunciation and phonetic transcription, grammatical information (word class, plural forms, and spelling variants), usage examples illustrating the usage of the lemma for each sense, and phrases have been excluded from the analysis because they did not contain pertinent data or because they were not included in the dictionary entries (see also section 5 for a brief discussion). Regarding examples, they are taken from the Oxford English Corpus (cf. Atkins and Rundell 2008).

As regards usage labels, special attention is paid to what Jackson (2013: 113) calls "effect labels", which "relate to the effect that a word or sense is intended by the speaker or writer to produce in the hearer or reader". According to Jackson (2013: 113), they are "derogatory" and "offensive", where the difference between the two typically reflects the effect intended and/or perceived by the people involved. Indeed, as Jackson clarifies (2013: 113), while *derogatory* means "intending to be disrespectful", *offensive* "may have intent on the part of the speaker or may be unconscious", but it "could be taken by a hearer as offensive, either racially or in some other way". Similarly, the dictionary itself defines *derogatory* as "showing a critical and disrespectful attitude" and *offensive* as

"causing someone to feel resentful, upset, or annoyed".

In the analysis, the 285 ethnophaulisms selected have been further classified according to three major criteria: (1) inclusion, (2) semantic relevance (ethnicity-related lemmas or senses of lemmas) and (3) offensiveness. In particular, semantic relevance and offensiveness have been assessed on the basis of the lexicographic data contained in effect labels, definitions, usage notes, and word origin, to evaluate whether and how the dictionary signals, and thus warns the user against, the racist power of ethnophaulisms.

4. The language of ethnic conflict in "powered by Oxford" Lexico.com

Based on the criteria mentioned in the previous section, three main groups of terms have been identified in the analysis of the English language of ethnic conflict in the dictionary. These groups include terms which are (1) included or excluded, (2) semantically relevant or irrelevant and (3) offensive, i.e., ethnophaulisms, or not offensive.

As concerns the first category, namely inclusion, as Figure 1 shows, the dictionary records 227 terms, which represent 80% of the 285 terms selected from Wikipedia's list of ethnic slurs.

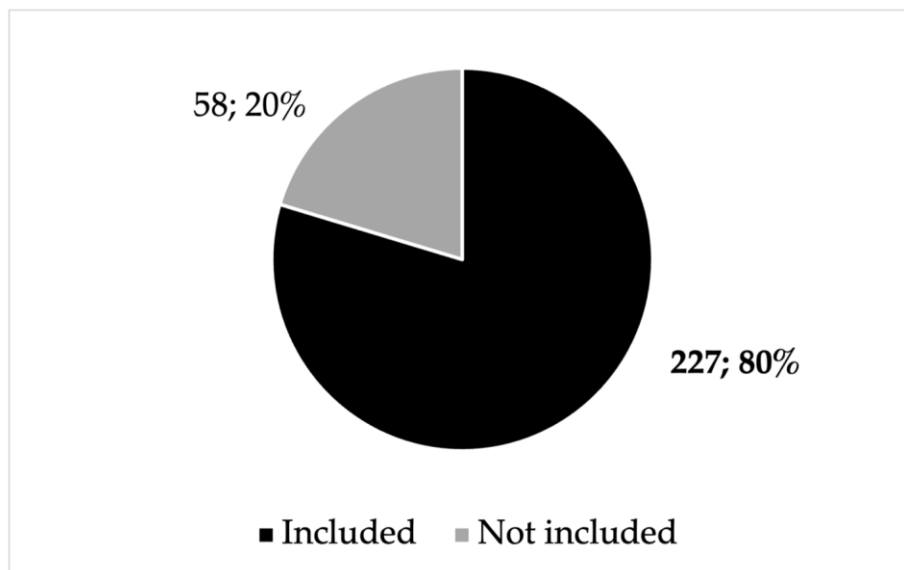


Figure 1: Terms included in "powered by Oxford" Lexico.com

Figure 1 shows that only 20% of Wikipedia's ethnic slurs (58 terms out of 285) are not recorded in the dictionary. These include, for example, *armo*, "a racial epithet" for "a white person of Armenian descent" (Dalton 2007: 139), *eight ball*, a name for "a dark-skinned black person" (Dalzell 2018: 261), *Leb*, *Lebo* or *Lebbo*, used derogatorily in Australian English to refer to "a Lebanese person, or any person from an Arabic background" (Dalzell and Victor 2013: 1375), and *nig nog*, used in British English to denote "any non-white person" (Dalzell and Victor 2013: 1580). As to their exclusion, it is possible to speculate that the dictionary does not record these ethnic slurs because they are no longer in use or because they are expressions confined within single varieties of English.

Regarding the second aspect, which is semantic relevance, as shown in Figure 2, only 48 lemmas, that is 21% of the terms included in the dictionary (227), do not present any pertinent ethnicity-related senses.

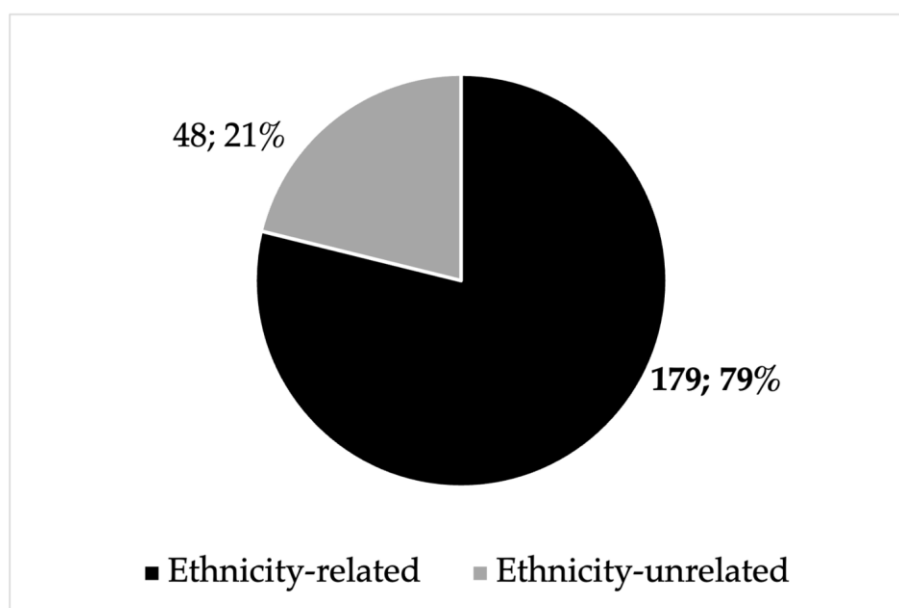


Figure 2: Ethnicity-related lemmas included in "powered by Oxford" Lexico.com

Semantically irrelevant or ethnicity-unrelated lemmas are mostly common nouns, some of which are polysemic, but the entries for them do not present any senses associated with nationality or ethnicity. Examples include the following lemmas which, according to Wikipedia's list of ethnic slurs (Wikipedia 2023b) target the ethnicity mentioned, although sometimes only in a variety of English: *ape* (US black people), *apple* (NAm native Americans), *banana* (NAm Asian people), *coconut* (US, UK, NZ Hispanics, or Latinos), *pancake* (Asian people), *snow-*

flake (US white people), and *teapot* (black people). Other interesting examples in this group are *goombah* and *shylock*. In the dictionary, **goombah** is not given as a US derogatory name for an Italian–American, as Dalzell (2018: 350) defines it, but only as an informal North American noun denoting "an associate or accomplice, especially a senior member of a criminal gang". Similarly, **shylock**, as an allusion to the Jewish moneylender in Shakespeare's *Merchant of Venice*, is only an offensive epithet for "a moneylender who charges extremely high rates of interest", but it is not said to be an anti-Semitic slur, as it is often considered and perceived today (Rothman 2014).

As shown in Figure 2, the most important group of terms identified in the analysis includes 179 ethnicity-related lemmas, which represents 79% of the lemmas included in the dictionary. In particular, based on the usage information the dictionary offers about the discriminatory potential of these words or of one of their senses, 146 ethnicity-related entries are treated as ethnophaulisms, as Figure 3 illustrates.

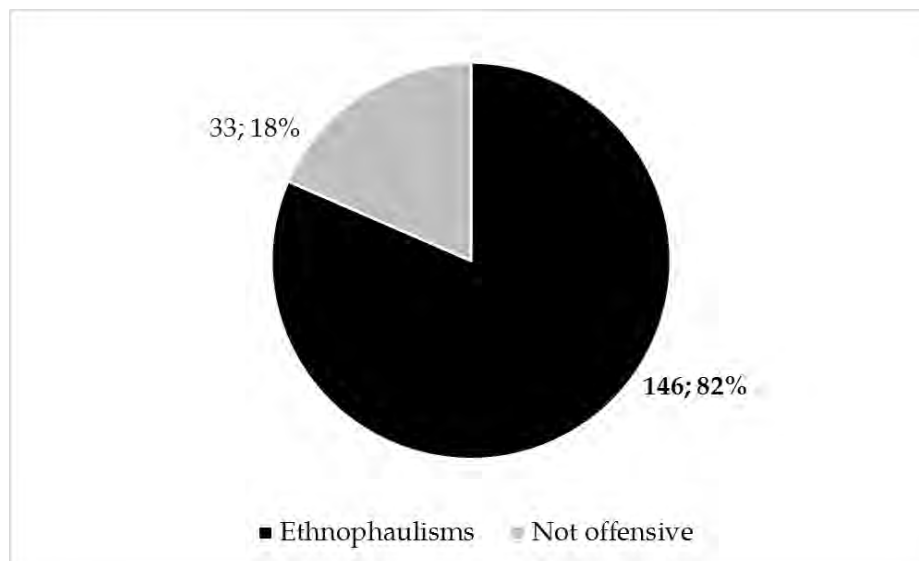


Figure 3: Ethnophaulisms in "powered by Oxford" Lexico.com

Ethnophaulisms represent 82% of the ethnicity-related lemmas. Thus, only 33 lemmas, the remaining ethnicity-related entries (18%), are not treated as ethnic slurs. Since potential offensiveness is not signaled, there is no indication in the dictionary of them belonging to the language of ethnic conflict. Instances of non-ethnophaulisms encompass terms like **ang moh**, used in Singapore English to refer to "A white person", **Indon**, an informal Australian noun for "A person from Indonesia", **Mr. Charlie**, which in African American usage means "A white

man", and also **rosbif**, an informal and humorous epithet originally used among French speakers to denote "An English person". As the examples show, although the senses of these entries are associated with ethnicity or nationality, since the dictionary does not provide any usage data about their potentially derogatory or offensive nature, users may not interpret them as ethnophaulisms.

To conclude this part of the analysis, before discussing the treatment of ethnophaulisms in more detail, it is worth highlighting that, overall, out of the total number of Wikipedia's ethnic slurs explored (285), the majority are not only included (80%) and with the relevant sense (63%), but more than half of them (51%) are treated as ethnophaulisms.

4.1 The treatment of ethnophaulisms

Concerning the treatment of ethnophaulisms in the dictionary, as Figures 4 and 5 illustrate, effect labels (L) are the major dictionary markers indicating offensiveness, either alone (108, 74%) or in combination with other sections of the entries (25, 17%), thus accounting for a total of 91% (L and L+). The combinations include label and definition (LD, 10%), label and usage note (LN, 5%), label, definition, and usage note (LDN, 1%), and label and word origin (LO, 1%). Moreover, to a limited extent, relevant usage information about offensiveness can also be found in definitions alone (D, 5%) and usage notes alone (N, 4%).

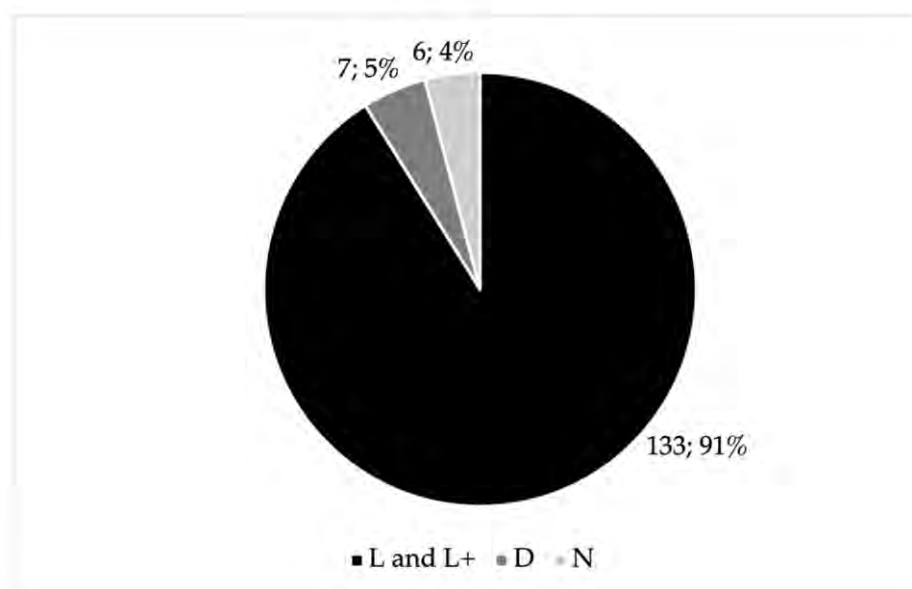


Figure 4: Markers of offensiveness in "powered by Oxford" Lexico.com

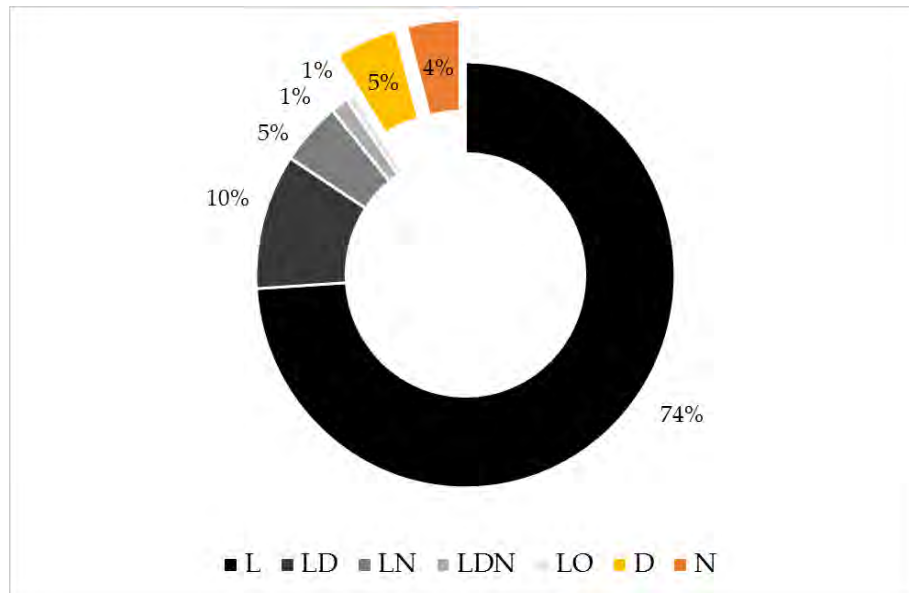


Figure 5: Effect labels and combinations in "powered by Oxford" Lexico.com

Effect labels are thus the first and most commonly used lexicographic information dictionary users find regarding offensiveness. Like all usage labels, they are highlighted in italics and placed at the beginning of the entry or of the sense they describe, depending on whether the lemma is monosemous or polysemous. Effect labels are assigned to 133 ethnophaulisms in the dictionary, and, as Figure 6 shows, they include 76 terms labeled *offensive* (57%), 50 terms labeled *derogatory* (38%), six presenting both labels (4%), and one labeled *derogatory* and *humorous* (1%). As regards the last-mentioned, the lemma is **gringo**, a noun characterized as derogatory and humorous meaning "(in Spanish-speaking countries and contexts, chiefly in the Americas) a person, especially an American, who is not Hispanic or Latino".

Going back to the two major effect labels, they can both occur with other usage labels, including temporal, stylistic, and geographical ones. Nevertheless, while *offensive* is not further specified, the *derogatory* effect of a term is also a matter of degree and frequency: on closer inspection, indeed, this label can be further qualified as *mainly derogatory* (10 occurrences) or *often derogatory* (8 occurrences), which are minor instances out of the total of 50 occurrences.

To cite some examples, the lemmas labeled *offensive* include **Abo** and **boong**, both meaning "an aboriginal person" in Australian English, **beaner** for "a Mexican or person of Mexican descent" in North American English, **bogtrotter** for "an Irish person", and **spook**, which presents an offensive sense labeled *US* and *dated* and meaning "a black person".

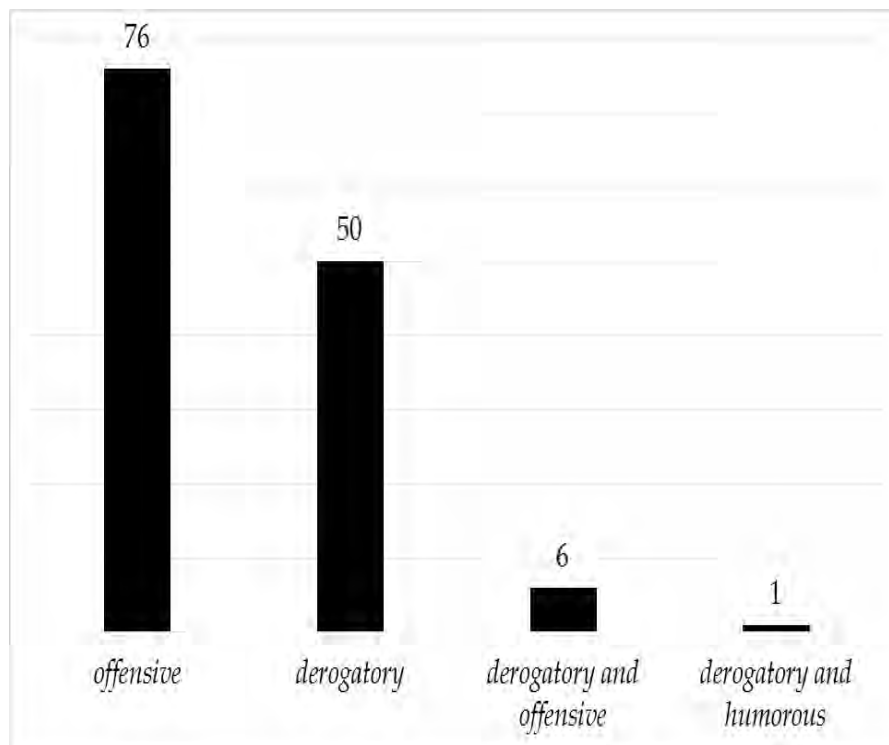


Figure 6: Effect labels in "powered by Oxford" Lexico.com

The lemmas labeled *derogatory* comprise **Argie**, an *informal British* expression for "a person from Argentina; an Argentinian", **goy**, which is used in informal language as "a Jewish name for a non-Jewish person", and **kraut**, an informal epithet for "a German". Moreover, *mainly derogatory* are, for example, **Limey** to name a British person and also **Jock**, **Paddy**, and **Taffy** meaning "a Scotsman", "an Irishman", and "a Welshman" respectively, all representing informal nouns "often used as a form of address". Labeled as *often derogatory* are, for instance, **seppo**, "an American person", and **pocho**, used in informal style for "a US citizen of Mexican origin; a culturally Americanized Mexican".

Another interesting subgroup of lemmas accompanied by effect labels are those defined as being both *derogatory* and *offensive*, including entries like **coon-ass**, **dothead**, **Jew boy**, **Rastus**, and **Uncle Tomahawk**. However, it must be clarified that in one entry only, i.e., the gendered **Jew bow**, meaning "a (typically young) Jewish male", is the *offensive* label not further associated with a geographical variety, which, in almost all cases, is US English. For example, **dothead** is said to be a slang, derogatory, offensive US term for "a person of South Asian origin or descent", while **coonass** is described as a dialect, derogatory, offensive US term for "a Cajun; a native inhabitant of Louisiana".

As shown in Figures 4 and 5, the second section warning the user against the discriminatory nature of these words in the dictionary entries is the definition, either alone (7 instances, 5%) or with other sections, namely usage labels (15 instances, 10%) or usage labels and usage notes (two, 1%). As to the marking of offensiveness in definitions only, the relevant information is always provided in brackets and corresponds to single labels, as in **frog-eater**, "especially (derogatory) a French person or a person of French descent", or it corresponds to longer usage descriptions as in **slant-eyed**, "(often used as an insult towards people of Japanese or Chinese origin)". When definitions reinforce the information also given in other sections, we can observe the following recurrent pattern in the phrasing of the descriptions: A/an + effect adjective + term for + a/an + ethnic adjective + person. The most frequently used effect adjective is 'contemptuous' in "a contemptuous term for" representing 59% of instances (ten occurrences out of 17), which always co-occurs with the label *offensive*, as in the entries for **Chink** (Chinese person), **coon** and **nigger** (black or dark-skinned person), **Jap** and **Nip** (Japanese person), **kike** and **sheeny** (Jewish person). Other effect adjectives in this phrasing pattern include *derogatory* and *offensive*, as in "A derogatory term for" and "An offensive term for".

The third section of the entry examined is the usage note, which, in the dictionary, is placed in its own box below the definition. This strategy includes instances in which the potential offensiveness is indicated in usage notes only (4%) or in usage notes combined with other sections, that is to say, combined with usage labels (5%) or with usage labels and definitions (1%). The usage notes in the dictionary vary in the quantity and quality of information offered: some focus on current usage only, while other notes, most of them, are more elaborate and also provide information relating to the origin of words.

For example, the note under **gypsy** states that "the word Gypsy is now sometimes considered derogatory or offensive and has been replaced in many official contexts by Romani or Roma, but it remains the most widely used term for members of this community among English speakers". More firmly, the note for **Indian**, examined in the sense "a member of any of the indigenous peoples of North, Central, and South America, especially those of North America" and not in the sense "a native or inhabitant of India, or a person of Indian descent", states that this term and *Red Indian* "are today regarded as old-fashioned and inappropriate, recalling, as they do, the stereotypical portraits of the Wild West". In addition, the note claims that, although *American Indian* is well established, if possible, users should refer to specific peoples, and finally it mentions European colonization with Columbus's journeys to the Americas as the origin of this sense of **Indian**.

In other entries, usage notes combine with effect labels or with labels and definitions to warn users against offensiveness. Like the ones mentioned earlier, these usage notes also vary in length and in the quantity and quality of information. An example is the entry for **nigger**, which is labeled *offensive*, defined as "a contemptuous term for a black or dark-skinned person". Accord-

ing to the usage note (original emphasis):

The word **nigger** has been used as a strongly negative term of contempt for a black person since at least the 18th century. Today it remains one of the most racially offensive words in the language. Also referred to as 'the n-word,' **nigger** is sometimes used by black people in reference to other black people in a neutral manner (in somewhat the same way that queer has been adopted by some gay and lesbian people as a term of self-reference, acceptable only when used by those within the community).

Lastly, as Figure 5 shows, the analysis has shown that the dictionary once also signaled offensiveness in the word origin box. More specifically, the lemma **mounseer**, labeled *slang*, *derogatory*, *rare*, and *archaic*, is defined as "a Frenchman". The etymological section reads as follows: "Mid 18th century; earliest use found in The Gentleman's Magazine. Representing an archaic pronunciation of French monsieur which survived as a colloquialism down to the 19th century, and occasionally appears in colloquial speech or in pejorative contexts with reference to English prejudice against foreigners".

5. Conclusions

As Gouws (2018: 215) states, "the era of Internet lexicography confronts lexicographers with challenges and opportunities to enhance the quality of the lexicographic practice and to produce dictionaries that help in satisfying the lexicographic needs" of their users. This is particularly true in users' wider sociocultural contexts, where ethnophaulisms prove to be one of those challenges and opportunities, especially for the "powered by Oxford" dictionary content, whose market-leading position cannot but influence the way Internet users deal with sensitive and taboo language issues in this digital age.

To summarize the main points and findings of this research, it seems reasonable to conclude that, despite the limitations of a pilot case study, the analysis shows that the online edition of the *Oxford Dictionary of English* hosted on the "powered by Oxford" Lexico.com quite clearly reflects the taboo nature of ethnophaulisms and quite consistently tends to warn the Internet user against the potentially racist and xenophobic power of these words. Indeed, the large majority of Wikipedia's ethnic slurs, that is 80%, is included in the dictionary (227/285), of which 79% are ethnicity-related entries (179/227), of which 82% are treated as ethnophaulisms (146/179).

Relevant and clear usage data tend to appear immediately before the definitions: labels indicate the either *offensive* or *derogatory* effect of the relevant lemmas or of one of their senses in 91% of entries, 57% of which define their use specifically as *offensive*, which means "causing someone to feel resentful, upset, or annoyed", according to the dictionary's own definition. Other important sections of the dictionary entries play a role too, although a minor one, either alone or combined with other sections. Definitions (16%) and usage notes (10%) con-

tribute to what seems to be a quite prescriptive approach of the dictionary to ethnophaulisms and, thus, to racial abuse, which might be interpreted as symptomatic of greater public awareness and sensitivity to possibly offensive racial references, while stressing the taboo nature of ethnophaulisms.

As Cloete suggests (2014: 482), the role of a dictionary, especially a general-purpose monolingual online dictionary, "is to reflect the language and thus the culture in which it exists, even if that culture is racist, sexist or in other ways politically incorrect". Cloete (2014: 848) notes further that "exclusion based on offensiveness is not acceptable", because it "might lead to ignorance and misuse". Moreover, as Cloete (2014: 484) claims, "to omit the racist words from a dictionary does not solve anything. Racist attitudes will not simultaneously be wiped out. By omitting these terms, the lexicographer loses the opportunity to warn the user against their hurtful nature". Based on preliminary findings, the online edition of the *Oxford Dictionary of English* hosted on the "powered by Oxford" Lexico.com website has not lost this opportunity.

However, to achieve the objective of the wider ongoing research project on the treatment of ethnophaulisms in "powered by Oxford" dictionary content, of which this pilot study is part, further research will be carried out on other "powered by Oxford" platforms, such as Google, Yahoo, and Bing search engines, and Microsoft and Apple preinstalled dictionaries, to compare findings across platforms in terms of dictionary content and user experience. In particular, given the clear and increasing tendency among Internet users to 'google' their language issues in this digital age (Jackson 2017: 540), special attention will have to be paid to the analysis of ethnophaulisms in the "powered by Oxford" Google's English dictionary (Ferrett and Dollinger 2021, Oxford Languages 2023). This study will be also further developed in order to cover the analysis of usage examples. In this regard, although they were excluded from this initial stage of research, it is relevant to mention that some illustrative examples are provided in 58 entries only, of which 30 are treated as ethnophaulisms, meaning that the dictionary does not exemplify the use of ethnic slurs in 80% of instances. This aspect seems to suggest an interesting tendency that deserves special attention, because examples are a fundamental and sometimes controversial lexicographic component of dictionary entries, as far as socioculturally sensitive issues are concerned (see Pettini 2021).

Endnotes

1. All lexicographic data cited and discussed in this pilot study are from the online edition of the *Oxford Dictionary of English* hosted on "powered by Oxford" Lexico.com (as of August 26, 2022), which is also referred to as simply "the dictionary" where applicable.

References

- Allan, K. and K. Burridge. 2006. *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.

- Allen, I.L.** 1983. *The Language of Ethnic Conflict: Social Organization and Lexical Culture*. New York: Columbia University Press.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: University Press.
- Béjoint, H.** 2016. Dictionaries for General Users: History and Development; Current Issues. Durkin, P. (Ed.). 2016. *The Oxford Handbook of Lexicography*: 7-24. Oxford: Oxford University Press.
- Benson, P.** 2001. *Ethnocentrism and the English Dictionary*. London/New York: Routledge.
- Burchfield, R.** 1980. Dictionaries and Ethnic Sensibilities. Michaels, L. and C. Ricks (Eds.). 1980. *The State of the Language*: 15-23. Berkeley: University of California Press.
- Cloete, A.E.** 2014. The Treatment of Sensitive Items in Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2014. *Dictionaries. An International Encyclopedia of Lexicography Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 482-486. Berlin/Boston: De Gruyter.
- Coffey, S.** 2010. 'Offensive' Items, and Less Offensive Alternatives. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress, Leeuwarden, 6–10 July 2010*: 1270-1281. Leeuwarden/Ljouwert: Fryske Akademy — Afûk.
- Croom, A.** 2015. The Semantics of Slurs: A Refutation of Coreferentialism. *Ampersand* 2: 30-38.
- Dalton, C.H.** 2007. *A Practical Guide to Racism*. Sheridan, WI: Gotham Books.
- Dalzell, T. (Ed.).** 2018. *The Routledge Dictionary of Modern American Slang and Unconventional English*. Second edition. London: Routledge.
- Dalzell, T. and T. Victor (Eds.).** 2013. *The New Partridge Dictionary of Slang and Unconventional English*. Second edition. London: Routledge.
- Dictionary.com.** 2023. 'About'.
<https://www.dictionary.com/e/about> (July 4, 2023)
- Faloppa, F.** 2020. L'hate speech, questo sconosciuto. *Speciale "Hate Speech"*. *Lingua Italiana Treccani*, October 05, 2020.
https://www.treccani.it/magazine/lingua_italiana/speciali/Hate_speech/01_Faloppa.html (April 22, 2023)
- Ferrett, E. and S. Dollinger.** 2021. Is Digital Always Better? Comparing Two English Print Dictionaries with Their Current Digital Counterparts. *International Journal of Lexicography* 34(1): 66-91.
- Filmer, D.** 2011. *Translating Racial Slurs: A Comparative Analysis of Gran Torino Assessing Transfer of Offensive Language between English and Italian*. Unpublished MA thesis. Durham, UK: Durham University.
- Filmer, D.** 2012. The 'gook' goes 'gay': Cultural Interference in Translating Offensive Language. *inTRAlinea* 14.
<http://www.intraline.org/archive/article/1829> (April 22, 2023)
- Gagliardone, I., D. Gal, T. Alves and G. Martinez.** 2015. *Countering Online Hate Speech*. Paris: UNESCO Publications.
- Gouws, R.** 2018. Internet Lexicography in the 21st Century. Engelberg, S., H. Kämper and P. Storjohann (Eds.). 2018. *Wortschatz: Theorie, Empirie, Dokumentation*: 215-236. Berlin: De Gruyter.
- Green, J.** 2005. Wash Your Mouth Out. *Critical Quarterly* 46(1): 107-111.
- Hauptfleisch, D.C.** 1993. Racist Language in Society and in Dictionaries: A Pragmatic Perspective. *Lexikos* 3: 83-139.
- Henderson, A.** 2003. What's in a Slur? *American Speech* 78(1): 52-74.

-
- Himma, K.E.** 2002. On the Definition of Unconscionable Racial and Sexual Slurs. *Journal of Social Philosophy* 33(3): 512-522.
- Hughes, G.** 2006. *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World*. Armonk: M.E. Sharpe.
- Hughes, G.** 2010. *Political Correctness: A History of Semantics and Culture*. Oxford: Wiley Blackwell.
- Iamartino, G.** 2020. Lexicography as a Mirror of Society: Women in John Kersey's Dictionaries of the English Language. *Textus: English Studies in Italy* 33(1): 35-67.
- Jackson, H.** 2013. *Lexicography: An Introduction*. London/New York: Routledge.
- Jackson, H.** 2017. English Lexicography in the Internet Era. Fuertes-Olivera, P.A. (Ed.). 2017. *The Routledge Handbook of Lexicography*: 540-553. London: Routledge.
- Krishnamurthy, R.** 1996. Ethnic, Racial and Tribal: The Language of Racism? Caldas-Coulthard, C. and M. Coulthard (Eds.). 1996. *Texts and Practices: Readings in Critical Discourse Analysis*: 129-149. London: Routledge.
- Lew, R. and G.-M. de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
- Lorentzen, H. and L. Theilgaard.** 2012. Online Dictionaries — How Do Users Find Them and What Do They Do Once They Have? Vatvedt Fjeld, R. and J.M. Torjusen (Eds.). 2012. *Proceedings of the XV EURALEX International Congress, 7–11 August, 2012, Oslo*: 654-660. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- 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: University of Ottawa Press.
- McCluskey, J.** 1989. Dictionaries and Labeling of Words Offensive to Groups, with Particular Attention to the Second Edition of the OED. *Dictionaries: Journal of the Dictionary Society of North America* 11: 111-123.
- Moon, R.** 2014. Meanings, Ideologies, and Learners' Dictionaries. Abel, A., C. Vettori and N. Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, Bolzano/Bozen, Italy, July 15–19, 2014*: 85-105. Bolzano/Bozen: EURAC Research.
- Müller-Spitzer, C. and A. Koplenig.** 2014. Online Dictionaries: Expectations and Demands. Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*. Lexicographica Series Maior 145: 143-188. Berlin/Boston: De Gruyter.
- Murphy, M.L.** 1991. Defining Racial Labels: Problems and Promise in American Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 13(1): 43-64.
- Murphy, M.L.** 1997. Afrikaans, American and British Models for South African English Lexicography: Racial Label Usage. *Lexikos* 7: 153-164.
- Murphy, M.L.** 1998. Defining People: Race and Ethnicity in South African English Dictionaries. *International Journal of Lexicography* 11(1): 1-33.
- Nissinen, S.** 2015. *Insulting Nationality Words in Some British and American Dictionaries and in the BNC*. Unpublished MA Thesis. Tampere, FI: University of Tampere.
- Norri, J.** 2000. Labelling of Derogatory Words in Some British and American Dictionaries. *International Journal of Lexicography* 13(2): 71-106.
- Oxford Languages.** 2023. Oxford Languages and Google.
<https://languages.oup.com/google-dictionary-en/> (April 22, 2023)
- Palmore, E.B.** 1962. Ethnophaulisms and Ethnocentrism. *American Journal of Sociology* 67(4): 442-445.

- Pettini, S.** 2021. "One is a Woman, So That's Encouraging Too": The Representation of Social Gender in "Powered by Oxford" Online Lexicography. *Lingue e Linguaggi* 44: 275-295.
- Pinnavaia, L.** 2014. Defining and Proscribing Bad Language Words in English Learner's Dictionaries. Iannaccaro, G. and G. Iamartino (Eds.). 2014. *Enforcing and Eluding Censorship: British and Anglo-Italian Perspectives*: 217-230. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Pinnavaia, L.** 2020. Tracing Political Correctness in Bilingual English-Italian Dictionaries. *Textus: English Studies in Italy* 33(1): 87-106.
- Pullum, G.K.** 2018. Slurs and Obscenities: Lexicography, Semantics, and Philosophy. Sosa, D. (Ed.). 2018. *Bad Words: Philosophical Perspectives on Slurs*: 168-192. Oxford: Oxford University Press.
- Rader, J.** 1989. People and Language Names in Anglo-American Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 11: 125-138.
- Reid, S.A. and G.L. Anderson.** 2010. Language, Social Identity, and Stereotyping. Giles, H., S.A. Reid and J. Harwood (Eds.). 2010. *The Dynamics of Intergroup Communication*: 91-104. New York: Peter Lang.
- Roback, A.A.** 1944. *A Dictionary of International Slurs (Ethnophaulisms)*. Cambridge, MA: Sci-Art Publishers.
- Rothman, L.** 2014. When Did 'Shylock' Become a Slur? *Time*, September 17.
<https://time.com/3394403/shylock-biden/> (April 22, 2023)
- Schutz, R.** 2002. Indirect Offensive Language in Dictionaries. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002*: 637-641. Copenhagen: Center for Sprogteknologi, Copenhagen University.
- Semrush.** 2023. Top Websites: Most visited Websites by Traffic in the World for All Categories, March 2023.
<https://www.semrush.com/website/top/> (April 22, 2023)
- Statcounter.** 2023. Search Engine Market Share Worldwide: March 2023.
<https://gs.statcounter.com/search-engine-market-share> (April 22, 2023)
- Van Dijk, T.A.** 2004. Racism, Discourse and Textbooks: The Coverage of Immigration in Spanish Textbooks. Paper for the Symposium on Human Rights in Textbooks, organized by the History Foundation, Istanbul.
<http://www.discursos.org/unpublished%20articles/Racism,%20discourse,%20textbooks.htm> (April 22, 2023)
- Wachal, R.S.** 2002. Taboo or Not Taboo: That is the Question. *American Speech* 77(2): 195-206.
- Wikipedia.** 2023a. Lexico.
<https://en.wikipedia.org/wiki/Lexico> (April 22, 2023)
- Wikipedia.** 2023b. List of Ethnic Slurs.
https://en.wikipedia.org/wiki/List_of_ethnic_slurs (April 22, 2023)
- Zgusta, L.** 2000. Some Developments in Lexicography, Past and Present. Mogensen, J.E., V.H. Pedersen and A. Zettersten (Eds.). 2000. *Symposium on Lexicography IX: Proceedings of the Ninth International Symposium on Lexicography, April 23-25, 1998 at the University of Copenhagen*. Lexicographica. Series Maior 103: 11-26. Tübingen: Max Niemeyer.
- Žugić, D. and M. Vuković-Stamatović.** 2021. Problems in Defining Ethnicity Terms in Dictionaries. *Lexikos* 31: 177-194.

Phonological Inclusion, and Exclusion, Regarding South African English in the Online OED

Alex Baratta, *Manchester Institute of Education,
University of Manchester, Manchester, UK*
(alex.baratta@manchester.ac.uk)

Abstract: This article approaches the subject of inclusivity and diversity from a phonological perspective, as applied to the online *Oxford English Dictionary* (OED). Specifically, the pronunciation guide within the OED entry for South African English (SAE) is investigated. On the one hand, that the OED provides a list of World Englishes is a step forward in terms of diversity, avoiding the linguistic hegemony of British and American English, for example. On the other hand, the pronunciation guide for SAE relies on a rather singular variety — that reflective of white South Africans. Moreover, the OED pronunciation guides for all varieties of English reflect "educated urban speakers of standard English". Therefore, the pronunciation guide is exclusive regarding race and class, and thus additional varieties of SAE are not represented. As such, while the OED, as a trusted source, has made efforts to address linguistic diversity, more work needs to be done in this area.

Keywords: PRONUNCIATION, ACCENT, WHITE SOUTH AFRICAN ENGLISH, GREAT TRICHOTOMY, EXCLUSIVITY, LANGUAGE ATTITUDES

Opsomming: Fonologiese insluiting, en weglating, rakende Suid-Afrikaanse Engels in die aanlyn OED. In hierdie artikel word inklusiwiteit en diversiteit, soos toegepas op die *Oxford English Dictionary* (OED), vanuit 'n fonologiese perspektief benader. Die uitspraakgids binne die OED-inskrywing vir Suid-Afrikaanse Engels (SAE) word spesifiek ondersoek. Aan die een kant is die insluiting van 'n lys van Wêreldengels deur die OED, in terme van diversiteit, 'n stap vorentoe om sodoende, byvoorbeeld, die linguïstiese hegemonie van Britse en Amerikaanse Engels te vermy. Aan die ander kant steun die uitspraakgids vir SAE op 'n enkele variëteit — wat 'n weerspieëling van wit Suid-Afrikaners is. Boonop reflekteer die OED-uitspraakgids vir alle variëteite "opgevoede stedelike sprekers van standaardengels". Hierop gebaseer, is die uitspraakgids eksklusief rakende ras en klas, en dus word addisionele variëteite van SAE nie verteenwoordig nie. Alhoewel die OED, as 'n vertroude bron, pogings aangewend het om linguïstiese diversiteit aan te spreek, sal meer werk op hierdie gebied verrig moet word.

Sleutelwoorde: UITSPRAAK, KLEM, WIT SUID-AFRIKAANSE ENGELS, GROOT DRIEDELING, EKSKLUSIWITEIT, TAALHOUDINGS

Introduction

British and American English have long since had a hold on the English language, notably in educational contexts such as EFL teaching (Baratta 2019). Beyond the implications this has for vocabulary teaching, there is also the issue of pronunciation to consider. Largely, this has focused on Received Pronunciation (RP) and General American as linguistic 'stand-ins' for British English and American English respectively (Kaur 2014; Carrie 2016; Wong 2018). While there is evidence of more regional British accents being appreciated for their 'difference' by EFL students (Baratta and Halenko 2022), if we look to dictionary entries then it is often the case that RP and GA are still the de facto guides for pronunciation. This is no exception within the online OED. While this suggests a certain exclusivity regarding pronunciation models for these two varieties of English, we might consider a more inclusive approach to language from an even broader perspective. This perspective in question pertains to the vast diversity of Englishes spoken around the world, those that go beyond the inner circle of native speakers (Kachru 1982) and involve speakers of English from countries such as India, Singapore and Ghana, as but three examples. Millions of people around the world have made the English language their own as part of World Englishes, varieties which are "unfairly judged based on exonormative norms set by more prestigious varieties such as British and American English" (Salazar 2023: 38). The features of World Englishes reflect differences — and not deficit — from the native speaker varieties in terms of grammar, lexis and pronunciation, and it is this latter category that is the focus of this article. While the central purpose of dictionaries is undoubtedly to present the vocabulary of a given language variety, there is perhaps a somewhat overlooked function — the role of dictionaries in providing pronunciation guides for their lexical entries. The OED has indeed taken steps to reflect the diversity of English, with its online dictionary providing entries for a variety of native (e.g., Canadian English) and non-native Englishes (e.g., Kenyan English), to include stigmatised native speaker varieties such as African American English (commonly known as Ebonics). Thus, the OED, in acknowledging the linguistic reality which involves all kinds of Englishes, is making positive linguistic strides, but there is perhaps additional room to incorporate further varieties within varieties. This article now focuses on this spirit of linguistic equality, and by extension *cultural* equality, by turning to the entry on SAE within the World Englishes project, which itself comprises a prominent section of the online OED.

South African English entry within the OED

In the entry for SAE (OED 2023a), there is a great deal of information on the topic, including recently recorded words (e.g., *ghoen* and *lightie*); a link from which individuals can submit new words for inclusion; and further links to relevant information, such as 'Introduction to South African English'. Of par-

ticular relevance to this article of course is the link to the pronunciation guide, which ostensibly reflects 'the South African accent'. There does not exist, of course, a singular South African accent. The OED acknowledges this, stating that "there is no single, reasonably uniform SAE accent", within the 'Introduction to South African English' section (Silva 2023). This then raises the question as to what accent will be provided via the pronunciation guide. Much like RP and GA representing the socially dominant accents of British and American English, and not, say, Liverpool and New York City, this raises the issue of linguistic exclusion.

Silva (2023) asserts that "no particular (model) is promoted over another" regarding pronunciation for (World) Englishes in the OED. This claim, however, is somewhat dubious regarding SAE (and perhaps other entries), for the following reasons. First, the pronunciation model for SAE within the online OED is based on white South African English, as clearly stated on the website (OED 2023b):

The focus for South African English is not the African variety widely termed 'Black South African English', nor the distinct 'Indian South African English', but the variety known as 'White South African English'.

This is not to suggest that white South African English should be replaced with, say, a pronunciation guide reflecting black South African English; rather, one can simply ask if room can be made for additional varieties in terms of pronunciation models and guides, such as Indian South African English.

That said, within the section 'Introduction to South African English' (Silva 2023), there is a great deal of information provided which discusses this variety, in terms of the origins of English in South Africa, the linguistic tug of war between English and Afrikaans, and in more specific terms regarding pronunciation, this section fully acknowledges, as referenced earlier, that there is no singular South African accent. Moreover, the point is made that ethnic differences regarding pronunciation became more pronounced due to the Apartheid system of separate schooling, for the most part, regarding the different groups (e.g., Black South African English, Indian South African English). However, Silva (2023) goes on to say that as children are now educated together, "ethnically determined differences in SAE are tending to break down". Though no phonological examples of this are provided, it is arguably still the case that distinctions on the basis of race, class and ethnicity can be detected in people's accents, and as such, perhaps this can be reflected in additional pronunciation guides on SAE.

Furthermore, the pronunciation guide for SAE is also very much reflective of social class, as are all the entries for (World) Englishes within the OED. The OED makes clear that its pronunciation guides reflect a "carefully-devised model", representing "educated urban speakers of standard English" (Silva 2023). This in turn suggests that working-class individuals, speakers of dialects and, more controversially, 'uneducated' speakers are not included regarding their models of pronunciation.

Furthermore, within the South African context, accents deemed to be 'broad' are not only associated with working-class individuals, but also associated with Afrikaans English phonology (Lass 2002); thus, not only are pronunciation models absent on the basis of race and class, but, to an extent, pronunciation models on the basis of ethnicity are also excluded, within the broader category of white South African English (OED 2023c):

Within White South African English, there are three groupings: the upper-class-associated 'Cultivated' form, the middle-class-associated 'General' form, and the working-class/Afrikaans-associated 'Broad'. The variety of focus for the OED model is the General form, although tending towards Cultivated over Broad.

To provide further background information, we need to consider the Great Trichotomy (Lass 2002), an approach taken in an attempt to classify South African accents. This consists of the three accents referenced above, cultivated, general and broad, each of which comes with specific connotations of the speaker, notably regarding class levels. Lass (2002), for example, points out that the broad accent is stigmatised, associated with those with low levels of education and low socio-economic status, and approximates Afrikaans English in terms of its phonology. Thus, we can see how class and ethnicity can intersect, and in this case, reflect a pronunciation — and by extension, accent — which does not receive perhaps a great deal of societal respect. Tellingly, Lass (2002) refers to the classifications within the Trichotomy as earlier involving terms such as 'respectable' and 'extreme' for the general and broad varieties respectively, further pointing out that the broad accent is looked down upon by cultivated and general speakers of white South African English in particular. Lass (2002: 110) refers to terms such as 'respectable' and 'extreme' as "nasty creations", and goes even further regarding his discussion of linguistic attitudes toward the three accents that make up the Trichotomy of white South Africans: "Type 2 speakers would not want to be caught dead really sounding like Type 3, but they don't sound all that much like Type 1 either, though many tend to think they do or wish they did."

That the OED entry for SAE pronunciation is based on 'General White South African English' without a "single, reasonably uniform SAE accent" (Silva 2023), might be a reflection of what Salazar (2023: 30) refers to as "potential discrepancies and compromises" within the OED's pronunciation entries for African Englishes overall. Bowerman (2008: 168), commenting on the Trichotomy, explains that the focus on white South Africans' use of English as the basis for the trichotomy is "not intended to reflect the apartheid classifications", but this particular variety might nonetheless create a sense of exclusion. Nonetheless, to use a variety of SAE associated with white South Africans, and to focus on the pronunciation of a particular social class, suggests that other accent varieties are marked and potentially subject to negative perceptions. Thus, there is the potential for exclusion on the basis of race and class, and arguably ethnicity, regarding South African English pronunciation guides within the online OED.

Going forward, online dictionaries in particular have the potential to address more fully the diversity seen within the English language from a phonological perspective. This is something Baratta (2022) argued for, in the specific context of online dictionaries compiled by individuals who themselves use a specific language variety, though not professional lexicographers themselves. This represents what Damaso (2005: 4) refers to as "democracy and equal access to meaning-making rights". We can see this reflection of linguistic diversity in South Africa, broadly, with the creation of the Kaaps dictionary, reflecting "a great democratic resource for developing understanding" (Haupt 2021). Further, The Dictionary Unit for South African English is another good example of linguistic inclusivity, as it is a resource committed to documenting the varieties of English within South Africa, and the various linguistic influences on SAE. Regarding pronunciation models specifically, further inclusion in online dictionaries could include non-white Englishes, such as Indian South African English, and the working-class/Afrikaans-associated broad accents. Lass (2002: 111) in fact states that non-white Englishes in South Africa, such as Indian South African English, "have their own internal varietal stratification", suggesting further variety within variety. That the OED is "still subject to ongoing review" regarding SAE pronunciation means that there is scope for further inclusion. It is with this purpose in mind — a democratic approach to SAE — that phonological inclusion should continue to be a main feature of dictionaries.

References

- Baratta, A. 2019. *World Englishes in English Language Teaching*. London: Palgrave Macmillan.
- Baratta, A. 2022. Stigmatised Dictionaries Housing a Stigmatised Variety of English: The Use of Korean English Online Dictionaries as a Teaching Tool within the EFL Classroom. *Lexikos* 32(1): 250-271.
- Baratta, A. and N. Halenko. 2022. Attitudes toward Regional British Accents in EFL Teaching: Student and Teacher Perspectives. *Linguistics and Education* 67: 1-11.
- Bowerman, S. 2008. White South African English: Phonology. Mesthrie R. (Ed.). 2008. *Varieties of English 4: Africa, South and Southeast Asia*: 164-176. Berlin: Mouton de Gruyter.
- Carrie, E. 2016. 'British is professional, American is urban': Attitudes toward English Reference Accents in Spain. *International Journal of Applied Linguistics* 27(2): 427-447.
- 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.
- Haupt, A. 2021. The First-ever Dictionary of South Africa's Kaaps Language has Launched — Why it Matters. *The Conversation*, August 29th, 2021.
- Kachru, B.B. 1982. *The Other Tongue. English across Cultures*. Urbana, Illinois, USA: University of Illinois Press.
- Kaur, P. 2014. Accent Attitudes: Reactions to English as a Lingua Franca. *Procedia — Social and Behavioral Sciences* 134: 3-12.
- Lass, R. 2002. South African English. Mesthrie R. (Ed.). *Language in South Africa*: 104-126. Cambridge: Cambridge University Press.

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1838> (Volle uitgawe / Full volume)
<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/33-1-1817> (Lexiconote)

Oxford English Dictionary (OED). 2023a. World Englishes. South African English.

<https://www.oed.com/discover/south-african-english/>

Oxford English Dictionary (OED). 2023b. South African English. Pronunciation Model.

<https://www.oed.com/information/understanding-entries/pronunciation/world-englishes/south-african-english>

Salazar, D. 2023. African Englishes in the *Oxford English Dictionary*. *Lexikos* 33(1): 24-40.

Silva, P. 2023. Introduction to South African English. *Oxford English Dictionary*.

www.oed.com/discover/introduction-to-south-african-english

Wong, R. 2018. Non-native EFL Teachers' Perception of English Accent in Teaching and Learning: Any Preference? *Theory and Practice in Language Studies* 8(2): 177-183.

A Further Look into the Use of a Dictionary APP in EFL Writing: A Replication Study

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

and

Suping Liu, *College of Foreign Languages,
Putian University, Fujian, P.R.C.*
(Corresponding Author, 260033359@qq.com)

Abstract: The study replicated the experiment by Chen and Liu (2022), investigating the effect of dictionary use on EFL writing. It involved the same research variables as the original study except for adopting a different dictionary. Sixty-two English majors took two writing tests, one without dictionary assistance, the other with access to a mobile phone dictionary application which features a combination of an L1–L2 and an L2–L1 dictionary for bidirectional search. The application can keep a record of users' search inputs and entry clicks. A questionnaire was also conducted to survey the students' evaluation of the dictionary application. Different from the negative results found by the original study, the replication revealed a non-significant effect of dictionary use on writing performance, providing solid evidence that a better dictionary leads to fewer consultation errors, although the improvement in writing scores brought about by dictionary use was only marginal. The study confirmed the original finding about the positive impact of dictionary use on lexical sophistication. It also identified some differences in dictionary lookup patterns between the participants of the replication and the original study in terms of search frequency, preference for language search, preference for search items, and use of source dictionaries. The implications of the study for dictionary making are discussed.

Keywords: REPLICATION, DICTIONARY USE, EFL WRITING, WRITING PERFORMANCE, LEXICAL SOPHISTICATION, LOOKUP PATTERNS

Opsomming: 'n Verdere kyk na die gebruik van 'n woordeboektoepassing in EVT-skryfwerk: 'n Repliseringstudie. In hierdie studie is die eksperiment van Chen en Liu (2022) waarin die effek van woordeboekgebruik op EVT-skryfwerk bestudeer is, gerepliseer. Buiten die gebruik van 'n ander woordeboek, het dit dieselfde navorsingsveranderlikes as die oorspronklike studie behels. Twee-en-sestig studente met Engels as hoofvak het twee skryftoetse afgeleë, een sonder die hulp van 'n woordeboek, die ander een met toegang tot 'n selfoonwoordeboektoepassing wat 'n kombinasie van 'n L1–L2- en L2–L1-woordeboek vir tweerigtingsoektogte bevat. Die

toepassing kan 'n rekord hou van gebruikers se soektogte en klikke op inskrywings. 'n Vraelys is ook voltooi om die studente se evaluering van die woordeboektoepassing te bepaal. Anders as die negatiewe resultate wat deur die oorspronklike studie verkry is, is daar in die replisering nie 'n beduidende effek van woordeboekgebruik op skryfprestasie nie, wat goeie bewyse verskaf dat die gebruik van 'n beter woordeboek tot minder naslaanfoute lei, alhoewel die verbetering in skryfprestasie deur die gebruik van 'n woordeboek slegs marginaal was. Die studie het die oorspronklike bevinding rakende die positiewe impak van woordeboekgebruik op leksikale sofistikasie bevestig. Dit het ook enkele verskille in woordeboeknaslaanpatrone tussen die deelnemers van die replisering en dié van die oorspronklike studie rakende soekfrekwensie, voorkeur vir taalsoektogte, voorkeur vir soekitems, en gebruik van bronwoordeboeke geïdentifiseer. Die implikasies van die studie vir woordeboekmaak word bespreek.

Sleutelwoorde: REPLISERING, WOORDEBOEKGEBRUIK, EVT-SKRYFWERK, SKRYFPRESTASIE, LEKSIKALE SOFISTIKASIE, NASLAANPATRONE

1. Introduction

1.1 Importance of replication studies

Within the empirical sciences, replication plays a major role in assessing the internal and external validity of findings and establishing predictable exceptions (Lindsay and Ehrenberg 1993, Gast 2009, Abbuhl 2012). It also helps to expose the weaknesses of the original study and improve the way we interpret empirical research (LTRP 2008: 1). However, such research is seldom attempted because it is difficult to successfully accomplish and it carries more risk than potential reward for both the replicator and the originator of the research (Park 2004: 194). In the field of second language acquisition (SLA), due to the lack of prestige and rewards associated with replication, it is not widely practiced either. In particular, replications in second language writing are virtually nonexistent (LTRP 2008).

As regards dictionary use research, there are also relatively few studies openly acknowledged to be replications of some previous investigations (Dziemianko 2012: 199). Yet, similar to other areas of SLA research, increasingly "diverse in scope and investigation of topics" and thus resulting in "divergent and at times fragmented research results" (LTRP 2008: 11), a replication of dictionary use research is even more needed, valued and encouraged today than before.

Dziemianko (2010, 2011, 2012, 2017) conducted a series of approximate replications¹ to evaluate the role of dictionary media in language learning which involved the same battery of tests, participants with the same English proficiency and linguistic background, and the same experimental setting. Regardless of the different, if not contradictory results, the series of replications exhibit steady improvement in research methodology and give fascinating in-

sights into the way dictionary form affects language comprehension, production and retention. Such self-replications, rare as they are, prove worthwhile due to their potential to motivate the researcher to higher standards of replicability, to learn from his own experience or mistakes, and to improve or even reconceptualize his own methods (see LTRP 2008: 6).

1.2 Introduction to the original study

Dictionary use in L2 writing has not received due attention from researchers. As indicated by Chen and Liu (2022), some efforts have been made to investigate how the availability of dictionary impacts on writing performance (e.g. Tall and Hurman 2002, East 2007, Qiao and Wang 2020, Lew 2016), how dictionary consultation affects lexical accuracy and lexical sophistication (e.g. Nesi and Meara 1994, Christianson 1997, East 2006, Qiao and Wang 2020), and what lookup patterns and strategies are employed by L2 writers (e.g. Boonmoh 2012, Chon 2009, Lai and Chen 2015), but scarce are endeavors to explore the use of electronic dictionary by Chinese EFL learners in writing.

A most recent study by Chen and Liu (2022) examined the effect of dictionary use on writing performance, lexical sophistication and the search patterns and strategies of English majors at a Chinese university. In the first week, the students familiarized themselves with Bing.dict, an online bilingual dictionary, and filled in a questionnaire on their dictionary consultation habits, preferences, perceptions on the role of dictionary in EFL writing, and needs for dictionary instruction. One week later, the students were asked to write, without any dictionary, a 200-word composition on a given topic on a word processor. In the following week, they were instructed to write on another given topic with Bing.dict. A screen-recorder was preinstalled in the computers to record how the students consulted the dictionary to assist their writing.

The study found that Bing.dict produced a significantly negative effect on the students' overall composition scores and the component scores for content and language use as well, although it did play a part in increasing the students' lexical richness. A variety of dictionary-based errors were committed in terms of lexicon, syntax and collocation due to the students' inadequate dictionary use skills and the unsatisfactory quality of the dictionary for language production. Screen recordings demonstrated that the students employed a range of poor strategies for dictionary consultation which brought about undesirable results.

1.3 Motivations for the present replication

The motivations to self replicate Chen and Liu's study were twofold. To our knowledge, the original study is the first one to reveal a substantially adverse

impact of dictionary use on writing performance. This result diverges from the conclusion from Tall and Hurman (2002), Lew (2016), East (2006, 2007), and Qiao and Wang (2020). Therefore, it is necessary to seek more evidence to test the original finding. In addition, Bing.dict is a commercially minded AED (alternative e-dictionary, Nesi 2012) which combines diverse resources such as dictionaries of different types and online resources which are not produced by lexicographers at all. Despite its wide popularity and high evaluation score according to Lew and Szarowska's (2017) Framework, it was not considered an ideal dictionary for language production due to its serious defects (Chen and Liu 2022: 486). The authors ascribed the negative role of dictionary use in writing partly to the unsatisfactory quality of Bing.dict, suggesting that users exercise caution when turning to AEDs for language encoding. This gives rise to an intriguing question: What would happen if other non-AEDs were utilized in a similar setting? Would that negative result hold true with another dictionary? To determine the generalizability of the original conclusion, replication seemed necessary.

In the next section, the design of the replication will be introduced, covering the research questions, the participants, the writing topics, especially the dictionary used for the study. Research methods and major procedure of the study will also be elaborated. Section 3 will report the results of the replication in terms of writing performance, lexical sophistication and dictionary look up behavior. It will also discuss and compare the outcomes of the replication with those of the original. Section 4 will contain a summary of the findings, making suggestions for improving dictionary compilation. Finally, section 5 will explain the limitations of the replication and propose some topics for future research.

2. Replication to be tried

2.1 Research questions

The replication to be tried seeks to address basically the same research questions as those in the original study. Specifically, the following questions are formulated.

- (1) Does the mobile application of *New Century English–Chinese Chinese–English Dictionary* (henceforth the APP) have a significant effect on the participants' writing scores?
- (2) To what extent does the APP contribute to increasing the participants' lexical sophistication as measured by lexical frequency profile (LFP)?
- (3) What differences occur in dictionary lookup patterns between the participants of the replication and the original study?

2.2 Participants and writing topics

To overcome one of the limitations of the original study, i.e. a relatively small number of participants, we increased the sample size from 34 to 62. These participants bore remarkable similarities with those original ones: they were English sophomores at the same university, shared the same linguistic and cultural background, and were about to take TEM4 (Test for English Majors, Band 4) in three months.

The writing topics from the original study were also taken. One was concerned with money saving, the other was regarding making friends online, on the prerequisite of the participants having no prior experience in writing on similar topics.

2.3 The dictionary used

We chose the APP for the replication out of several considerations. First, as demonstrated by the original research, Bing.dict has undeniable defects, so it is reasonable to try another dictionary to test whether a change of dictionaries would lead to different outcomes. Second, the APP is the first of its kind in China to integrate a prestigious L2–L1 dictionary with a quality L1–L2 dictionary, both produced by the same publisher. It would be interesting to gauge the effectiveness of this new type of dictionary application for language production. Third, unlike the online version of Bing.dict accessed via a computer, the APP is installed in the user's mobile phone. According to dictionary surveys (e.g. Li 2015, Fan 2018, Gao and Yao 2020), mobile dictionary applications have gained immense popularity among Chinese EFL learners. We believe that introducing such a dictionary to the replication can bring the participants closer to their daily dictionary use scenario.

Developed jointly by Foreign Language Teaching and Research Press (FLTRP) and Shanghai Haidi Digital Publishing Technology Co., Ltd., the APP was launched in 2018. It combines *New Century English–Chinese Dictionary* (2016) (ECD)² with *New Century Chinese–English Dictionary* (2nd edition, 2016) (CED)³, representing a new type of "two in one" application which enables users to search words bidirectionally via a "jump" facility between the two source dictionaries.

Take the L1 search word "词典" (*cídiǎn*, dictionary) for example. When users log into the interface of the APP (see Screenshot 1) and enter in the search box the Chinese word "词典", a guide-page (see Screenshot 2) instantly pops up showing three English equivalents to "词典", i.e. *dictionary*, *lexicon*, *wordbook* and other L1 words and phrases containing "词典" and their corresponding English translations. A tap on whichever of the three equivalents to "词典" will lead users to the information page in the L1–L2 source dictionary of the APP, i.e. CED as demonstrated in Screenshot 3 which includes information about the pronuncia-

tion of "词典", its part of speech, meanings/equivalents, and auditory examples containing "词典". Users can further hit on each of the three equivalents to retrieve its specific information in the other source dictionary of the APP, i.e. ECD (L2-L1). For instance, if users tap on *dictionary*, the interface of the APP will "jump" from CED to ECD (see Screenshot 4), switching immediately from L1-L2 search to L2-L1 search (see Screenshot 5). ECD comprises phonetic information (both symbols and auditory), semantic information (both English explanation and Chinese equivalent/translation), and examples (both phrases and sentences) for the headword *dictionary*. It also gives lexical information on its part of speech and inflected form. Moreover, it includes frequency information, specifying that the headword *dictionary* belongs to the vocabulary for TEM 4, University Entrance Examinations and Graduate Admission Examinations, all being essential tests for different levels of students in China. Generally speaking, the design of the APP is simple and clear, offering convenient access routes and making dictionary consultation easy.⁴



Screenshot 1: Interface of the APP



Screenshot 2: Guide-page for "词典"



Screenshot 3: Information given by CED 1



Screenshot 4: "Jump" facility of the APP



Screenshot 5: Information given by ECD

2.4 Methods

Since users' search inputs and entry clicks are documented automatically in the Search Records, it is convenient to collect the data about what and how many words they retrieved and in what order. For example, it is evident from Screenshot 6 that the user looked up 16 words in the APP (9 English and 7 Chinese words) and that the search items were inclusive of both individual words and four-character Chinese idioms. By cross-checking the participants' Search Records and their compositions, we can learn about how they retrieved and applied dictionary information to their writing. In addition, a questionnaire survey was also undertaken to obtain feedback from the participants.



Screenshot 6: Search Records: Example

2.5 Procedure

The replication was implemented in the same experimental setting as the original study. In the first week, the students downloaded the APP into their mobile phones and received a brief training session about its structure, layout and usage⁵. Administered in the following week was Test 1 in which the students had 50 minutes to write a 200-word composition about money saving without any reference tool. One week later, the students were instructed, in Test 2, to write on another topic with the APP at hand (no other reference tool allowed).

After writing, the participants filled in a questionnaire about the usefulness, strengths and weaknesses of the APP and what they valued most for a good writing dictionary (see the Appendix).

2.6 Composition marking and analysis instruments

The compositions were sent to the two same evaluators in the original study and marked according to the same procedure and the same TEM 4 composition scoring rubric. Each composition was given an overall score (max = 20 points) together with separate scores for the three marking components, i.e. content (max = 10 points), structure (max = 3 points) and language use (max = 7 points). To check the inter-evaluator agreement, Pearson correlation coefficients were computed, as presented in Table 1.

Table 1: Correlation coefficient of scores on Test 1 and Test 2 between the two evaluators

	r (content)	r (language use)	r (structure)	r (overall)
Test 1	0.75	0.72	0.60	0.91
Test 2	0.75	0.81	0.62	0.92

Like in the original study, we employed RANGE 32, a lexical analysis tool developed by Heatley, Nation and Coxhead (2002) to compare the words in the compositions with the word lists for reference, including Base word 1 (approximately 1000 most-commonly-used English word families), Base word 2 (approximately 1000 second-commonly-used English word families), and Base word 3 (approximately 570 English word families). LFPs generated by RANGE contain information about (1) tokens, that is, all words in the composition; (2) types, that is, different words in the composition, and (3) families, that is, the base word, its inflections and its most common derivations (Laufer 2005), which can objectively reflect the students' choice and range of lexis. SPSS 20 software was also utilized for statistical processing.

3. Results and discussion

Results of the replication are analyzed and discussed from three aspects. Firstly, the effect of dictionary use on writing performance is examined. Secondly, the impact of the APP on lexical sophistication is evaluated. Thirdly, the differences in dictionary lookup behavior between the participants of the replication and the original study are explained.

3.1 Dictionary use and writing performance

In this subsection, the participants' writing scores in the two tests are computed to ascertain whether the differences are statistically significant. Dictionary-based errors are discussed in relation to what was found in the original study. The questionnaire data are also interpreted regarding the participants' evaluation of the APP.

3.1.1 Analysis of scores

Statistics demonstrated that in Test 1, the lowest and highest overall scores were 11.5 and 18.5 points respectively while in Test 2, the overall scores ranged between 10.0 to 17.0 points. The highest percentage of students (27.4%) in both tests scored between 15.0–15.9 points (see Table 2).

Table 2: Distribution of students' overall scores in the two tests (Max=20 points, N=62)

Overall scores	10.0-11.9	12.0-12.9	13.0-13.9	14.0-14.9	15.0-15.9	16.0-16.9	17.0-17.9	18.0-18.9
Number (Test 1)	2	3	14	7	17	14	4	1
Percentage (Test 1)	3.2%	4.8%	22.6%	11.3%	27.4%	22.6%	6.5%	1.6%
Number (Test 2)	1	4	9	13	17	9	9	0
Percentage (Test 2)	1.6%	6.5%	14.5%	21.0%	27.4%	14.5%	14.5%	0.0%

As displayed in Figure 1, the largest distribution difference in the overall scores across the two tests lies in the area of 14.0–14.9 points, with Test 2 scoring higher than Test 1 by 9.7% (21.0%–11.3%). However, in the areas between both 13.0–13.9 points and 16.0–16.9 points, Test 1 surpassed Test 2 by 8.1% (22.6%–14.5%). It seems that some students performed better when using the APP whereas for others, this was not the case.

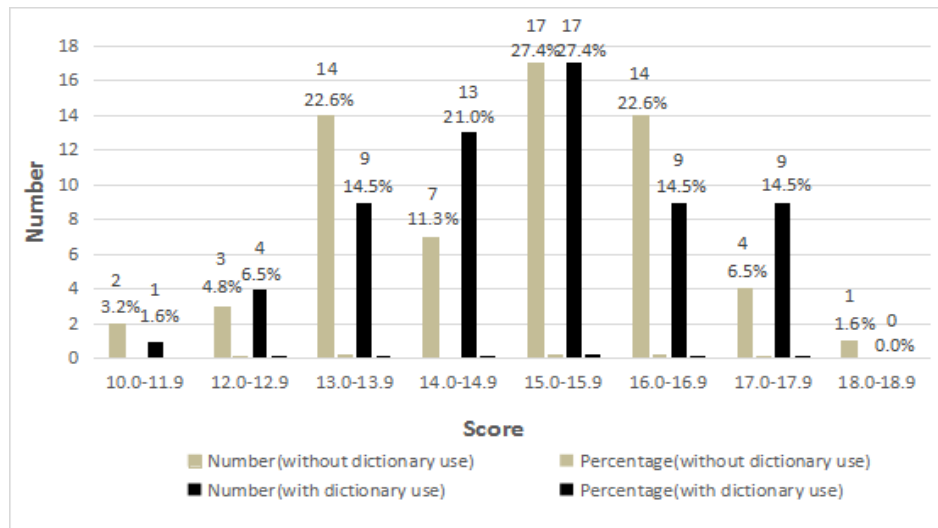


Figure 1: Distribution of the students' overall scores

Paired-Samples T-Tests were run to compute the students' scores (except structure scores) on Test 1 and Test 2. Table 3 indicates that the overall scores on Test 2 ($M=14.95$) are slightly higher than those on Test 1 ($M=14.88$), so are the scores for language use ($M = 5.35$ vs. $M = 5.29$), and the scores for content in the two tests look close ($M=6.99$ vs. $M= 6.95$).

Table 3: Descriptive statistics of scores on Test 1 and Test 2 ($N=62$)

		Mean	Std. Deviation	Std. Error Mean
Pair 1	Content score (Test 1)	6.95	0.68	0.09
(Max= 10)	Content score (Test 2)	6.99	0.67	0.08
Pair 2	Language score (Test 1)	5.29	0.61	0.08
(Max= 7)	Language score (Test 2)	5.35	0.64	0.08
Pair 3	Overall score (Test 1)	14.88	1.49	0.19
(Max= 20)	Overall score (Test 2)	14.95	1.50	0.19

Results of Paired-Samples T-Tests revealed that there was hardly any significant difference between the overall scores across the two tests [$t(61)=-0.355$,

$p=0.724$, two-tailed] (see Table 4). The difference between component scores on the two tests did not reach a significant level either for content [$t(61)=-0.354$, $p=0.725$, two-tailed] or for language use [$t(61)=-0.731$, $p=0.468$, two-tailed]. Apparently, the APP only made a marginal contribution to the students' writing performance.

Table 4: Paired-Samples T-Tests of scores on Test 1 and Test 2

		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1	Content score (Test 1) - Content score (Test 2)	-0.322	0.718	-0.215	0.150	-0.354	61	0.725
Pair 2	Language score (Test 1) - Language score (Test 2)	-0.056	0.608	-0.211	0.098	-0.731	61	0.468
Pair 3	Overall score (Test 1) - Overall score (Test 2)	-0.645	1.433	-0.428	0.299	-0.355	61	0.724

In comparison with the original finding about a markedly negative impact of dictionary use on writing performance, the replication exhibited a more helpful role of dictionary use in writing. As will be illustrated, the APP induced very few dictionary-based errors and received favorable evaluation on its usefulness for writing. Nevertheless, it still failed to exert a significantly positive impact on writing scores. By performing Pearson Correlation analysis, we noticed a weak correlation between the students' overall scores on Test 2 and their frequency of APP consultation [$r=0.25$, $p=0.06 > 0.05$], implying that the higher scorers did not necessarily search more words.

3.1.2 Dictionary-based errors

A cross-examination of the students' Test 2 compositions and their Search Records uncovered 35 dictionary-based errors, inclusive of 14 collocation errors, 10 lexical errors, 8 syntactic errors and 3 other errors (see Table 5)⁶.

Table 5: Distribution of dictionary-based errors

Dictionary-based errors (35)	Lexical errors	10	28%
	Syntactic errors	8	23%
	Collocation errors	14	40%
	Other errors	3	9%

(1) Lexical errors: Some students opted for inappropriate or incorrect English equivalents to express their ideas out of confusion or misunderstanding of the semantic difference between synonymous English equivalents listed in the entries. The following are some examples taken from their compositions.

*"Initially, online dating is easy to be cheated." (The student was confused about the difference between *initially* and *first* when s/he looked up "首先" [shǒuxiān, in the beginning] in the APP.)

*"We should enjoy the convenience and *sake* that online dating brings to us and at the same time keep vigilant." (The student failed to notice the difference between *sake* and *benefit* when they both appeared in the dictionary guide-page for "好处" [hǎochù, benefit].)

*"Recently, the debate on whether it is wise to make friends online has *thrashed out*." (The student misunderstood the meaning of *thrash out*.)

(2) Syntactic errors: A few students committed syntactic errors when applying the retrieved items to writing, ignoring the part of speech or syntactic properties of words. For example:

*"If you *addict* to the friends online, you will be disjointed with people around you."

*"... if both of us realize we don't *fit to* each other."

*"Sometimes we come across private problems in real life and the Internet is a good platform for us to *vent*."

*"We *are too immersed ourselves in* the virtual online dating world, which will affect our interpersonal relationships in real life."

(3) Collocation errors: Some students used the retrievals correctly in grammar, yet the combination of words sounded unnatural. Like the original study, this type of errors made up the majority of the total, for instance:

*"Others make the use of *vulpine communication skills* to earn their trust."

*"Some people will over *indulge in friends* in the network."

*"Not only can it expand our friends circle but it can also *relieve our anxiety anonymously when talking to people* we are familiar with *trickily*."

*"Recent surveys unveil that a large number of young man are more ready to be solitary due to *indulging to the talking* online."

*"We would encounter a variety of persons that could *embrace frauds*."

(4) Other errors: A couple of students made use of words stylistically inappropriately, blind to the style annotations in the entry. To illustrate, one student wrote, "You'll *forlese* the ability of associating with others in reality." It seemed when the student checked the APP, she overlooked the style annotation for *forlese*, i.e. 〈废〉 (*fèi*, obsolete) which indicates that the word dropped out of use. Another case in point is the awkward sentence, *"When you feel *ennuied* and got nothing to do at home ...". The dictionary does proffer an annotation "〈文〉" (*wén*, literary) to specify the style of the word *ennuied*, but the student obviously didn't notice it.

From the analysis above, it can be observed that some of the errors could be attributed to the students' inadequate skills of dictionary use such as choosing equivalents without further looking for their semantic difference, unable to model on dictionary examples to produce natural collocations, and ignoring dictionary annotations or other useful information. Nevertheless, instances of such inappropriate strategies of dictionary use were comparatively rare, most probably thanks to the clear and user-friendly design of the APP. Without the distraction of a multitude of web-crawled lexicographical information, the students had easy access to the reliable information from the two source dictionaries, hence fewer errors. Some of the dictionary-based errors were related to the students' English proficiency, especially their shaky grammatical foundation or weak awareness of collocation.

It should be noted that only two errors were induced by the problems inherent in the APP itself, one of which was due to inaccurate lexical information offered by the APP. It translates *sonnetize* into "沉迷于 (*chénmí yú*, indulge, be addicted, be obsessed with); 把...写入十四行诗 (*bǎ...xiě rù shí sì háng shī*)", the former part being incorrect, thus misleading one student to write, *"I used to *sonnetize* in chatting with congenial net friends" when she searched an equivalent for "沉迷于". The other error resulted from insufficient dictionary examples. The APP renders two translations for *venturesome*, i.e. "好冒险的 (*hǎo màoxiǎn de*, venturesome); 大胆的 (*dàdǎn de*, daring)" and "有风险的 (*yǒu fēngxiǎn de*, risky); 危险的 (*wēixiǎn de*, dangerous)", without any examples to support its detailed use. Consequently, one wrote *"... we are *venturesome* to make friends online."

Compared with the original study in which 34 participants made 106 dictionary-based errors, the replication reported more optimistic data, with only 35 errors from the 62 participants. In other words, the average error rate was 0.56 per person in the replication, much lower than the figure in the original (3.12). With two more reliable source dictionaries and a clearer interface design, the APP

serves users with more accurate lexical information and easier access to dictionary data than Bing.dict, thus causing much fewer errors. The compelling evidence presented by the replication points to the fact that a better dictionary leads to fewer consultation errors, which highlights the importance of dictionary quality.

3.1.3 Responses to the questionnaire

Seven students were excluded from analysis due to incomplete or self-contradictory feedback, leaving a sample size of 55 for the questionnaire survey. According to the survey, none of the students had been familiar with the APP before. Youdao and Eudict are the two applications used most frequently by about 49% and 45% of the sample respectively, followed by Powerword (four students), Baidu Translate (two students), and Collins (one student).

As regards dictionary evaluation, Table 6 manifests that the APP was considered as very useful in rendering assistance for writing, as the mean score reached 9.3 out of a maximum of 10 points. It was also highly rated by a majority of the respondents (Mean=9.2) with respect to its convenience for dictionary search. In terms of the accuracy of dictionary information, the APP earned an average of 8.4 points. It seemed the APP was not as positively evaluated in terms of richness of dictionary information as in other evaluation dimensions, for it received a relatively low score (M=7.9). Generally speaking, the APP gained favorable recognition from the students.

Table 6: Evaluation scores of the APP (Max=10 points, Min = 0 points)

Evaluation score	Highest	Lowest	Mean
Accuracy of information	10	4	8.4
Richness of information	10	3	7.9
Convenience for word search	10	2	9.2
Usefulness for writing	10	2	9.3

This overall evaluation conforms to the responses from the students when asked about the strengths of the APP in comparison with other applications they have used. Half of the students agreed that the APP is more comprehensive in content, featuring rich ancillary learning resources. About 40% of the sample deemed it as more trustworthy due to its accurate lexical information, and some (28%) commented that it is more convenient in use, as it has a neat interface design.

According to over one third (35%) of the respondents, the major weakness of the APP lies in failing to fulfill their need for sentence translation. Some students complained of limited dictionary examples (30%) and lack of access route for phrase and collocation search (28%). Some (26%) responded that the lexical coverage is not wide enough and several students felt it to be expensive.

When it comes to how the APP can improve, 36% of the students hoped for more abundant information, especially on word disambiguation. Roughly one third (34%) expected a more user-friendly APP with more flexible search routes for multiword searches. About 28% desired more accessible examples and phrases. A few (15%) called for specialized columns like a writing guide.

When inquired about their ideas of an ideal writing dictionary, the students' opinions varied. The quality of lexical information was placed at the top by 31% of the sample, followed by a wide lexical coverage (20%), convenient access through keywords (17%) and dictionary brand (15%). A small number of students also maintained that a good writing dictionary should incorporate applicable sentence examples (7%), high quality sentence translations (5%) and should be free-downloadable, ad-free, and upgradable (2%).

3.2 Dictionary use and lexical sophistication

This subsection explores how the APP affected the participants in their choice of lexis during writing through an analysis of LFPs.

Like the original study, the two sets of compositions in the replication were also put into analysis, employing RANGE 32 to examine the students' choice of lexis in terms of tokens, types and families. As illustrated in Table 7, a majority of lexis were taken from Base word 1 with 84.02% for Test 1 and 80.25% for Test 2, indicating that the students mostly relied on the 1,000 most-commonly-used English word families. In Test 2, the tokens from the three Base words decreased respectively from 84.02% to 80.25%, from 5.86% to 4.41% and from 5.44% to 5.27%, whereas those from outside the lists boosted by 5.39% (from 4.68% to 10.07%). Likewise, in Test 2 the types from the three Base words presented a uniformed decreasing tendency compared with Test 1 while those from outside the lists went up from 20.81% to 27.75%. With regard to the families of lexis, except for Base word 2, there was an increase in both Base word 1 and Base word 3 by 38 and 15 families respectively. All this suggests that the students made use of less basic words and preferred more academic ones when accessible to the APP. By cross examining the students' Search Records and the results of LFP analysis for outside the lists in Test 2, we made a list of words retrieved from the APP like "abyss", "alienate", "authenticity", "celebrity", "detrimental", "harassment", "intangible", and "recap" etc. An overwhelming proportion of the words were searched only once and except for several misspelt words, there were no unattested or non-existent words.

Table 7: LFPs of the two sets of compositions

Profiles	Base word 1		Base word 2		Base word 3		Not in the lists	
	Test1	Test 2	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
Tokens/%	13153/ 84.02	13350/ 80.25	917/ 5.86	733/ 4.41	852/ 5.44	877/ 5.27	732/ 4.68	1675/ 10.07
Types/%	977/ 49.00	1025/ 45.07	299/ 14.99	290/ 12.75	303/ 15.02	328/ 14.42	415/ 20.81	631/ 27.75
Families	559	597	223	209	204	219	?	?

As Independent-Samples T-Tests (see Table 8) revealed, the differences in both tokens and types from "not in the lists" reached a statistically significant level ($p < 0.01$), implying that the students tended to use more advanced and sophisticated words when the APP was available. In addition, the tokens from Base word 2 between the two sets of lexis also differed remarkably from each other ($p < 0.01$). Although no substantial difference was found between the two sets of lexis from Base word 1 and Base word 3, there was an observable fall in the use of high-frequency words and a noticeable increase in more complex ones.

Table 8: Independent-Samples T-Tests on LFPs

	Base word 1	Base word 2	Base word 3	Not in the lists
Tokens				
t	-0.961	2.672	-0.579	-6.728
df	122	122	122	122
p	0.338	0.009**	0.564	0.000**
Types				
t	-1.892	0.492	-0.590	-2.593
df	122	122	122	122
p	0.061	0.623	0.556	0.004**
Families				
t	-1.204	0.267	-0.843	
df	122	122	122	
p	0.231	0.790	0.401	

** $p < .01$ (two-tailed).

The replication confirmed the original finding about the impact of dictionary use on lexical sophistication. However, despite the enhanced lexical range, scores on Test 2 were only marginally higher than those on Test 1 (see Table 3). In other words, the extent of richer lexis was not large enough to have a significant effect on the scores.

3.3 Dictionary lookup behavior

This subsection identifies the differences in dictionary lookup behavior between the participants of the replication and the original study, looking into the questions about who consulted the dictionary more frequently, what the preferred language input was, L1 or L2, what kind of lexical items were searched most often, and how the participants made use of the source dictionaries differently.

3.3.1 Frequency of dictionary searches

The data about the frequency of dictionary consultation was gathered from the Search Records which encompassed the words entered in the search bar and the items in the guide-page or the entries tapped for further or cross-reference. As displayed in Table 9, the students looked up 884 lexical items altogether with an average of 14.3 per person. The frequency of dictionary lookup varied from 2 to 47. Evidently, the students in the replication turned to the APP more frequently than those in the original study.

Table 9: Frequency of dictionary searches

	Total number	The average	The maximum	The minimum
The original (n=34)	405	11.9	31	0
The replication (n=64)	884	14.3	47	2

As evidenced in Table 10, the students performed more L2–L1 than L1–L2 consultation (542 vs. 342). A dominant number (92%) of L2–L1 searches were individual English words with one student consulting as many as 41 items. L2 multiple-word combinations accounted for no more than 8%, mostly phrasal verbs such as *bear upon*, *wear down*, and *fan out*, etc. Among the L1–L2 searches, individual Chinese words constituted the bulk (85%) with a maximum of 17 words, followed by 14% of four-character Chinese idioms and phrases such as "不知所措" (*bùzhīsuǒcuò*, all at sea), "喜怒哀乐" (*xǐnùāilè*, joy, anger, sorrow

and happiness), "难言之隐" (*nányánzhīyǐn*, *painful secret*), "随时随地" (*suíshísuǐdì*, *anytime and anywhere*), "网上聊天" (*wǎngshàng liáotiān*, *cyber chat*), "主旋律" (*zhǔ xuánlǜ*, *theme*) etc. There were only three search cases for Chinese sayings like "不怕一万, 只怕万一" (*bù pà yī wàn, zhǐ pà wàn yī*, *be prepared for the one risk in a million*). Most students switched between L1–L2 and L2–L1 searches, with six performing L2–L1 consultation exclusively.

Table 10: Descriptive statistics of the students' Search Records (N=62)

Total searches	L1–L2 searches			L2–L1 searches	
884	342			542	
Mean=14.3	L1 words	L1 idioms and phrases	L1 sayings	L2 words	L2 multiple-word combinations
	290/85%	49/14%	3/1%	500/92%	42/8%
	Mean=4.68	Mean=0.79	Mean=0.05	Mean=8.06	Mean=0.68
	Max=17	Max=7	Max=1	Max=41	Max=5
	Min=0	Min=0	Min=0	Min=0	Min=0

It is noteworthy that only two students applied all the words and phrases they retrieved from the APP to their writing. Of the 884 search items, 541 were actually put into the compositions. The average use of lexicographical information was about 61%. Most students utilized 40–80% of their retrievals, with the exception of three who used only about a tenth of their lookups.

3.3.2 Preference for language search

In the original study, the participants mostly attempted to obtain L2 equivalents or translations with L1 inputs dominating the scene. No one carried out L2–L1 searches exclusively. In contrast, the replication reported that the frequency of L2–L1 searches was 1.6 times that of L1–L2 searches (see Table 10). Six students conducted L2–L1 searches solely without a single L1 input. This difference might result from data collection method. The APP only records the lexical items that appear in the headwords or examples in the two source dictionaries. Consequently, inputs of some Chinese phrases, collocations and sentence fragments such as "潮湿的空气" (*cháoshī de kōngqì*, *moist air*), "开展活动" (*kāizhǎn huódòng*, *carry out activities*), "谨防上当受骗" (*jǐnfáng shàngdāngshòupiàn*, *beware of being cheated*) cannot be documented in the Search Records unless the translations of such items are included in the APP either as headwords or as entry examples. In other words, the APP cannot keep track of the Chinese phrases,

collocations or other multiword expressions which go beyond the coverage of the source dictionaries.

3.3.3 Preference for search items

One notable lookup pattern discovered by the original study was that many participants tended to seek English translation for Chinese sentences. This formed a striking contrast with the replication where the searches of individual words made up the bulk. Bing.dict features a multitude of web-crawled lexicographical information as well as automatic machine translation, which renders sentence translation possible, but in many cases the translation is of poor quality, if not ridiculous, and only misleads users. By comparison, since the source dictionaries in the APP are essentially the electronic versions of the original print dictionaries without fundamental changes in content and structure, they are unable to cater for users who try to look for sentence translation. Moreover, the link directing users to Google Translate when the lookup items go beyond the lexical coverage of the APP is currently inaccessible in Mainland China.

Another difference consists in the consultation of basic words. Some participants in the original study looked up high frequency Chinese words like "通常" (*tōngcháng*, usually), "其次" (*qícì*, next), and "第二" (*dìèr*, secondly). However, the replication identified only two such instances. This divergence might arise from the participants' overall English proficiency. Despite the similar linguistic proficiency of the participants, students in the replication achieved a higher, though not significantly, average score on Test 2 than those in the original study ($M = 14.9$ vs. $M = 12.8$), which implied that they can use English more competently and may not feel the need to look up high frequency words.

3.3.4 Use of source dictionaries

The original participants mostly depended on Internet-generated lexicographical information, neglecting the source dictionaries in Bing.dict. Only 9 out of 32 participants further clicked on them for cross-reference and no one ever hit on the tabs in the bilingualized source dictionary to read examples. The replication showed a different picture, for as many as 54 students switched between L1-L2 and L2-L1 searches, meaning the majority of the students made use of both source dictionaries.

This difference can be ascribed to the interface design of the dictionaries involved. In Bing.dict, the Internet-generated translations and web-crawled sentence examples are posted in a conspicuous spot, dwarfing the source dictionaries on that score. In contrast, the APP is based on only two source dictionaries without accessible links to extra lexicographical resources. Its interface looks clean and clear, making easy the "jump" from one dictionary to the other. Without access to extra lexicographical information from the Internet, users have no alternative but to focus on the two source dictionaries.

4. Conclusion

Three findings emerge from the replication. Firstly, the use of the APP has a non-significant effect on the participants' writing performance, distinct from the original conclusion about the negative role of dictionary in EFL writing. The APP proves to be more helpful for encoding, as it gave rise to a much smaller number of dictionary-based errors than the original dictionary, suggesting that the better a dictionary is, the fewer consultation errors it will cause.

Secondly, dictionary use did enhance the participants' lexical sophistication, although this advantage was not significant enough to make a marked difference in writing scores. This conforms to the original conclusion.

Thirdly, some differences in dictionary lookup behavior were detected between the participants of the replication and the original study. Students in this study consulted the APP more frequently. They entered or tapped on more L2 items than L1 ones, mainly looking up individual words and paying more attention to the source dictionaries than the original participants. Moreover, they committed far fewer dictionary-based errors, chiefly owing to the more authoritative source dictionaries and the well-designed dictionary interfaces.

The study shows that the APP, as a pioneering "two in one" product at the Chinese lexicographic market, is more effective than Bing.dict for EFL writing, yet it did not exert a significantly positive effect on writing performance. The questionnaire survey reflected that the APP was highly ranked in terms of usefulness for writing, accuracy of dictionary information and convenience in use. However, it was also perceived to have some weaknesses such as lack of search function for some phrases, collocations and sentences, limited lexical coverage, and insufficient dictionary examples, etc. The participants expressed their hopes for more useful lexical information such as word disambiguation, a wider vocabulary coverage, richer dictionary examples, and easier access routes to multiword search. To better satisfy users' needs and expectations, improvements should be made to the APP with regards to the above-mentioned areas.

According to the questionnaire survey, the participants held a variety of opinions concerning the criteria for an ideal electronic dictionary for EFL writing. It seemed the quality of dictionary information was prioritized by most students, followed by multiple and convenient access routes. Practical information categories such as derivatives, synonym disambiguation, collocations, sentence examples, and a writing guide were also among the list. Conceivably, the participants' feedback can serve as useful advice for dictionary optimization.

ECD and CED are rated among the best bilingual dictionaries in China, receiving positive recognition from lexicographical experts and users (Wang et al. 2019). However, it should also be pointed out that they are general linguistic dictionaries in nature, different from learners' dictionaries in at least four major aspects, i.e. the target users, lexical coverage, sense arrangement and the amount of lexical information crucial for language production such as collocations and examples. Take lexical coverage for example. The APP covers

low-frequency words and archaic or even obsolete words, which may lead users to select unfamiliar words due to the misconception that the rarer the word, the better. By comparing collocations and examples in the APP with those of LDOCE (*Longman Dictionary of Contemporary English*)⁷, one of the most well-known English learners' dictionaries in the world, we believe that the former leaves much to be desired in these aspects. The APP could have a more beneficial effect on writing if it had been equipped with a better encoding function. It is a pity that despite the remarkable progress in China's practical lexicography, there is still a long way to go in the compilation of production-oriented L1–L2 learners' dictionaries.

5. Limitations of the study and suggestions for future research

The study is not without limitations. Since the Search Records can only keep track of search items which fall within the lexical coverage of the two source dictionaries in the APP, we were unable to know what and how many invalid searches were performed by the participants. Such a problem could have been avoided if we had relied on screen recording for data collection as we did in the original study. Besides, due to the restriction of the research method, we could not learn about the cognitive aspects of dictionary consultation such as what prompted a particular search, how the participants chose among equivalents, how they dealt with lexical issues when failing to retrieve needed information from the APP, and why many L2 lookups were not used in writing. Think-aloud protocols or a follow-up interview would help to elicit some interesting information about the students' cognitive processes and strategy use.

The dictionary per se is a crucial variable when testing the effect of dictionary use. To gain more insights into the impact of dictionary use on writing, it is advisable to carry out more replications. Future research may try another type of dictionary, especially production-oriented dictionaries with user-centered design. Dictionary use competence constitutes another important factor influencing the outcomes of dictionary consultation. Therefore, it is necessary to involve participants of different proficiency levels such as English-majoring MA students or skilled dictionary users in further studies. In addition, other types of writing tasks (e.g. free-topic writing), or other forms of language production (e.g. L1–L2 translation), are also considerable. Finally, more explorations can be attempted to develop writing assistants and check their effectiveness for language production.

Endnotes

1. Approximate (also known as partial or systematic) replication involves repeating the original study exactly in most respects, but changing one of the non-major variables so as to allow for comparability between the original and replication study (Abbuhl 2012: 298).

2. Based on the Collins Corpus, ECD is a general linguistic dictionary with a coverage of over 250,000 words and 350,000 senses. It encompasses concise and accurate definitions, rich and comprehensive information and a wide coverage of new words and senses, representing the status quo of the English language.
3. CED is hailed as the first work of the fourth generation Chinese–English dictionaries. It covers more than 150,000 headwords, highlighting linguistic information and including encyclopedic information. It is an official dictionary used for China Accreditation Test of Translators and Interpreters. Due to its innovation in terms of lexical coverage, definition, examples, translation, and part-of-speech tagging, it has won several national awards.
4. The APP has remarkable user-friendly features such as a multitude of functions for customization in language learning and abundant learning resources including the Chinese ideological and cultural terminology, special Chinese–English columns (such as Chinese four-character idioms, proverbs and particularized sayings), and special English–Chinese columns (such as English phrases and idioms, usage notes, cultural columns and collocations). Since those features were irrelevant to the replication, they are mentioned briefly here.
5. Thanks to the general support of FLTRP, all students in the experiment had a three-month free access to the APP.
6. Due to the complexity and challenges involved in the classification of errors, some of the errors might fall into more than one category, some might be borderline cases and some might be hard to categorize.
7. *Longman Dictionary of Contemporary English Online* can be accessed via <https://www.ldoceonline.com/>.

Acknowledgements

This research is part of the project "Constructing an Evaluation Framework for Bilingual Dictionary APPs in the Digital Era" (No. FJ2023B031) funded by the Fujian Social Science Foundation. We would like to extend our sincere gratitude to the anonymous reviewers who dedicated their precious time to reading our paper and making insightful comments and suggestions. We are also grateful to those students who participated in the experiment. In addition, heartfelt thanks also go to our colleagues, Professor Hou and Ms. He, who helped to proofread our paper.

References

- Abbuhl, R. 2012. Why, When, and How to Replicate Research. Mackey, A. and S.M. Gass (Eds.). 2012. *Research Methods in Second Language Acquisition: A Practical Guide*: 296-312. Oxford: Blackwell Publishing.
- Boonmoh, A. 2012. E-dictionary Use under the Spotlight: Students' Use of Pocket Electronic Dictionaries for Writing. *Lexikos* 22: 43-68.
- Chen, Y.Z. and S.P. Liu. 2022. Exploring the Use of an Online Bilingual Dictionary in EFL Writing. *International Journal of Lexicography* 35(4): 468-490.

- Chon, Y.V.** 2009. The Electronic Dictionary for Writing: A Solution or a Problem? *International Journal of Lexicography* 22(1): 23-54.
- Christianson, K.** 1997. Dictionary Use by EFL Writers: What Really Happens? *Journal of Second Language Writing* 6(1): 23-43.
- 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. *ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives*: 92-101. Kyoto: Asian Association for Lexicography.
- Dziemianko, A.** 2012. Why One and Two Do Not Make Three: Dictionary Form Revisited. *Lexikos* 22: 195-216.
- Dziemianko, A.** 2017. Dictionary Form in Decoding, Encoding and Retention: Further Insights. *ReCALL* 29(3): 335-356.
- East, M.** 2006. The Impact of Bilingual Dictionaries on Lexical Sophistication and Lexical Accuracy in Tests of L2 Writing Proficiency: A Quantitative Analysis. *Assessing Writing* 11(3): 179-197.
- East, M.** 2007. Bilingual Dictionaries in Tests of L2 Writing Proficiency: Do They Make a Difference? *Language Testing* 24(3): 331-353.
- Fan, K.** 2018. An Investigation of English Major Students' Use of English Dictionaries. *Lexicographical Studies* 6: 24-32.
- Gao, X. and K.X. Yao.** 2020. Survey and Prediction of the Usage of College Students' Mobile Phone Dictionary. *Jiang Su Technology and Information* 37(19): 63-65.
- Gast, D.L. (Ed.).** 2009. *Single Subject Research Methodology in Behavioral Sciences*. New York: Routledge.
- Heatley, A., P. Nation and A. Coxhead.** 2002. RANGE and FREQUENCY Programs [EB/OL]. <http://www.vuw.ac.nz/lals/staff/Paul-Nation>. [2021-03-20]
- Lai, S.-L. and H.-J. Howard Chen.** 2015. Dictionaries vs Concordancers: Actual Practice of the Two Different Tools in EFL Writing. *Computer Assisted Language Learning* 28(4): 341-363.
- Language Teaching Review Panel (LTRP).** 2008. Replication Studies in Language: Learning and Teaching: Questions and Answers. *Language Teaching* 41: 1-14.
- Laufer, B.** 2005. Lexical Frequency Profiles: From Monte Carlo to the Real Word. A Response to Meara. *Applied Linguistics* 26(4): 582-588.
- Lew, R.** 2016. Can a Dictionary Help You Write Better? A User Study of an Active Bilingual Dictionary for Polish Learners of English. *International Journal of Lexicography* 29(3): 353-366.
- Lew, R. and A. Szarowska.** 2017. Evaluating Online Bilingual Dictionaries: The Case of Popular Free English-Polish Dictionaries. *ReCALL* 29(2):138-159.
- Li, J.Y.** 2015. A Survey of Mobile-phone Dictionary Use by University Students in China's Remote Minority Regions. *Lexicographical Studies* 3: 38-47.
- Nesi, H.** 2012. Alternative e-Dictionaries: Uncovering Dark Practices. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 369-378. Oxford: OUP.
- Nesi, H. and P. Meara.** 1994. Patterns of Misinterpretation in the Productive Use of EFL Dictionary Definitions. *System* 22(1): 1-15.
- Park, C.** 2004. What Is the Value of Replicating Other Studies? *Research Evaluation* 13(3): 189-195.

- Qiao, L.T. and W.Y. Wang.** 2020. An Empirical Study of the Impact of Dictionary Use on Learners' Lexical Performance in L2 Writing: Chinese L2 Learners as an Example. *Foreign Languages Research* 183(5): 36-42.
- Tall, G. and J. Hurman.** 2002. Using Dictionaries in Modern Language GCSE Examinations. *Educational Review* 54(3): 205-217.
- Wang, R.Q., Z.Z. Huo and J. Deng.** 2019. A Study of the Representation Strategies of Heterosemy in *A New Century Chinese-English Dictionary* (2016). *Foreign Languages and Literature* 35(2): 11-22.

Appendix

A questionnaire on the use of the application of *New Century English–Chinese and Chinese–English Dictionary* (the APP)

Instructions: Please write an answer to or put a tick at the answer of the following questions.

1. Was it the first time you used this APP? Yes. No.
2. Please write the name of the mobile phone APP you use most often.

3. Please rate your satisfaction of the APP (from low to high, the full score is 10) according to your dictionary use experience.
 - ☆ the accuracy of dictionary information: 1 2 3 4 5 6 7 8 9 10
 - ☆ the richness of dictionary information: 1 2 3 4 5 6 7 8 9 10
 - ☆ the convenience for dictionary research: 1 2 3 4 5 6 7 8 9 10
 - ☆ the usefulness of the APP for writing: 1 2 3 4 5 6 7 8 9 10
4. Compared with other electronic dictionaries you use, what do you think are the advantages of the APP?
5. In your opinion, what are the weaknesses of the APP?
6. In what ways can the APP improve?
7. Do you think the APP is affordable or not at a price of 138 RMB?
8. What do you think are the features of an ideal electronic dictionary for writing?

Developing Dictionary Skills through Monolingual and Bilingual English Dictionaries at Tertiary-level Education in Hungary

Katalin P. Márkus, *Institute of English Studies,
Department of English Linguistics, Károli Gáspár University
of the Reformed Church in Hungary, Budapest, Hungary*
(p.markus.kata@kre.hu)

and

Ida Dringó-Horváth, *ICT Research Centre / Educational
Technology Training Centre, Károli Gáspár University of the
Reformed Church in Hungary, Budapest, Hungary*
(dringo.horvath.ida@kre.hu)

Abstract: Among reference works, dictionaries are particularly important in foreign language learning. Dictionaries provide language learners with a wide range of data, however, wading through the mass of data and information can be a daunting task. Mastering dictionary skills should be important in the language learning process; however, in official educational documents in Hungary, there are no clear guidelines on how to develop these skills. By integrating dictionary skills into the curriculum and teaching them explicitly in lessons, teachers could play an important role in bridging the gap between lexicographers and dictionary users. In the present study, we report on our methods of teaching lexicography and dictionary skills to students at a Hungarian university. The authors are speaking from experience, the discussion and accompanying material are based on more than 15 years of teaching practice. To ensure that the training can meet the expanding needs of young students, a longitudinal study was launched in 2020 to examine students' changing habits and needs. The aims of the article are threefold: first, to encourage the teaching of lexicography at university level by providing concrete methods, then to highlight the importance of dictionary skills, and finally, to emphasise the importance of integrating the effective teaching of the use of electronic dictionaries into dictionary didactics. In this context, the article underlines the need to incorporate new evaluation criteria as well as to develop new skills for digital dictionaries, different from those for print dictionaries, into education.

Keywords: TEACHING LEXICOGRAPHY, DICTIONARY DIDACTICS, REFERENCE SKILLS, DICTIONARY SKILLS, DICTIONARY USE, UNIVERSITY COURSE DESIGN, COURSE EVALUATION, ONLINE DICTIONARIES, EVALUATION OF DICTIONARIES

Opsomming: Die ontwikkeling van woordeboekvaardighede m.b.v. eentalige en tweetalige Engelse woordeboeke in die onderwys op tersiêre vlak in Hongarye.

Onder naslaanwerke is woordeboeke besonder belangrik in die aanleer van 'n vreemde taal. Alhoewel woordeboeke taalaanleerders van 'n wye reeks data voorsien, kan dit 'n enorme taak wees om deur die massa data en inligting te worstel. Ofskoon dit belangrik behoort te wees om woordeboekvaardighede in die taalaanleerproses te bemeester, bestaan daar geen duidelike riglyne in amptelike onderwysdokumente in Hongarye oor hoe om hierdie vaardighede te ontwikkel nie. Deur woordeboekvaardighede in die kurrikulum te integreer en dit eksplisiet in lesse te onderrig, kan onderwysers 'n belangrike rol speel om die gaping tussen leksikograwe en woordeboekgebruikers te oorbrug. In hierdie studie word verslag gelewer oor die metodes wat ons gebruik om leksikografie en woordeboekvaardighede aan studente by 'n Hongaarse universiteit te onderrig. Die outeurs beskik oor baie ervaring en die bespreking en bygaande materiaal is op meer as 15 jaar se onderrigpraktyk gebaseer. Om te verseker dat die opleiding in die toenemende behoeftes van jong studente kan voorsien, is 'n longitudinale studie in 2020 van stapel gestuur om die veranderende gewoontes en behoeftes van studente te bestudeer. Die doel met hierdie artikel is drieërlei van aard: eerstens, om die onderrig van leksikografie op universiteitsvlak aan te moedig deur konkrete metodes te verskaf, daarna, om die belangrikheid van woordeboekvaardighede uit te lig, en laastens, om die belangrikheid van die integrasie van effektiewe onderrig van die gebruik van elektroniese woordeboeke in die woordeboekdidaktiek te beklemtoon. Teen hierdie agtergrond word die behoefte aan die inkorporering van nuwe evaluasiekriteria sowel as die ontwikkeling van nuwe vaardighede vir digitale woordeboeke, verskillend van dié vir gedrukte woordeboeke, in die onderwys beklemtoon.

Sleutelwoorde: LEKSIKOGRAFIEONDERRIG, WOORDEBOEKDIDAKTIEK, NASLAAN-VAARDIGHEDE, WOORDEBOEKVAARDIGHEDE, WOORDEBOEKGEBRUIK, ONTWERP VAN UNIVERSITEITSKURSUSSE, KURSUSEVALUERING, AANLYN WOORDEBOEKE, EVALUERING VAN WOORDEBOEKE

Introduction

In today's fast-changing online world, we need to have a wide range of knowledge — but not necessarily in our heads. Students need an education that builds a solid foundation for life-long learning to compete with advanced professional skills in our rapidly changing society, in which new knowledge must constantly be acquired and integrated (European Commission 2019; RFCD 2018). The dictionary is one of the first reference sources students should learn to use and in turn, these skills will help to prepare them for university-level work since it is one of the most appreciated and widely used resources for learning a language (cf. Lew 2016; Nied Curcio 2022). In the hands of a skilled learner, a dictionary is an invaluable resource and once we learn how to use it successfully, it opens the way to autonomous learning, which puts the power in the students' hands. What is more, these skills play a crucial role in supporting lifelong learning (Campoy-Cubillo 2015).

When talking about dictionaries, first it is important to clarify that there are many types of dictionaries available today, each of which aims to meet a different need (see Engelberg and Lemnitzer 2009; Wiegand et al. 2010). As far as the form of publication is concerned, one can distinguish between print and electronic dictionaries. Typologies of electronic dictionaries can be developed on the basis of different features (cf. Wiegand et al. 2010), but it is fundamental to distinguish between electronic dictionaries for the use of people or of machines/software (e.g., translation software) — the latter is not the subject of this article. Read (2023) notes the other specific types of data dictionaries may contain (e.g., pronunciation, grammatical forms, etymologies, syntactic peculiarities, variant spellings) and the various roles they can play in different contexts, emphasising the diversity of dictionaries from the viewpoint of their objectives and content. Read (2023) also highlights the role of the dictionary in the learning process, stating that "dictionaries can encourage schoolchildren to learn about language". Dictionaries, especially monolingual and bilingual learner's dictionaries, are still considered to be very useful language learning tools in today's digital world where we are surrounded by technology such as AI (Artificial Intelligence), and MTs (Machine Translators) (cf. Campoy-Cubillo 2015; Lew and De Schryver 2014; Nied Curcio 2022). At the same time, the role of dictionaries is constantly changing, with more and more digital tools entering the educational process. In the future, we will have to learn to collaborate and train students on how and when to use the wide range of resources effectively (cf. De Schryver and Joffe 2023; Jakubiček and Rundell 2023; Lew 2023; Lew and De Schryver 2014). Students who have mastered the use of reference works will find it simpler to use a variety of reference materials as they progress from primary school to secondary school and then to university.

In the language learning process, it should be the task of language teachers to draw students' attention to the effective use of monolingual and bilingual learner's dictionaries and other reference works, which assist language learners in reading, writing, and translation activities (cf. Campoy-Cubillo 2002, 2015; Nied Curcio 2022). For this reason, teachers should be prepared for this task within the framework of teacher training. The use of reference works should also be emphasised as part of the language learning process in universities to foster lifelong learning and learner autonomy, substituting for the human teacher in addressing language problems. All these background events clearly outline some important steps to be taken in order to provide training for teachers to fill this gap (cf. Atkins and Varantola 1998; Cowie 1999, Chi 2003; Dringó-Horváth 2017; Lew 2011; Lew and Galas 2008; P. Márkus 2020a; P. Márkus and Pődör 2021; P. Márkus et al. 2023; Sinclair 1984; Yamada 2010).

With this article, we would like to contribute to the enrichment of studies on the teaching of lexicography (and dictionary skills) as an academic subject. The more widely we share our academic experiences, the more insight we can provide into different practical teaching methodologies and their effectiveness (for more on the subject, see for example Bae 2011; Béjoint 1989; Chi 1998;

Hartmann 2001; Magay 2000; Martynova et al. 2015; Nkomo 2014; Prčić 2020; P. Márkus 2020b). The main idea of this article was inspired by Prčić (2020), who describes a university course in lexicography that was specifically designed and developed for advanced EFL (English as a Foreign Language) students. Prčić advocates for the teaching of both theoretical and practical lexicography to university students by offering specific recommendations. In the present study, a similar approach is attempted, with the difference that the course is complemented by the development of dictionary skills and its methodology. In light of the above-mentioned, the aims of the article are threefold: first, to encourage the teaching of lexicography at university level by providing concrete methods, then to highlight the importance of dictionary skills, and finally, to emphasise the importance of integrating the effective teaching of the use of electronic dictionaries into dictionary didactics. In the following section, the related research will be outlined, which has been conducted to support the teaching of "Lexicology and Lexicography" as an academic subject at Károli Gáspár University of the Reformed Church in Hungary.

1. Researching didactic lexicography and dictionary use

Many research projects at the Faculty of Humanities and Social Sciences at Károli Gáspár University of the Reformed Church in Hungary are focused on dictionary didactics, dictionary use and how dictionaries could best be employed in the context of language teaching and learning. As members of these projects, the authors have been investigating EFL and GFL (German as a Foreign Language) students' dictionary use habits and strategies in the Hungarian context since the early 2000s. The research projects were launched to help develop teaching materials for lexicography courses. Based on the results, we can offer courses tailored to the needs of students.

Márkus and Szöllősy (2006) used a questionnaire in conjunction with a set of tasks to test whether secondary school EFL students could use their dictionaries effectively. Participants (n=122) were asked about their dictionary use habits and a set of practical exercises was used to identify the factors that make it difficult to use a dictionary. The tasks focused on the following main topics: interpreting abbreviations and symbols used in dictionaries; using grammatical information; finding meanings, collocations, idioms, and phrasal verbs. The findings revealed that students' language awareness and dictionary use awareness were very low. As a result, it was suggested that dictionary training be included in school and academic curricula (Márkus and Szöllősy 2006). The survey provided valuable insights that could be implemented in teaching aids, such as workbooks, and study pages in coursebooks to develop dictionary skills (P. Márkus 2020b; P. Márkus 2023; P. Márkus et al. 2023).

In her survey of undergraduates (n=80), Dringó-Horváth (2017) indicated that online dictionaries were becoming more and more popular. Approxi-

mately half of the respondents claimed that they used online dictionaries to look up the meaning of a word or expression, whereas just a quarter reported that they used print dictionaries for this reason. When looking for pronunciation information, the dominating influence of online dictionaries appeared to be obvious, with almost 60% of respondents reporting that they used online dictionaries. Due to the difficulty of interpreting phonetic symbols, print dictionaries have already played a very minor role in this activity. The final section of the survey was particularly important since it addressed the learning of dictionary skills. The vast majority of respondents stated that they acquired those skills in a self-taught way (Dringó-Horváth 2017). The results appear to confirm the research of Márkus and Szöllősy (2006), who found that only a small percentage of respondents received training in dictionary use at school and university.

In her study, P. Márkus (2020a) discussed the state of dictionary culture and dictionary didactics in Hungary by analysing educational documents (*Common European Framework of Reference for Languages, National Core Curriculum, Framework Curriculum, syllabuses*). The analyses showed that in Hungary 'dictionary awareness' was generally rather low and that more attention to the teaching of dictionary skills was needed in the curricula for foreign language learning. The study also demonstrated the use of a dictionary workbook to remedy the situation. Designed to accompany an English–Hungarian/Hungarian–English Learner's Dictionary, the workbook contained exercises and activities which aimed to help language learners in two ways: first, by teaching the basic dictionary skills that students need to locate headwords or expressions and their meanings in the dictionary; and second, by showing how the dictionary can be used as a tool and a source of information about the English and Hungarian languages. The workbook was piloted by teachers and students at primary and secondary schools, and it was designed to be used both in the classroom and for self-study. In 2020, P. Márkus et al. (2023) launched a longitudinal research project on the dictionary use habits of EFL and GFL students, as well as their attitudes towards learning and teaching dictionary skills in the L2 classroom. The aim of the project is to assess students' dictionary use habits and needs at three-year intervals in order to tailor dictionary training to those needs.

The results of international surveys and monitoring of dictionary use seem to show a dismal picture of language learners' reference skills (see for example Atkins and Varantola 1998; Bogaards 1998; Chan 2012; Dringó-Horváth 2017; Márkus and Szöllősy 2006; Müller-Spitzer et al. 2018; Nesi and Hail 2002; Nesi and Meara 1994; Nied Curcio 2022). Among other things, it is a common problem that language learners are not aware of where to start looking up the information they require; where to find answers to their questions about words, fixed expressions, synonyms and grammar — or if they do find the answer, they often cannot interpret the data found (cf. Frankenberg-Garcia 2011; Müller-Spitzer et al. 2018). Usual issues with interpretation are mostly caused by the lexicographical codes, symbols and abbreviations used in dictionary

practice. Moreover, a lot of previous research reports that very little training in dictionary use takes place in schools (cf. Dringó-Horváth 2017; Márkus and Szöllősy 2006; Nied Curcio 2022; P. Márkus et al. 2023).

The development of dictionary skills should start in primary school and be progressively developed in secondary school as the language level rises, in order to ensure quality knowledge and skill development. Even university students, especially teacher trainees, need to learn to use (online) dictionaries properly (cf. Campoy-Cubillo 2015; Nied Curcio 2022) because if students learn how to use dictionaries effectively, they will be able to teach it more successfully. It is encouraging to see that more and more universities are recognising the importance of teaching lexicography and dictionary use (cf. Bae 2011; Hartmann 2013; Magay 2000; Nied Curcio 2022; P. Márkus 2020a; P. Márkus and Pődör 2021; Prčić 2020).

With the given objectives in mind, focus now shifts to the "Lexicology and Lexicography" course.

2. Lexicography and dictionary use as a university course at Károli Gáspár University of the Reformed Church in Hungary

2.1 About the course in general

In the area of pedagogical lexicography, there is widespread agreement that dictionaries should be more and more user-oriented and at the same time, users should be trained to use dictionaries successfully (see for example Bae 2011; Lew 2016; Nied Curcio 2022; P. Márkus et al. 2023). The aims and findings outlined in section 1 suggest that it would be useful to complement theoretical lexicographic teaching in the university context with dictionary didactics, and the teaching of dictionary use. In the following section, a course in "Lexicology and Lexicography", which has been taught since the academic year 1996/1997 (at the Department of English Linguistics), will be outlined. The lexicology section of the course serves as a foundation for the various lexicographic topics that follow. The goal is to introduce the theoretical and practical connections between these two key areas and to provide all the necessary elements needed for practical training in the use of dictionaries.

Károli Gáspár University was founded in 1993 and it was the first among Hungarian universities to introduce the teaching of lexicography (for more information on the original course design, see Magay 2000). In 1996, Tamás Magay devised the course, which since his retirement has been taught by Dóra Pődör and Katalin P. Márkus, the developers of the course and authors of the course's teaching materials. "Lexicology and Lexicography" is a compulsory elective course. It lasts 12 weeks and has two classes per week for between 20 and 30 English and German as a Foreign Language teacher trainees. During the semester, students learn all the basic concepts of lexicology and lexicography, gain

insight into research on dictionary use, dictionary analysis, and learn about print and electronic dictionaries relevant to language learning. After exploring the different types of dictionaries, the emphasis is placed on general synchronic descriptive (English monolingual and bilingual) dictionaries, and therefore no in-depth study of etymological or specialised dictionaries is undertaken. When the students leave university, they will need to know for what purpose, for what age group, and what type of dictionary to use, so they should have sufficient knowledge of reliable monolingual and bilingual English dictionaries. In addition, they will need to be familiar with methods of teaching dictionary use. In selecting the course material, we have to take into account that on the theoretical and practical level of lexicography, we have four main actors: the *dictionary-maker* compiles the dictionary, which is purchased by the *user*, who is often supported by the *teacher* who teaches dictionary skills, and finally, we have to mention the *researcher*, whose work supports the development and ensures continuous quality (cf. Hartmann 2001).

Every one of the four roles must be covered during the training. In the theoretical part, the dictionary editing process is outlined. After learning about the micro-, macro- and megastructure of dictionaries, the topic of dictionary use is introduced, and it is worthwhile to accompany it with vocabulary and grammar exercises so that different language problems can be illustrated by answering them with a dictionary. This brings us to the area of dictionary didactics, where we may discuss methods of teaching dictionary use, illustrated by practical exercises. Finally, when students have a broader picture of the various fields of lexicography, they learn about possible methods and tools of dictionary research (cf., e.g., Fóris 2018; Lew 2016; Nied Curcio 2022; P. Márkus 2020b; P. Márkus and Pődör 2021).

2.2 The syllabus

In accordance with the aforementioned objectives, the syllabus of this course has been designed to present a balanced picture of lexicology and lexicography to students at an advanced level of English proficiency (focusing in particular on English monolingual and English–Hungarian/Hungarian–English bilingual learners' dictionaries). It has been developed to give teacher trainees an understanding of the field of lexicographic research as well as a variety of methodological tools and didactic materials for teaching dictionary use. The syllabus (Table 1) is divided into three thematic sections (Lexicology; Lexicography; Dictionary use and dictionary skills), each of which is further subdivided into several thematic units. Here are the complete thematic sections and their associated units:

Table 1: Syllabus of the course "Lexicology and Lexicography"

Section I. — Lexicology
<ul style="list-style-type: none"> — Unit 1: Introduction; discussion of the requirements for obtaining a grade; differences between lexicology and lexicography: basic concepts; lexicology defined; the notion of a/the word; word meaning. — Unit 2: Polysemy; homonymy; the structure of English vocabulary; sense relations (synonymy, hyponymy, meronymy, antonymy, etc.). — Unit 3: Inflection and derivation; word-formation processes (compounding, affixation, conversion, back-formation, blending, etc.); historical development of English vocabulary.
Section II. — Lexicography
<ul style="list-style-type: none"> — Unit 4: Lexicography defined: basic concepts; differences between a dictionary and an encyclopaedia; encyclopaedic dictionaries; encyclopaedic information in a dictionary; the classification/typology of dictionaries. — Unit 5: The historical development of lexicographic traditions; the microstructure, macrostructure, and megastructure of dictionaries I. — Unit 6: The microstructure, macrostructure, and megastructure of dictionaries II. — Unit 7: Corpus-based lexicography: definition of corpus; types of corpora; uses of corpora in language study. — Unit 8: Dictionary analysis.
Section III. — Dictionary use and dictionary skills
<ul style="list-style-type: none"> — Unit 9: Dictionary skills exercises, activities, games (for English monolingual dictionaries); keeping vocabulary books; creating card systems. — Unit 10: Dictionary skills exercises, activities, games (for English–Hungarian/Hungarian–English bilingual dictionaries). — Unit 11: Dictionary skills exercises, activities, games, and picture dictionaries. — Unit 12: Research on dictionary use.

2.3 Course requirements and grading policy

To successfully complete the course, students are required to complete two written assignments for grading. First, they are asked to choose a dictionary from a given list (the list includes dictionaries published by renowned inter-

national and Hungarian publishers — English monolingual and English–Hungarian/Hungarian–English bilingual dictionaries) and analyse it according to a set of guidelines prepared by Dóra Pődör (for an abridged version see Appendix 1). An important aspect of the guide was to cover every detail of the dictionary so that students could become acquainted with it thoroughly. This home assignment represents 50% of the final grade. Second, students must complete a project in which they create a workbook (worksheets) for the development of dictionary skills. This assignment represents 50% of the final grade. To receive a grade, students must score at least 51% on each of these assignments.

The primary aim of the dictionary analysis is to assist students in thoroughly exploring their chosen dictionary using the criteria provided (see Appendix 1). It is essential because even during their university years, most students use dictionaries only superficially and are unaware of the wealth of data they contain (see for example Müller-Spitzer et al. 2018; Nied Curcio 2022; P. Márkus et al. 2023). To assist students in better comprehending the structure of the dictionary and the richness of information it contains, a substantial number of questions are included in the guide. If they feel the necessity, students may employ the criteria at any point during their classroom practice. Teacher trainees need to be prepared to teach dictionary use, so the analysis is followed by the compilation of a dictionary workbook, which is usually attached to their teaching portfolio (which is required to be completed and submitted to graduate).

Before delving into more depth, it is crucial to understand and distinguish between two fundamental forms of dictionary use: active and passive. Teachers need to develop exercises to demonstrate and practise both types of dictionary use if we want to develop a conscious use of dictionaries in learners. In the seminars, methods and exercises for the different dictionary elements are demonstrated, and students are given the opportunity to prepare their own exercises based on the examples presented. The exercises cover each part of the entry, practising the identification of each type of data. The activities aim to demonstrate effective dictionary use and to develop dictionary competence.

From a methodological point of view, there are several key aspects to keep in mind regarding the tasks. Here are only a few examples because it is not possible to provide a complete picture within the scope of this study: recognition; observation; comparison; creativity; and cooperation. When designing the exercises, it is critical to emphasise the process of *recognition* by using exercises that focus on selecting the correct information (e.g., finding the correct meaning) or on the interpretation of the data in the dictionary (e.g., interpreting signs and abbreviations). Tasks describing the structure of the entry and interpreting the dictionary's instructions for use reinforce the process of *observation*. Reading the external texts of dictionaries can be a significant aid to effective use, especially if the dictionary type is new to the user or if the use of the dictionary is required for more complex tasks. Both are typical of the activities of foreign language students in higher education. That's why the importance of reading

the introduction to the dictionary should be stressed and the correct interpretation of abbreviations and phonetic symbols should be practised. The process of *comparison* is developed, for example, when students compare a dictionary entry from an electronic dictionary with a print dictionary (or a dictionary entry from a previous edition with the revised edition) as a comparison and analysis task. Tasks that require *creativity* and *cooperation* (when learners create similar tasks independently) will have huge motivational power in the language learning process (cf. Gonda 2009; P. Márkus and Pődör 2021).

2.4 General feedback and future aims

In line with the Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG 2015), our university has its own quality assurance system (Student Evaluation of Teacher Performance). Students rate the success of the course on a scale of 1 to 5 (where 5 means extremely or very satisfied; 1 means extremely or very dissatisfied). Completion of the questionnaire is not mandatory; the results provide feedback to the university on how well the course is meeting the requirements of the ESG. Based on the results of the previous four years, the average rating of the course was 4.8 out of 5.

When asked, students exhibit a willingness to learn more about lexicographic tools. They want to know which dictionaries are accessible to language learners, which are the most reliable, how they are built and structured, and how to recognise information that is reliable. Most students reported that the course was very useful, they enjoyed it, and that they were now more familiar with learner's dictionaries and lexicographic tools. The feedback from students suggests that they start to look at dictionaries with "different eyes".

Regarding course modifications in the future, we should place a greater emphasis on digital dictionaries as well as related resources (electronic translators, or the combination of translators, AI, and dictionaries) in the course material. Digital competence is one of the eight key competences for lifelong learning developed by the European Union — a core competence that facilitates the acquisition and development of other competences (such as languages) and is in fact considered one of the core competences for the 21st century (European Commission 2019). As the foundations of our modern culture are increasingly based on and powered by the digital world, info communication and digital tools are increasingly present in a wide range of contexts (such as language learning, language exams, dictionary use or dictionary writing), and digital tools are therefore an essential part of schools. Developing digital competence is essential, as knowing how to use a computer does not necessarily mean that students can quickly access and process the information they are looking for. Knowing the layout and basic properties of data structure is essential to accessing the right information. Students need to know the internal structure (whether it is a dictionary entry or a corpus) of the textual content that

they read in print or on the internet, otherwise the retrieval of information will take a lot of time or fail completely (cf. Tarp and Gouws 2020; M. Pintér 2019). The next section will outline the directions set by future challenges and share the materials that are in the preparatory phase.

3. New challenges: Specific skills for the teaching of electronic dictionary use

Research shows that electronic and especially web-based online dictionaries are becoming increasingly popular among language learners (Dringó-Horváth 2012; Nied Curcio 2015; Töpel 2015). The significant decline in user interest in print dictionaries as well as the high costs compared to electronic dictionaries has led to a continuous decrease in the production of print dictionaries while the online offer continues to expand (cf. Töpel 2015). Several studies of user research show a lack of knowledge of electronic dictionaries (cf. Nied Curcio 2015; P. Márkus et al. 2023), which indicates a great deal of uncertainty when selecting from the growing offer.

Since electronic dictionaries differ in many respects from print dictionaries, new examination criteria and new skills of use are needed to be established (for textual differences and terminological difficulties, see Müller-Spitzer 2014; details of new evaluation criteria and the differences in usage and successful skills can be found in Kemmer 2010; Dringó-Horváth 2012, 2021; Müller-Spitzer, Koplenig and Wolfer 2018; Müller-Spitzer et al. 2018). The relevant quality features presented below offer a practical orientation aid in the modern dictionary landscape (also presented in a checklist form in Appendix 3). Furthermore, concrete task suggestions show how specific features of digital dictionaries can be included in dictionary didactics.

3.1 Changing quality characteristics as drivers of new dictionary skills

In addition to the established, primarily content-based criteria that are transferable to electronic dictionaries, other, new quality characteristics should also be used that are only valid in the electronic learning environment. The reason for this lies essentially in the changed structure and functioning of this new form of publication: "Appropriately designed electronic dictionaries differ from print dictionaries not only in terms of media but above all in the variety of linguistic information offered or in the revolutionary way in which the linguistic data are presented. In this respect, unfortunately, one finds serious differences in quality in the landscape of electronic dictionaries" (Dringó-Horváth 2012: 35). Accordingly, it is extremely important to know appropriate features, including newly established ones, for selection and effective use in foreign language teaching.

3.1.1 Traditional quality characteristics

Many criteria, especially those related to content, can be applied — possibly in a slightly modified manner — to electronic dictionaries as well. The basic criterion is still the examination of the data set: the number of entries or the number of equivalents and the quality of information in the entries. The former can hardly be checked in the case of electronic dictionaries, so in this respect the user is entirely dependent on the data provided by the authors/publishers. In the presentation of entries, it is necessary to check to what extent the structural indicators make the information clear and unambiguous, or whether they are designed in a way that is suitable for the digital context (possibility of personalisation, use of colours, highlighting or other media such as images or sounds, cf. Figure 1).

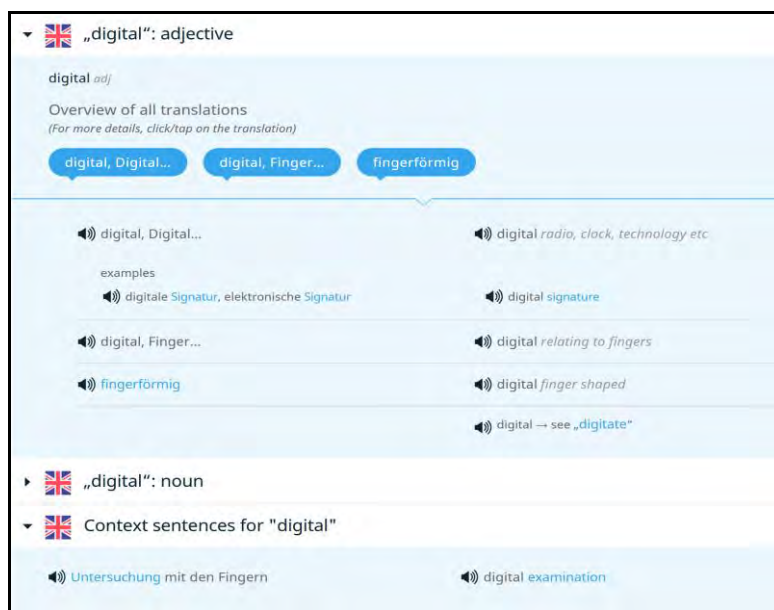


Figure 1: Use of colours, images, and sounds. Personalisation of the entry: the user can open the white segments by clicking on the arrows
Source: <https://de.langenscheidt.com/>

It is important to note, however, that research on dictionary use has shown that not all users value the above-listed features (cf. Müller-Spitzer and Koplenig 2014; Kosem et al. 2018).

Elements that make the text more compact should be used less since in the digital environment one does not necessarily require these summaries thanks

to the increased storage capacity and the flexible presentation possibilities of data. This increased capacity allows for the presence of as many external dictionary texts as possible that promote foreign language acquisition, such as notes on abbreviations, dictionary grammar or tables on the morphology and syntax of lexemes, which in turn makes it an important quality criterion. A new feature is that data and external dictionary texts can be adapted to the element being searched for (e.g., conjugation tables for each unit), and are generated automatically on a daily basis (Figure 2).



Figure 2: Clickable data and external dictionary texts, including a daily generated infographic on typical compounds
Source: <https://www.duden.de>

In addition, the quality of the content must be checked: To what extent can the editorial team be described as reputable, or is it indicated if the dictionary is based on the print version? A distinction must be made between products that can be described as electronic versions of print dictionaries and those that have been designed directly for the electronic environment (cf. Wiegand et al. 2010).

3.1.2 New quality characteristics

Among the new aspects, there are quality features that can make working with the dictionary much more convenient and faster for the user compared to print dictionaries but do not provide any advantage in terms of content. Such features are, for example, options that make using the product easier or even possible in the first place: installation and handling of the clear user interface as well as author/publisher support. This group also includes aspects that enable an adequate purchase/use decision, such as the corresponding user note (information on the most important characteristics of the dictionary with pictures and training exercises to practise adequate use) and the licence information with special consideration for the possible school use of the product.

Other new features bring advantages in terms of content compared to print dictionaries. These include the different search functions, the type and number

of which offer a decisive criterion in the evaluation. Here, we can also distinguish search functions that only contribute to facilitating and accelerating the retrieval of searched information for dictionary users from functions that also provide an advantage in terms of content (Table 2):

Table 2: Search functions with accelerating and content advantage

acceleration	
input based search	Type in the entry you are looking for in an input field.
incremental search	When typing in individual letters, the list of lemmas with the entered letter order is displayed search.
content profitable	
whole-text-search	Finding elements in other dictionary texts, in examples or definitions
search with logical operators	Use of logical connectors, such as * (any complement), OR (either/or search elements), AND (search multiple entities together).
search in the target language	Possibility to switch between the source and target language of the dictionary.
search without manual text entry	Search via the cursor (displays in speech bubbles) or via audio (sound input), possibly even via looks, facial expressions and gestures (cf. Töpel 2015).
find conjugated forms	Automatically reducing inflected forms of a word to its base form.
search in corpora	Directly on the WWW or in specifically prepared collections so that further (current, authentic) sample can be displayed.

A self-report questionnaire survey found that participants do not seem to take advantage of the various search techniques (P. Márkus et al. 2023), but research based on screen recording in conjunction with a thinking-aloud task showed, that users adapt their search strings to the particular tool they use, so in this respect they "are quite aware of the different functionalities of search engines, translation tools and dictionaries" (Müller-Spitzer et al. 2018: 310).

Other quality features that can be beneficial in terms of content or learning psychology are above all the following:

- Multimedia: The possibility of integrating different types of media (such as images, graphics, animation, and video).
- Flexibility of the technical side: Possibility of updating the management software or the online interface, changing and possibly saving the data presentation options according to user wishes (setting of colours or displayed elements), or the possibility of creating bookmarks, markings to directly reach the elements that are important for the user.

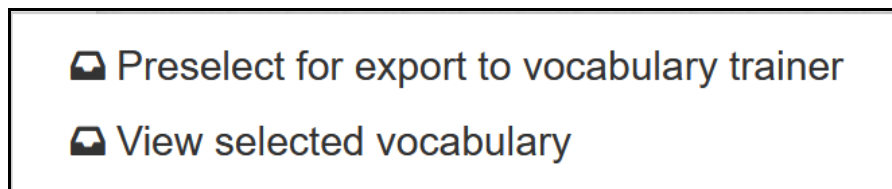


Figure 3: Marking vocabulary lists for training
Source: <https://de.pons.com>

- Flexibility of content: Possibility to expand the data by the publisher as well as by the user. This creates personalised dictionaries for individual as well as community use (in the case of publicly editable built-up dictionaries, with or without editorial review).
- Hypermediality: the meaningful linking of individual dictionary entries to each other as well as to external dictionary texts (abbreviations, grammar tables, etc.) and to other dictionaries or corpora in order to increase the number of hits and the number of (authentic, up to date) examples (see Figure 4).
- Functionality: In addition to the dictionary function, electronic dictionaries often deliver other functions that are usually helpful for foreign language learners, such as translator function, vocabulary trainer or pronunciation practice function (see Figure 5).
- Author/publisher support: In the case of digital, updatable products, the need for continuous mutual contact and information exchange increases from both sides to make technical and content updates as well as further user-induced development of the dictionary possible.

The screenshot displays the PONS Dictionary interface for the word 'collaboration'. At the top, there is a search bar with 'collaboration' entered, and language dropdowns for 'English' and 'German'. Below the search bar, there are two tabs: 'Dictionary' and 'Usage Examples', with the latter being selected. A blue header bar reads 'Examples from the PONS Dictionary (editorially verified)'. Below this, two example sentences are shown in a table-like format, each with a speaker icon on the left and a speaker icon on the right. The first example is 'to do sth in collaboration with sb' in English and 'etw in Zusammenarbeit mit jdm tun' in German. The second example is 'many benefits will flow from our collaboration' in English and 'unsere Zusammenarbeit wird viel Gutes mit sich bringen' in German. Below these examples is an 'AD' (Advertisement) placeholder. Another blue header bar reads 'Examples from the Internet (not verified by PONS Editors)'. Below this, four example sentences are shown in a two-column layout, each with a source URL 'www.giz.de' at the bottom. The first column contains two examples: '[...] The aim of the collaboration is the development of both national strategies and the capacity of health services in order to cater for the specific needs of men and thus provide them with better access to health care. [...]' and '[...] Financed by the Dutch Government, and in collaboration with EUPOL, GIZ has been running two-week training courses since 2011 to promote better cooperation between the police and the public prosecutor's office. [...]'. The second column contains two examples: '[...] Ziel der Zusammenarbeit ist die Erarbeitung von nationalen Strategien und die Qualifizierung von Gesundheitsdiensten, um die spezifischen Bedürfnisse von Männern zu berücksichtigen und ihnen somit einen besseren Zugang zur Gesundheitsversorgung zu verschaffen. [...]' and '[...] Finanziert durch die niederländische Regierung und in Zusammenarbeit mit EUPOL fördert die GIZ seit 2011 eine bessere Zusammenarbeit von Polizei und Staatsanwaltschaft durch zweiwöchige Trainingsmaßnahmen. [...]'].

Figure 4: External but not verified examples
 Source: <https://de.pons.com>

Quality features that provide an advantage in terms of content or learning psychology compared to print dictionaries should be valued more highly. It could be used for quick orientation of foreign language teachers and learners as well as for teaching purposes.

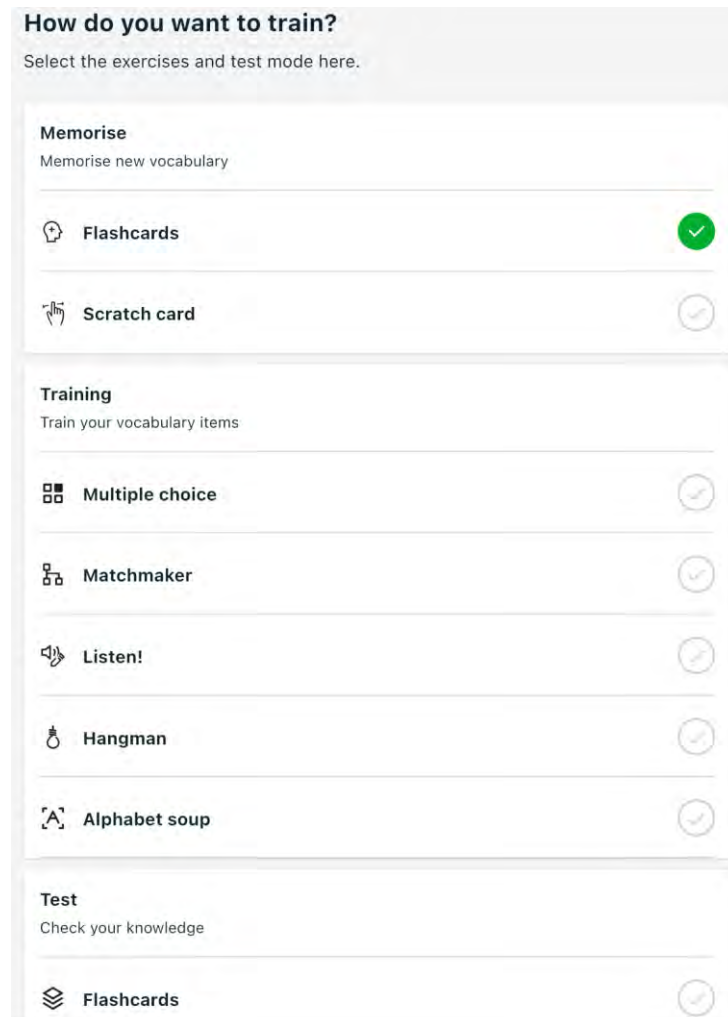


Figure 5: Tasks to learn and test vocabulary with pre-recorded vocabulary and/or with additional, selected word lists
Source: <https://de.pons.com>

3.2 Adapted dictionary didactics when teaching the use of electronic dictionaries

The new, enhanced quality features of dictionaries in the digital learning environment also have an impact on dictionary didactics, i.e., on the development of concepts and working methods for teaching effective dictionary use. Fol-

lowing a basic description by Zöfgen (2010), the main functions and tasks of dictionary didactics can be formulated as follows: Learners should

- know the different types of dictionaries and the data they contain;
- acquire knowledge about the structure, arrangement principles and text-typical features of dictionaries;
- be enabled to master the basic skills of searching and retrieving lexicographic data and the technique of looking up information;
- be encouraged to use dictionaries responsibly, which includes both the ability to make selections and to contribute to the development of dictionaries that are appropriate for the target group and are user-friendly.

In the following section, ideas as well as concrete task suggestions are given on how specific features of digital dictionaries can be included in this respect.

— **Knowledge about dictionary types**

When acquiring bibliographic knowledge, the problem arises that, on the one hand, some previously familiar terms lose their original meaning (pocket dictionary, unabridged dictionary) and, on the other hand, new phenomena have to be named, whereby the lack of a uniform terminology makes orientation difficult. From the perspective of foreign language teaching, the reliability of the data stock (see above) and completeness (dictionaries that can no longer be expanded, or advanced dictionaries that can be expanded by the editors and/or users, cf. Wiegand et al. 2010) are also important.

Some traditional type designations common for print dictionaries can be used well (monolingual, bilingual or multilingual dictionary, thematic dictionary, learner's dictionary, specialised dictionary, etc.), but a clear-cut distinction is becoming increasingly difficult as a result of the growing complexity of these products: dictionary portals or complex software with multiple, coupled functions can search several dictionaries simultaneously.

— **Knowledge about structure**

With the respective publication form of a dictionary — even if it is the same dictionary in printed or electronic form — the structure, the structural elements and the characteristic features inevitably change (cf. Dringó-Horváth 2012; Müller-Spitzer et al. 2018). Consequently, learners must be made aware that certain data in electronic dictionaries can also be found in new, unusual or even impossible forms in print dictionaries and they must discover how they work.

— Basic skills in searching

When thinking about making successful searches in the digital environment, one must also include the appropriate handling of the user interface. The recognition and adequate use of different search functions are also important; therefore, in the lessons, the focus should be on tasks for getting to know and practising individual search options. However, if elements can be retrieved quickly and effortlessly again and again, one (subconsciously) makes less effort to memorise them (cf. Rüschoff and Wolff 1999). Thus, lesson instruction should also include awareness raising activities that emphasise the conscious applying of found information to practice.

— Responsible use

Since electronic dictionaries are likely to contain a wide range of data sets, the ability to make selections is particularly important, especially in the case of collaboratively editable built-up dictionaries. At the same time, these make it possible for users to really contribute actively and effectively to the production of dictionaries that are user-friendly and appropriate for the target audience. The shared responsibility for this work should also be emphasised in the classroom. As a further contribution possibility, users can regularly inform publishers about deficiencies in dictionaries.

4. Conclusion

In this article, we aimed to describe in detail a university course in Lexicology and Lexicography. In addition to the theoretical and practical aspects, the description includes additional information on related research projects which have been conducted to support the design of the course. By providing the syllabus and a set of course requirements, we may be able to provide inspiration for other universities planning similar courses. The materials attached in the appendix may also provide ideas for resources for lexicographic and dictionary skills training. The article also looks at future challenges and possible course modifications, since we need to understand how the role of dictionaries is changing and be able to adapt to a constantly changing world. The incredible pace of technological progress is taking lexicography and dictionary didactics into new areas, changing dictionaries and thus dictionary use. At the same time, in our modern world, the role of dictionaries and lexicons as information carriers is becoming increasingly important. The information literacy skills developed through the use of dictionaries can be useful in many areas of the information society — for example, independent learning, digital literacy, problem-solving and information retrieval (DCF 2020; ECA 2021; European Commission 2019). Acquiring all these will be difficult if reference works, dictionaries and ency-

clopaedias are left out of the educational process (cf. DCF 2020). The teaching of dictionary use should be given a more prominent role in future public education to meet the challenges of the 21st century, hence teacher training at university level should prepare teacher trainees for this. Working and learning with digital dictionaries is — despite many similarities — marked by significant differences compared to working with print dictionaries. Hopefully, this article can help foreign language teachers and learners to become aware of these differences and to include them more and more in their dictionary use.

Acknowledgements

For constructive comments, the authors wish to thank the anonymous reviewers. Additionally, they extend their appreciation and thanks to Mr André du Plessis for his valuable suggestions and support.

References

- Atkins, B.T.S. and K. Varantola. 1998. Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 21-81. Tübingen: Max Niemeyer. <https://doi.org/10.1515/9783110929997.21>
- Bae, S. 2011. Teacher-Training in Dictionary Use: Voices from Korean Teachers of English. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX2011 Proceedings, LEXICOGRAPHY: Theoretical and Practical Perspectives*: 45-55. Kyoto: Asian Association for Lexicography.
- Béjoint, H. 1989. The Teaching of Dictionary Use: Present State and Future Tasks. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989. *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Vol. 1: 208-215. Berlin/New York: Walter de Gruyter.
- Bogaards, P. 1998. Scanning Long Entries in Learner's Dictionaries. Fontenelle, T., P. Hilgsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *EURALEX '98 Actes/Proceedings*: 555-563. Liège: English and Dutch Departments, University of Liège.
- Campoy-Cubillo, M.C. 2002. Dictionary Use and Dictionary Needs of ESP Students: An Experimental Approach. *International Journal of Lexicography* 15(3): 206-228.
- Campoy-Cubillo, M.C. 2015. Assessing Dictionary Skills. *Lexicography ASIALEX* 2: 119-141. <https://doi.org/10.1007/s40607-015-0019-2>
- Chan, A. 2012. Cantonese ESL Learners' Use of Grammatical Information in a Monolingual Dictionary for Determining the Correct Use of a Target Word. *International Journal of Lexicography* 25(1): 68-94.
- Chi, M.L.A. 1998. Teaching Dictionary Skills in the Classroom. Fontenelle, T., P. Hilgsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *EURALEX '98 Actes/Proceedings*: 565-577. Liège: English and Dutch Departments, University of Liège.

- Chi, M.L.A.** 2003. *An Empirical Study of the Efficacy of Integrating the Teaching of Dictionary Use into a Tertiary English Curriculum in Hong Kong. Vol. IV: Research Reports* (Ed. G. James). Hong Kong: Language Centre, Hong Kong University of Science and Technology.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press.
- DCF = Office of the Secretary-General of the European Schools.** 2020. *Digital Competence Framework for the European Schools*.
<https://www.eursec.eu/BasicTexts/2020-09-D-51-en-2.pdf>
- De Schryver, G.-M. and D. Joffe.** 2023. *The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks*. Conference paper. 20th CODH Seminar, Center for Open Data in the Humanities, Research Organization of Information and Systems, National Institute of Informatics, Tokyo, Japan, 27 February 2023.
<http://codh.rois.ac.jp/seminar/lexicography-chatgpt-20230227/>
- Dringó-Horváth, I.** 2012. Lernstrategien im Umgang mit digitalen Wörterbüchern [Learning Strategies in Dealing with Digital Dictionaries]. *Fremdsprache Deutsch* 46: 34-40.
- Dringó-Horváth, I.** 2017. Digitális szótárak — szótárdidaktika és szótárhasználati szokások [Digital Dictionaries — Dictionary Didactics and Dictionary Use Habits]. *Alkalmazott Nyelvtudomány* 17(5): 1-27.
- Dringó-Horváth, I.** 2021. Digitale Wörterbücher — Auswahlkriterien und angepasste Wörterbuchdidaktik. *Fremdsprache Deutsch* 64.
https://fremdsprachedeutschdigital.de/download/fd/FD_64_online_Dringo-Horvath.pdf
- Dringó-Horváth, I., K. P. Márkus and B. Fajt.** 2020. Szótárhasználati ismeretek vizsgálata német és angol szakot végzettek körében [Dictionary Skills in Teaching English and German as a Foreign Language in Hungary]. *Modern Nyelvtanítás* 26(4): 16-38.
- ECA = European Court of Auditors.** 2021. EU Actions to Address Low Digital Skills. Review no. 2.
https://www.eca.europa.eu/Lists/ECADocuments/RW21_02/RW_Digital_skills_EN.pdf
- Engelberg, S. and L. Lemnitzer.** 2009. *Lexikographie und Wörterbuchbenutzung*. 4th edition. Tübingen: Stauffenburg.
- European Commission, Directorate-General for Education, Youth, Sport and Culture.** 2019. Key Competences for Lifelong Learning. Publications Office.
<https://data.europa.eu/doi/10.2766/569540>
- Fóris, Á.** 2018. *Lexikológiai és lexikográfiai ismeretek magyar (mint idegen nyelv) tanároknak* [Lexicology and Lexicography for Teachers of Hungarian (as a Foreign Language)]. Budapest: Károli Gáspár University — L'Harmattan.
- Frankenberg-Garcia, A.** 2011. Beyond L1–L2 Equivalents: Where Do Users of English as a Foreign Language Turn for Help? *International Journal of Lexicography* 24(1): 97-123.
- Gonda, Zs.** 2009. A szótárhasználati kompetencia elsajátítása és fejlesztése [Competence Acquisition and Development in the Use of Dictionaries]. *Anyanyelv-pedagógia* 2.
<http://www.anyanyelv-pedagogia.hu/cikkek.php?id=160>
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Hartmann, R.R.K.** 2013. Resources. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 373-390. London: Bloomsbury.

- Jakubíček, M. and M. Rundell.** 2023. The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-editing Lexicography? Medved', M., M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček and S. Krek (Eds.). 2023. *Electronic Lexicography in the 21st Century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 Conference, Brno, 27–29 June 2023*: 518-533. Brno: Lexical Computing CZ s.r.o.
- Kosem, I., R. Lew, C. Müller-Spitzer, M. Ribeiro Silveira, S. Wolfer et al.** 2018. 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.
[doi:10.1093/ijl/ecz002](https://doi.org/10.1093/ijl/ecz002)
- Lew, R.** 2011. Studies in Dictionary Use: Recent Developments. *International Journal of Lexicography* 24(1): 1-4.
<https://doi.org/10.1093/ijl/ecq044>
- Lew, R.** 2016. Dictionaries for Learners of English. *Language Teaching* 49(2): 291-294.
- Lew, R.** 2023. *ChatGPT as a COBUILD Lexicographer*.
<https://doi.org/10.31219/osf.io/t9mbu>
- Lew, R. and G.-M. de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
<https://doi.org/10.1093/ijl/ecu011>
- Lew, R. and K. Galas.** 2008. Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-School-Level Polish Learners of English. Bernal, E. and J. DeCesaris. (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress Barcelona, 15–19 July 2008*: 1273-1285. Barcelona: Universitat Pompeu Fabra.
- Magay, T.** 2000. Teaching Lexicography. Heid, U., S. Evert, E. Lehmann and Ch. Rohrer (Eds.). 2000. *Proceedings of the 9th EURALEX International Congress EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 443-451. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Márkus, K. and É. Szöllősy.** 2006. Angolul tanuló középiskolásaink szótárhasználati szokásairól [The Dictionary Use Habits of Hungarian Secondary School Students Learning English as a Foreign Language]. Magay, T. (Ed.). 2006. *Szótárak és használói* [Dictionaries and their Users]. *Lexikográfiai füzetek* 2: 95-116. Budapest: Akadémiai Kiadó.
- Martynova, I.N. et al.** 2015. Modern Approaches to Teaching English Lexicography at the University of Freiburg, Germany. *Journal of Sustainable Development* 8(3): 220-226.
- M. Pintér, T.** 2019. Digitális kompetenciák a felsőoktatásban [Digital Competencies in Higher Education]. *Modern Nyelvoktatás* 25(1): 47-58.
- Müller-Spitzer, C.** 2014. Textual Structures in Electronic Dictionaries Compared with Printed Dictionaries: A Short General Survey. Gouws, R., U. Heid, W. Schweickard and H. Wiegand (Eds.). 2014. *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 367-381. Berlin/Boston: De Gruyter Mouton.
<https://doi.org/10.1515/9783110238136.367>
- Müller-Spitzer, C. and A. Koplenig.** 2014. Online Dictionaries: Expectations and Demands. Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*: 143-188. *Lexicographica Series Maior* 145. Berlin/Boston: Walter de Gruyter.
[doi:10.1515/9783110341287.143](https://doi.org/10.1515/9783110341287.143)

- Müller-Spitzer, C., A. Koplenig and S. Wolfer.** 2018. Dictionary Usage Research in the Internet Era. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 715-734. London: Routledge.
<https://doi.org/10.4324/9781315104942-45>
- Müller-Spitzer, C., M.J. Domínguez Vázquez, M. Nied Curcio, I.M. Silva Dias and S. Wolfer.** 2018. Correct Hypotheses and Careful Reading Are Essential: Results of an Observational Study on Learners Using Online Language Resources. *Lexikos* 28: 287-315.
- Nesi, H. and R. Haill.** 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.
<https://doi.org/10.1093/ijl/15.4.277>
- Nesi, H. and P. Meara.** 1994. Patterns of Misinterpretation in the Productive Use of EFL Dictionary Definitions. *System* 22(1): 1-15.
- Nied Curcio, M.** 2015. Wörterbuchbenutzung und Wortschatzerwerb. Werden im Zeitalter des Smartphones überhaupt noch Vokabeln gelernt? [Dictionary Use and Vocabulary Acquisition. Is Vocabulary Still Learned at All in the Age of the Smartphone?]. *Info DaF* 5: 445-468.
- Nied Curcio, M.** 2022. Dictionaries, Foreign Language Learners and Teachers. New Challenges in the Digital Era. Klosa-Kückelhaus, A., S. Engelberg, Ch. Möhrs and P. Storzjohann (Eds.). 2022. *Dictionaries and Society: Proceedings of the XX EURALEX International Congress*: 71-84. Mannheim: IDS-Verlag.
- Nkomo, D.** 2014. Teaching Lexicography at a South African University. *Per Linguam* 30(1): 55-70.
- P. Márkus, K.** 2020a. A szótárhasználat jelene és jövője a közoktatásban — a nyelvtanítást szabályozó dokumentumok és segédanyagok tükrében [The Present and Future of Dictionary Use in Public Education — In the Light of Documents Governing Language Teaching]. *Modern Nyelvtanítás* 26(1-2): 59-79.
- P. Márkus, K.** 2020b. Szótárhasználati munkafüzet — Angol tanulószótár [Dictionary Skills Workbook — English Learner's Dictionary]. Szeged: Maxim.
- P. Márkus, K.** 2023. *Teaching Dictionary Skills*. Budapest: L'Harmattan.
- P. Márkus, K. and D. Pődör.** 2021. Lexikográfia és szótárdidaktika a Károli Gáspár Református Egyetem Angol Nyelvészeti Tanszékén [Lexicography and Dictionary Didactics at the Department of English Linguistics — Károli Gáspár University of the Reformed Church in Hungary]. *Iskolakultúra* 31(5): 92-107.
- P. Márkus, K., B. Fajt and I. Dringó-Horváth.** 2023. Dictionary Skills in Teaching English and German as a Foreign Language in Hungary: A Questionnaire Study. *International Journal of Lexicography* 36(2): 173-194.
<https://doi.org/10.1093/ijl/ecad004>
- Préi, T.** 2020. Teaching Lexicography as a University Course: Theoretical, Practical and Critical Considerations. *Lexikos* 30(1): 293-320.
<https://doi.org/10.5788/30-1-1597>
- Read, A.W.** 2023. Dictionary. *Encyclopedia Britannica*.
<https://www.britannica.com/topic/dictionary>
- RFCD = Council of Europe.** 2018. *Reference Framework of Competences for Democratic Culture. Volume 2. Descriptors of Competences for Democratic Culture*. Council of Europe: Strasbourg.
- Rüschhoff, B. and D. Wolff.** 1999. *Fremdsprachenlernen in der Wissensgesellschaft: zum Einsatz der Neuen Technologien in Schule und Unterricht* [Foreign Language Learning in the Knowledge Society: On the Use of New Technologies in Schools and Classes]. Ismaning: Hueber.

- Sinclair, J.** 1984. Lexicography as an Academic Subject. Hartmann, R.R.K. (Ed.). 1984. *LEX'eter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 3–12. Tübingen: Max Niemeyer.
- Tarp, S. and R. Gouws.** 2020. Reference Skills or Human-centered Design: Towards a New Lexicographical Culture. *Lexikos* 30: 470–498.
- Töpel, A.** 2015. Das Wörterbuch ist tot — es lebe das Wörterbuch?! [The Dictionary is Dead — Long Live the Dictionary?!] *Info DaF* 5: 515–534.
- Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer and W. Wolski (Eds.).** 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung*. Band 1 [Dictionary of Lexicography and Dictionary Research. Volume 1]. Berlin: De Gruyter.
- Yamada, S.** 2010. EFL Dictionary Evolution: Innovations and Drawbacks. Kernerman, I. and P. Bogaards (Eds.). 2010. *English Learners' Dictionaries at the DSNA 2009*: 147–168. Tel Aviv: K Dictionaries.
- Zöfgen, E.** 2010. Wörterbuchdidaktik [Dictionary Didactics]. Königs, F.G. and W. Hallet. 2010. *Handbuch Fremdsprachendidaktik* [Handbook of Foreign Language Didactics]: 107–110. Seelze-Velber: Klett Kallmeyer.

Appendix 1

Guidelines for preparing your analysis and questions to be answered (an abridged version)

(The evaluation guide was compiled by Dóra Pődör.)

<i>Analyse the dictionary that has been assigned to you according to the following points (note that in most cases the information conveyed by the category in question as well as its place in the dictionary and its typographical form — presentation — have to be considered): Please place the dictionary into the general theoretical framework you have learnt about (e.g., whether general, descriptive, synchronic, etc.)</i>		
1. Front matter		
What kind of information is given in the Front matter?		
What sections does it contain?		
Is there a Preface/Foreword/Introduction?	✓	✗
Are the purpose, intended user group, approach, organisation, and scope of the dictionary stated?	✓	✗
Are there instructions for use (e.g., "Guide to the dictionary" or "How to use the dictionary")?	✓	✗
Is there any grammatical information given in the Front matter?	✓	✗
2. Entry		
What is the basic structure of an entry? (e.g., headword – pronunciation – inflections – part of speech – sense pattern with usage/register labels, cross-references, synonyms, constructions information, collocations, etc.; or any other structure.)		
3. Headwords		
What is the typographical form of the headword like?		
How are homonyms and conversion pairs distinguished? Do they appear in separate entries?		
How are variant spellings presented?		

Is there any information on word division?	✓	✗
Does the dictionary list as headwords irregular grammatical forms of the language?	✓	✗
4. Pronunciation		
Is it shown on the headword or separately?	✓	✗
What is its typographical form?		
What transcription system is used?		
If a system other than the IPA is used, describe it and comment on it.		
5. Inflection		
In what typographical form does it appear?		
What kind of information does it convey?		
How is the change in the headword indicated?		
How are inflected forms that are identical with the headword indicated?		
6. Part of speech		
What is conveyed by this type of information?		
What is its position and its typographical form?		
7. Constructions (syntactic valency)		
What kind of information is given on constructions?		
Are examples given?	✓	✗
What is the position and the typographical form of this kind of information?		

8. Collocations and idioms		
How are these presented?		
Do proverbs appear in the dictionary?	✓	✗
9. Definitions (in monolingual dictionaries)		
How are these presented?		
How extensive is the information given?		
Are there any guidewords? (Note that 'guideword' is used in at least two senses in lexicography.)	✓	✗
Does the dictionary use a 'defining vocabulary'?	✓	✗
10. Equivalents (in bilingual dictionaries)		
How are these presented?		
How extensive is the information given?		
Are there any guidewords? (Note that 'guideword' is used in at least two senses in lexicography.)		
What is the arrangement of meanings like?		
11. Subject field labels and register labels (usage labels)		
How many subject field labels and register labels (style, time, geography, speaker's attitude) are used?		
Are the abbreviations in the source language or the target language?	✓	✗
12. Other types of information (e.g., encyclopaedic information, illustrations, etymology, cross-references, guidewords, frequency):		
Are these present in the dictionary?	✓	✗
What kind of information do they convey?		
What is their format?		

13. Back matter (Appendices)		
What kind of information is contained in the back matter?		
How is it presented?		
14. Please state your personal opinion about the dictionary.		

Appendix 2

Guidelines for compiling a dictionary skills workbook

Dictionary skills are important because they transfer to the use of other reference books that students will use in their future studies. Unfortunately, many students are not confident about using dictionaries. The *Dictionary Skills Workbook* should be designed to familiarise students with the information included in the dictionary (you have chosen), as well as to help them quickly and effectively locate words, find meanings, etc.

- compile a workbook for teaching dictionary use with different types of exercises, tasks (focusing on different parts of the dictionary, e.g., pronunciation, grammar, meanings)
- an exercise should naturally include several questions: e.g., if you want to practise phonetic transcription, do it with about 10 words, not just one
- the reproducible exercises/worksheets should provide progressive instruction on topics such as alphabetizing, phonetic spellings, guide words, meanings, etc.
- the workbook will help students learn to use any dictionary successfully

Suggested framework

Topic: e.g., using dictionaries

Level: (use *The Common European Framework of Reference for Languages (CEFR)*)

Aims: e.g., to develop students' dictionary skills

Age group:

Time:

Material:

Sources:

Contents:

- I. Headword
- II. Pronunciation
- III. Part of speech
- IV. Inflection
- V. Equivalents
- VI. Grammar
- VII. Collocations and idioms
- VIII. Appendix

The fundamental rule for avoiding plagiarism is to always list your sources. Never use someone else's worksheet unless you explicitly state so in your own workbook.

Appendix 3

Checklist for the evaluation of electric dictionaries (Following Dringó-Horváth 2021)

NEW FEATURES WITHOUT ADDED CONTENT- OR LEARNING THEORY RELATED VALUE		
Installation and Handling		
— User-friendly installation/user-friendly subscription	✓	✗
— Changeable navigation language	✓	✗
— Clear navigation interface	✓	✗
User guide		
— User advice (apart from technical use also about dictionary contents)	✓	✗
— Training exercises for effective use	✓	✗
Licensing options and price		
— Price for one user		
— Further licensing options with discounts (depending on the user, for teaching purposes, etc.)	✓	✗
— Product protection e.g., by product access code	✓	✗
NEW FEATURES WITH ADDED VALUE IN TERMS OF CONTENT OR LEARNING PSYCHOLOGY		
Support by the authors/publishers		
— Product support on technical and content-related questions	✓	✗
— Information exchange possibilities between users and manufacturers or between users and each other	✓	✗

Search functions		
— Input-based search	✓	✗
— Incremental search (parallel display of the word list while typing)	✓	✗
— Search for multiple terms with logical connectors (+/AND, OR -/NOT etc.)	✓	✗
— Full text search	✓	✗
— Search for conjugated forms	✓	✗
— Search without manual text input (via sound, or cursor: pop-up function)	✓	✗
— Search in data collections (corpora)	✓	✗
— Search in several dictionaries simultaneously	✓	✗
— Other		
Flexibility, expansion possibilities		
(a) Flexibility of the technical side		
— Technical updates from the publisher	✓	✗
— Personalisation of data presentation	✓	✗
— Bookmarks/markers	✓	✗
(b) Flexibility of the content		
— Content updates by the publisher	✓	✗
— Extension of existing entries by the user	✓	✗
— Addition of new words/word lists by the user	✓	✗
— Other expansion options (e.g., own images, sound, and video)		

Multimedia		
— Text	✓	✗
— Image/graphic	✓	✗
— Moving image (animation, video)	✓	✗
— Sound	✓	✗
Hypermediality		
— Hyperlinks within the dictionary text (e.g., synonyms)	✓	✗
— Hyperlinks to external dictionary texts (e.g., conjugation tables)	✓	✗
Data output		
— Save data as a file or print it	✓	✗
Other connected functions		
— Translator	✓	✗
— Vocabulary trainer	✓	✗
— Pronunciation exercise	✓	✗
— Miscellaneous		

Nuwe raamstruktuurtypes in aanlyn woordeboeke*

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

Opsomming: In die leksikografiepraktyk lei die oorgang vanaf gedrukte na aanlyn woordeboeke tot talle veranderinge. Hierdie veranderinge moet in die leksikografieteorie beskryf word. Woordeboekstrukture is ook 'n teken van veranderinge, met ou strukture wat aangepas en nuwe strukture wat ontwikkel word. In hierdie artikel is die fokus op verskeie aspekte van een van die nuwe struktuurtypes in aanlyn woordeboeke, naamlik die skermkootstruktuur. Hierdie struktuur huisves onder meer individuele woordeboekartikels, artikeldede, beperkte artikels asook artikeldeeltrajekte. Naas die kernartikel laat die skermkootstruktuur ook uitgebreide artikels toe wat 'n nuwe tipe raamstruktuur vertoon, te wete 'n artikelraam. 'n Artikelraam kan die gevolg van horisontale uitbreiding na links en/of regs asook opwaartse en/of afwaartse vertikale uitbreiding wees. Geïntegreerde en niegeïntegreerde artikelbuite tekste bied huisvesting aan leksikografies relevante data wat nie in die kernartikel verskyn nie. Veral in die kommersiële leksikografie raak nieleksikografiese data in woordeboeke al hoe belangriker en die skermkootstruktuur moet so beplan word dat hierdie data gehuisves word sonder om data-oormatiging tot gevolg te hê. Dit kan geplaas word in 'n soekraam wat in die skermkootraamstruktuur voorsien word.

Sleutelwoorde: AANLYN WOORDEBOEK, ARTIKELRAAM, ARTIKELRAAMSTRUKTUUR, BUITETEKSE, DATAVERSPREIDING, RAAMSTRUKTUUR, SKERMKOOTRAAM, SKERMKOOTSTRUKTUUR, SOEKRAAM, SPYSKAARTTEKSBLOK

Abstract: New Types of Frame Structures in Online Dictionaries. In the lexicographic practice the transition from printed to online dictionaries resulted in numerous changes. These changes have to be accounted for in lexicographic theory. Dictionary structures also are a target of changes, with some old structures being adapted and new structures being developed. The focus in this article is on various aspects of one of the new structures in online dictionaries, namely the screenshot structure. This structure accommodates among other individual dictionary articles, partial articles, restricted articles as well as partial article stretches. Besides the core article the screenshot structure also allows extended articles that display a new type of frame structure, namely an article frame. An article frame can be the result of horizontal expansion to the left and/or right as well as upward and/or downward vertical expansion. Integrated and non-integrated arti-

* Hierdie artikel is 'n verwerking van 'n referaat gelewer by die 27ste Jaarlikse Internasionale AFRILEX-kongres, aangebied deur die Departement Suid-Afrikaanse Gebaretaal en Dowe-studies aan die Universiteit van die Vrystaat (UV) in Bloemfontein, Suid-Afrika, 26–29 Junie 2023.

cle outer texts accommodate lexicographically relevant data that are not included in the core article. Especially in commercial lexicography non-lexicographic data in dictionaries become increasingly important and the screenshot structure should be planned in such a way that these data can be accommodated without resulting in data-overload. Non-lexicographic data can be allocated to a search frame prevailing in the screenshot frame structure.

Keywords: ARTICLE FRAME, ARTICLE FRAME STRUCTURE, DATA DISTRIBUTION, FRAME STRUCTURE, MENU TEXTBLOCK, ONLINE DICTIONARY, OUTER TEXT, SCREENSHOT FRAME, SCREENSHOT STRUCTURE, SEARCH FRAME

1. Ter agtergrond

In die verhouding tussen teorie en praktyk in die leksikografie speel die teorie dikwels 'n navolgersrol om 'n beskrywing te gee van prosedures wat in die praktyk uitgevoer is. Die koms van die digitale era het daartoe gelei dat 'n verskeidenheid aanlyn woordeboeke tot stand gekom het en kenmerke vertoon wat nog nie in die teorie beskryf is nie. Die taak van die metaleksikograaf is nie net om vernuwend te dink en voorstelle vir aanpassings in die leksikografiepraktyk voor te stel nie, maar ook om krities na bestaande woordeboeke te kyk en dit te beskryf waarvoor daar nog nie in die teorie voorsiening gemaak is nie. Dit geld ook die benutting van woordeboekstrukture. Hierdie artikel poog om enkele aspekte van een van hierdie struktuurtypes oorsigtelik te bespreek.

Die oorgang vanaf gedrukte na aanlyn woordeboeke het tot talle veranderinge in die struktuur en aanbieding in aanlyn woordeboeke gelei. Hierdie veranderinge kom ook op die gebied van woordeboekstrukture voor. 'n Oorsig oor woordeboekstrukture wys dat sommige daarvan slegs in gedrukte woordeboeke voorkom, ander kom in sowel gedrukte as aanlyn woordeboeke voor, maar ondergaan wel bepaalde aanpassings wanneer hulle in aanlyn woordeboeke gebruik word. Daarbenewens is daar strukture wat slegs in aanlyn woordeboeke voorkom.

Woordeboekstrukture het onder meer ten doel om leksikografiese data, byvoorbeeld aanduiders in woordeboekartikels asook ander data wat nie in artikels optree nie, te huisves en om die gebruiker te help om die data te kan opspoor. Daar is 'n verskeidenheid plekke waar data deur die benutting van 'n uitgebreide dataverspreidingstruktuur geplaas kan word en wat as soekplekke dien waaraan inligting deur middel van 'n woordeboek se soekstruktuur onttrek kan word. Die verskillende soekplekke is onder meer:

- die *soeksone*, dit is 'n gleuf in 'n woordeboekartikel waar 'n bepaalde aanduiders of datatipe geplaas word;
- die *soekgebied*, dit is die individuele woordeboekartikel;

- die *soekveld*, dit is die woordeboek se sentrale teks wat die verskillende artikeltrajekte bevat;
- die *soekstreek*, dit is die volle woordeboek met sy sentrale teks plus die verskillende buitetekste;
- die *soekdomein*, dit is die portaal waarbinne aanlyn woordeboeke geplaas kan word; en
- die *soekuniversum*, dit is die omgewing buite die soekdomein, byvoorbeeld die internet of ander woordeboekportaal-eksterne bronne waartoe gebruikers via die woordeboek toegang het.

In gedrukte woordeboeke vertoon die soekstreek dikwels 'n raamstruktuur met buitetekste in die voor- en agtertekste-afdelings wat die sentrale teks aanvul (Kammerer en Wiegand 1998). 'n Onderskeid word gemaak tussen 'n volle raam, dit is waar sowel voor- as agtertekste voorkom, en 'n halwe raam, dit is waar die sentrale teks slegs deur óf voor- óf agtertekste aangevul word. Aanlyn woordeboeke kan wel buitetekste hê, maar hierdie buitetekste, oftewel buitekenmerke volgens Klossa en Gouws (2015) omdat dit nie net tekste is nie, maar ook onder meer video- en oudioknipsels mag wees, is nie lineêr posisioneel gebonde met betrekking tot die sentrale teks nie. Daarom vertoon aanlyn woordeboeke normaalweg nie 'n raamstruktuur nie.

Een van die struktuurtypes wat uitsluitlik in aanlyn woordeboeke voorkom, is die skermkootstruktuur, vergelyk Gouws (2014: 165). Afhangend van die aard van die uitleg van 'n aanlyn woordeboek kan 'n skermkootstruktuur soms meer as een artikel bevat, maar dikwels slegs 'n enkele artikel, 'n beperkte artikel of 'n artikeldeel. Naas die leksikografiese data kan 'n skermkootstruktuur ook nieleksikografiese data insluit (Gouws en Tarp 2017: 403). Alhoewel die skermkootstruktuur reeds as een van die strukture van aanlyn woordeboeke geïdentifiseer is, is daar in die metaleksikografiese gesprek nog min aandag aan die spesifieke aard en inhoud van hierdie struktuur gegee. Die verhouding tussen die skermkootstruktuur en die artikelstruktuur as komponent van 'n skermkootstruktuur is ook nog nie na behore bespreek nie. In die bespreking van die struktuur van woordeboekartikels is daar wel reeds verwys na 'n dinamiese veelvlakkige artikelstruktuur. Dit lei daartoe dat die gebruiker in die raadpleging van die bewerking van 'n bepaalde lemma toegang het tot verskillende skermkote wat telkens slegs enkele soeksones van die betrokke artikel vertoon, vergelyk Gouws (2014). Dít blyk uit die volgende drie skermkote uit *lexiko*. Figuur 1 wys 'n deel van die openingskermkoot van die lemma *Farbe* se bewerking. Figuur 2 bied die skermkoot wat volg nadat daar op die struktuurmerker "weiter" van figuur 1 se eerste semantiese subkommentaar geklik word, en figuur 3 is die skermkoot wat volg op 'n klik op figuur 2 se data-aanduider "Grammatik".

Farbe

Lesartenübergreifende Angaben

Orthografie
Normgerechte Schreibung: Farbe
Worttrennung: Far|be

Wortbildungsprodukte
(automatisch ermittelt) [weiter >](#)

Lesartenbezogene Angaben

Lesart 'optischer Eindruck' [weiter >](#)
Mit *Farbe* bezeichnet man die Eigenschaft eines Stoffes, Licht in einer bestimmten Wellenlänge zu reflektieren und dadurch von anderen Stoffen optisch unterscheidbar zu sein. So besteht zum Beispiel ein Regenbogen aus verschiedenen *Farben*.

Lesart 'Färbemittel' [weiter >](#)
Mit *Farbe* bezeichnet man ein Mittel, das bewirkt, dass Dinge (z. B. eine Wand, ein Gemälde) optisch, hinsichtlich ihrer Färbung, voneinander unterscheidbar sind.

Lesart 'Kartenreihe' [weiter >](#)
Mit *Farbe* bezeichnet man eine Reihe von Karten eines Kartenspiels, die durch das gleiche Symbol (z. B. Herz, Karo) gekennzeichnet sind.

Lesart 'Art und Weise eines Tons' [weiter >](#)
Mit *Farbe* bezeichnet man in der Musik die spezifische Art und Weise, wie ein Ton klingt.

Zum Zusammenhang der Lesarten
Die Lesart '*optischer Eindruck*' stellt das Abstraktum dar. Die Lesart '*Färbemittel*' ist hierzu eine Konkretisierung. Die Lesarten '*Kartenreihe*' und '*Art und Weise eines Tons*' sind Metonymien der Lesart '*optischer Eindruck*'.

Figuur 1: Uit *ellexiko*

Farbe

Lesart: '*optischer Eindruck*'

[zur Übersichtseite](#) [Lesarten im Überblick](#)

Bedeutungserläuterung **Kollokationen** **Konstruktionen** **Sinnverwandte Wörter** **Gebrauchsbesonderheiten** **Grammatik**

Erläuterung der Bedeutung / Funktion

Mit **Farbe** bezeichnet man die Eigenschaft eines Stoffes, Licht in einer bestimmten Wellenlänge zu reflektieren und dadurch von anderen Stoffen optisch unterscheidbar zu sein. So besteht zum Beispiel ein Regenbogen aus verschiedenen **Farben**.

[Hinweis anzeigen >](#) [Belege anzeigen >](#)

Sachinformationen

Weitere Informationen

"**Farbe**, *allgemein*: eine durch Licht bestimmter spektraler Beschaffenheit ausgelöste und durch das Auge vermittelte Sinnesempfindung [...]; die wahrgenommene Farbe ist keine physikalische Eigenschaft der Gegenstände, wird jedoch im Sprachgebrauch für die auslösende Lichtstrahlung selbst oder für stoffliche Farbmittel verwendet. An den bunten Farben nimmt man einen Farbton wahr (Gelb, Rot, Blau, Grün), der verschieden stark ausgeprägt sein kann (Sättigung), sowie eine Helligkeit. Fehlt der Farbe der Farbton, ist die Sättigung null, und es liegt eine unbunte Farbe vor (Schwarz, Weiß und ihre Mischungen, die Farben der Graureihe)." *Brockhaus in Text und Bild Edition (2002)*. Mannheim (CD-ROM).

Wortklasse: Eigenschaftsprädikator

Figuur 2: Uit *ellexiko*



Figuur 3: Uit *elexiko*

Hier beweeg die gebruiker in sy/haar soektog na bepaalde data van die een skermkoot na 'n volgende. Gouws (2018: 52 e.v.) dui daarop dat daar in hierdie tipe veelvlakke artikelstruktuur onderskei moet word tussen die omvattende artikel, dit is die volle artikel wat weliswaar nie as 'n geheel in 'n enkele skermkoot verskyn nie, en die beperkte artikels, dit is daardie dele van die omvattende artikel wat in elke afsonderlike skermkoot verskyn. Gouws (2014) wys aan die hand van die kriteria vir woordeboekartikels daarop dat hierdie beperkte artikels wel as artikels benoem kan word, onder meer omdat elkeen steeds die lemmateken bevat en 'n leksikografiese bewerking bied van die leksikale item wat deur die lemmateken verteenwoordig word. In so 'n veelvlakke artikel wys Gouws (2022) ook op die bestaan van 'n soektonnel, dit is waar 'n kommentaar of subkommentaar in 'n beperkte artikel deur opeenvolgende stappe uitgebrei word na daaropvolgende beperkte artikels met die soektog wat deur middel van die opeenvolgende datamerkers, die soektonnel, uitgevoer kan word. Die interne toegangstruktuur bepaal hierdie spesifieke soekroete.

Reeds uit die voorbeelde in figuur 1–3 is dit duidelik dat alle skermkote nie dieselfde uitleg en struktuur het nie. In hierdie artikel word enkele aspekte van die skermkootstruktuur asook die struktuur van artikels en beperkte artikels wat per skermkoot aangebied word, bespreek met die doel om 'n metaleksikografiese uiteensetting van enkele kenmerke van hierdie onbespreekte woordeboekstrukture te bied.

2. Skermkoot van 'n tradisionele artikelstruktuur

Anders as in *elexiko*, figuur 1–3, vertoon sommige aanlyn woordeboeke 'n tradi-

sionele artikelstruktuur, dit is 'n struktuur vergelykbaar met dié van 'n gedrukte woordeboek. Vergelyk figuur 4 uit die *Pharos Afrikaans–Engels/English–Afrikaans Woordeboek*.

werk¹

werk [werk] 1 work, labour, effort, exertion; 2 task, job; 3 work, deed, handiwork; 4 work, creation; 5 work, employment, job; 6 duty, function, task, business; 7 walk of life; 8 work, office, place of employment; 9 motion (*of the stomach, a machine, etc.*); 10 works (*of a watch etc.*)
aan die ~ wees **a** be working, be at work **b** be up and doing;
hard aan die ~ wees be hard at work;
aan die ~! go to it!
~ aanneem contract for work;
~ afskeep botch/scamp one's work;
agterstevoor/verkeerd te ~ gaan go the wrong way about it;
dit is alles sy/haar ~ it is all his/her doing;
die ~ het begin work has started;
baie ~ aan iets bestee spend much work/labour/effort on s.t.;
aan die ~ bly a keep on working **b** work away;
'n boel ~ (infm.) stacks of work;
by die ~ wees be at work;
by/in die ~ wees be on duty;
dis nie (sommers) elkeen se ~ nie (infm.) it takes some (or a lot of) doing;
aan die ~ gaan/spring start working, fall/go/set to work;
hoe om te ~ te gaan how to proceed (or go about it);
reg/verkeerd met iets te ~ gaan set about s.t. (in) the right/wrong way;
~ toe gaan go to work;
ek moet ~ gedaan kry I must put in some work;
iem. ~ gee give s.o. work, employ s.o.;
daar word ~ van gemaak s.t. is being done about it, it's in the works (*infm.*);

Figuur 4: Uit *Pharos*

Dié skermkootdeel van die artikel van die lemmateken *werk* bied 'n tradisionele artikelstruktuur. Hier is geen sprake van beperkte artikels of 'n veelvlak-kige artikelstruktuur nie. Die artikelstruktuur bevat ook slegs die kernartikel en geen ander aanvullende segmente of 'n soektonnel nie. Waar 'n aanlyn woordeboek hierdie soort artikelstruktuur handhaaf, is die fokus in die beplanning van die skermkootstruktuur primêr op die beskikbare vertoonruimte van die skerm van die betrokke apparaat — 'n rekenaar, tablet of selfoon — en die besetting daarvan deur die artikel-interne leksikografiese data. 'n Aspek van artikel-dinamika wat hier ter sprake kan wees, is waar die artikel te lank is om ten volle op die skerm van die betrokke apparaat sigbaar te wees en waar die gebruiker op die skerm moet afbeweeg om die verdere artikeldele te kan sien. Hier is nie sprake van 'n veelvlak-kige artikelstruktuur of van skermkote wat beperkte artikels bevat nie. So 'n artikel word steeds in 'n enkele skermkoot, weliswaar soms 'n verlengde skermkoot, aangebied met 'n eenvoudige skermkootstruktuur. Die inhoud van die skermkoot word hiërargies georden, dit verskyn nie in beperkte artikels nie, maar in een aaneenlopende artikel sonder enige bykomende data hetsy leksikografies hetsy nieleksikografies.

In hierdie artikel word 'n eenvoudige skermkootstruktuur wat slegs 'n tradisionele artikelstruktuur vertoon nie verder bespreek nie. Aandag sal wel gegee word aan 'n skermkootstruktuur waar die tradisionele artikelstruktuur vergesel word van bepaalde artikel-eksterne inskrywings.

3. Nuwe raamstruktuurtypes

3.1 'n Artikelraamstruktuur

3.1.1 Horisontale uitbreiding

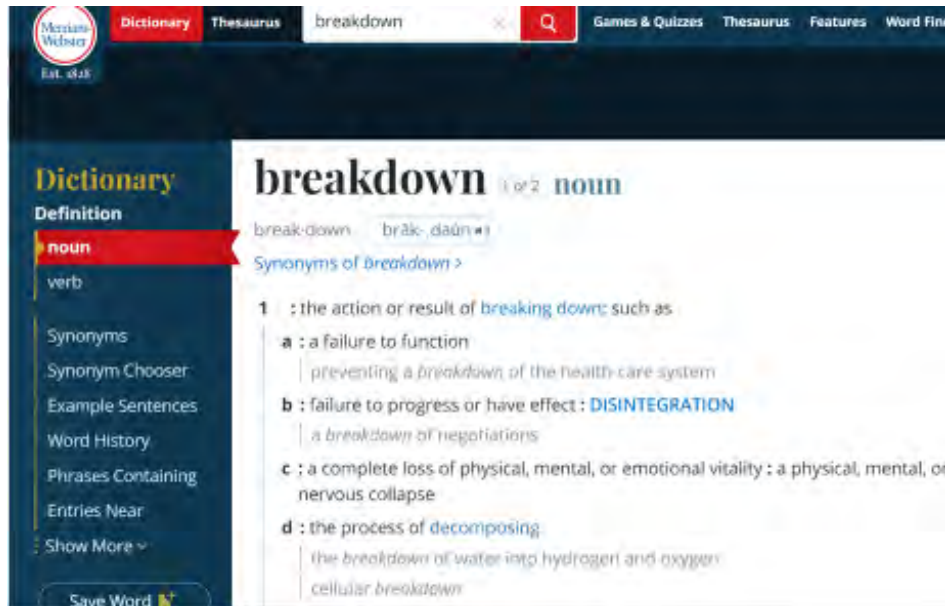
Die skermkoot in figuur 5 vertoon 'n deel van 'n beperkte artikel van die lemma *Kritik* in die *Duden Online*.



Figuur 5: Uit *Duden Online*

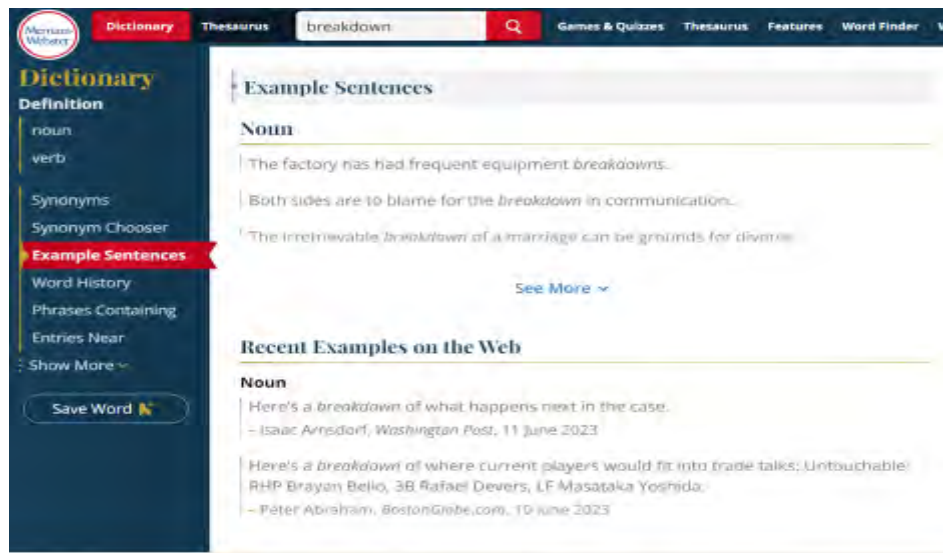
Die beperkte artikel bevat 'n kerngedeelte wat bestaan uit die lemmateken, aanduiders en struktuurmerkers wat aangebied word as deel van die leksikografiese bewerking van die woord *Kritik*. Daarnaas is daar op die linkerflank 'n teksblok wat 'n spyskaart is ter aanduiding van die datatipes wat in die artikel aangebied word. Die spyskaart is deel van die woordeboek se kitstoegangstruktuur en 'n klik op 'n datamerker lei die gebruiker na 'n beperkte artikel wat die tersaaklike soeksone en aanduiders bevat. Die spyskaartteksblok tree nie as deel van die kernartikel op nie. Dit bevat ook nie aanduiders nie, maar slegs datamerkers.

Figuur 6, 'n skermkootdeel uit die *Merriam-Webster* (MW) bied 'n beperkte artikel met 'n vergelykbare struktuur.



Figuur 6: Uit Merriam-Webster

'n Klik op die spyskaartteksblok se datamerker "Example sentences" lei tot die voorstelling in figuur 7.



Figuur 7: Uit Merriam-Webster

In figure 5–7 is dit duidelik dat die inhoud van die spyskaartteksblok nie verander wanneer daar van een beperkte artikel na 'n ander beweeg word nie. As spyskaart is dit ter sake vir elke artikel met 'n verpligte mikrostruktuur — vergelyk Hausmann en Wiegand (1989: 346). In figuur 6 en 7 word die datamerkers wat die soeksone(s) aandui wat die fokus van die beperkte artikel is, wel in die spyskaartteksblok gemerk. Die spyskaartteksblok staan telkens links van die kernartikel. Vanuit die perspektief van die artikelstruktuur met elke artikel wat uit minstens twee basiese kommentare, dikwels 'n vormkommentaar en 'n semantiese kommentaar, bestaan (Hausmann en Wiegand 1989: 354), lyk die spyskaartteksblok met 'n eerste oogopslag na 'n pre-kommentaar. 'n Pre-kommentaar is 'n onmiddellike stuk van 'n woordeboekartikel met 'n linksuitgebreide mikrostruktuur wat onmiddellik voor die vormkommentaar optree, die linkerrandstruktuur van die woordeboekartikel vorm en meestal een of meer aanduidersimbole bevat (Wiegand 1989: 474; Wiegand et al. 2020: 581). Die inskrywings in 'n pre-kommentaar is, gewoonlik deur 'n prosedure van lemmatiese adressering, na regs geadresseer. In figuur 8, 'n artikeldeeltrajek uit die *Kenkyusha's Lighthouse*, lei 'n linksuitgebreide mikrostruktuur tot die plasing van gebruiksfrekwenasierkers, die asteriske, in 'n prekommentaar waar hulle lemmaties na regs geadresseer is. In die artikel van die lemma *Europe* dui die drie asteriske daarop dat die woord *Europe* 'n hoër gebruiksfrekwenisie het as die woord *European* met slegs twee asteriske in die prekommentaar terwyl die afwesigheid van hierdie aanduidersimbole in die artikel van die lemma *European Communities* daarop dui dat die leksikale item wat deur hierdie lemma voorgestel word 'n nog laer gebruiksfrekwenisie as die voorafgaande twee woorde het.



Figuur 8: Uit *Kenkyusha's Lighthouse*

Die spyskaartteksblok in figure 5–7 bevat nie aanduidersimbole as inskrywings nie, maar wel datamerkers. Hierdie teksblokke is nie prekommentare nie, maar tree telkens as voorste rand van 'n uitgebreide artikel op. Die kernartikel word aangevul deur 'n artikelbuiteteks, die spyskaartteksblok, as voorste rand. Die optrede van die buiteteks het 'n artikelraamstruktuur tot gevolg. 'n Artikelraamstruktuur impliseer die bestaan van 'n artikelraam, hetsy 'n volle hetsy 'n halwe raam. In die voorafgaande voorbeelde bevat die raamstruktuur telkens slegs 'n buiteteks in die voortekste-afdeling en vertoon die artikel 'n halwe raam.

Aangesien die datamerkers in die spyskaartteksblok van die genoemde voorafgaande artikels die gebruiker help om by spesifieke soeksones in die omvattende artikel uit te kom, kan hierdie spyskaartblok telkens as 'n geïntegreerde buiteteks van die artikel beskou word.

'n Artikelraamstruktuur bevat nie noodwendig net geïntegreerde buitetekste nie. Vergelyk figuur 9 uit *lexiko* waar die artikel 'n volle raam vertoon met sowel voor- as agtertekste. Hierdie buitetekste is egter nie in die kernartikel, d.w.s. in die bewerking van die woord wat deur die lemmateken voorgestel word, geïntegreer nie. Dit ondersteun nie die gebruiker om spesifieke inligting aan die data in die woordeboekartikel te onttrek nie.

The screenshot displays the OVID lexiko web interface. At the top, there is a search bar with the text 'Suchen' and 'Erweiterte Suchen'. Below the search bar, a list of filters is visible, including 'Auschwitz', 'ausdrücklich', 'Auskunft', 'Ausland', 'Ausländer', 'ausländisch', 'Ausnahme', 'Ausschuss', 'außen', 'Außenminister', 'äußern', 'äußerst', 'äußerst-', 'Aussicht', 'Ausstellung', 'Australien', 'Auswirkung', 'Auto', and 'Autobahn'. The main content area shows the entry for 'Auto' with the following details:

- Orthografie:** Normgerechte Schreibung: Auto; Worttrennung: Au|to
- Wortbildung:** Wortbildungsart/-typ: unisegmentales Kurzwort (Typ: Kopfwort); Gekürztes Wort: Automobil (Nomen)
- Herkunft und Wandel:** Etymologische Angaben: anzeigen >; Wandel 1700 bis 1945: -; Wandel seit 1945: -
- Wortbildungsprodukte:** (automatisch ermittelt) weiter >

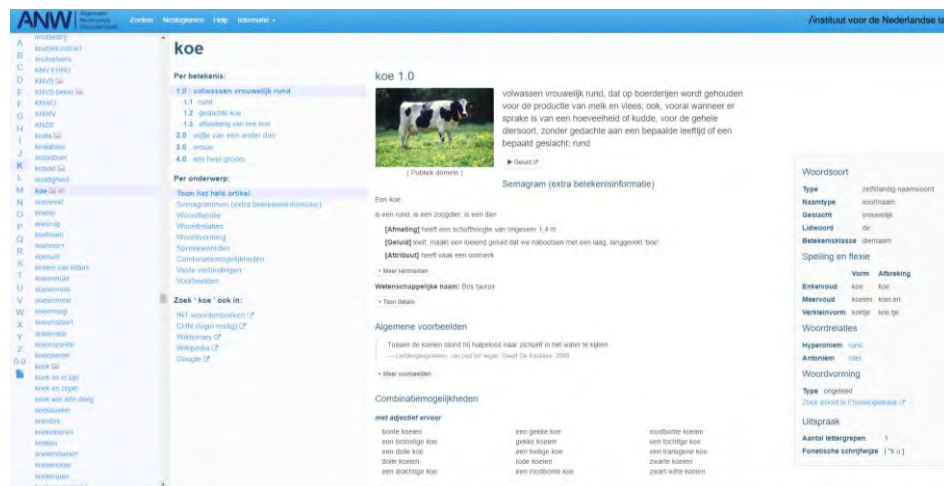
Below the main entry, there is a section titled 'Lesartenbezogene Angaben' with a link to 'Lesart "Fortbewegungsmittel" weiter >'. The text states: 'Mit Auto bezeichnet man ein Fortbewegungsmittel, und zwar ein durch Motor angetriebenes Straßenfahrzeug auf meist vier Rädern zum Transport von Personen.'

On the right side of the interface, there is a sidebar with a list of links: 'lexiko', 'OWID', 'lexiko', 'Startseite', 'Wortartikel', 'Projekt', 'Benutzungshinweise', 'Glossar', 'Erweiterte Suche', 'Feste Wortverbindungen', 'Sprichwörterbuch', 'Kommunikationsverben', 'Verlaufsformen', 'Neologismenwörterbuch', 'Schulddiskurs 1945–55', 'Protestdiskurs 1967/68', 'OBELEX meta', 'OBELEX dict', and 'Korpussuche'.

Figuur 9: Uit *lexiko*

Hierdie artikel is én na links én na regs uitgebrei. Die buitetekste is wel leksikografies relevant vir die artikel. Hulle bied kontekstualiserende hulp, met die voorteks wat die lemma binne die tersaaklike deeltrajek plaas en die agterteks wat die artikel binne 'n soekreeks, 'n spesifieke woordeboek, in die woordeboekportaal, die soekdomein, plaas. Dit gaan hier om niegeïntegreerde artikelbuitetekste.

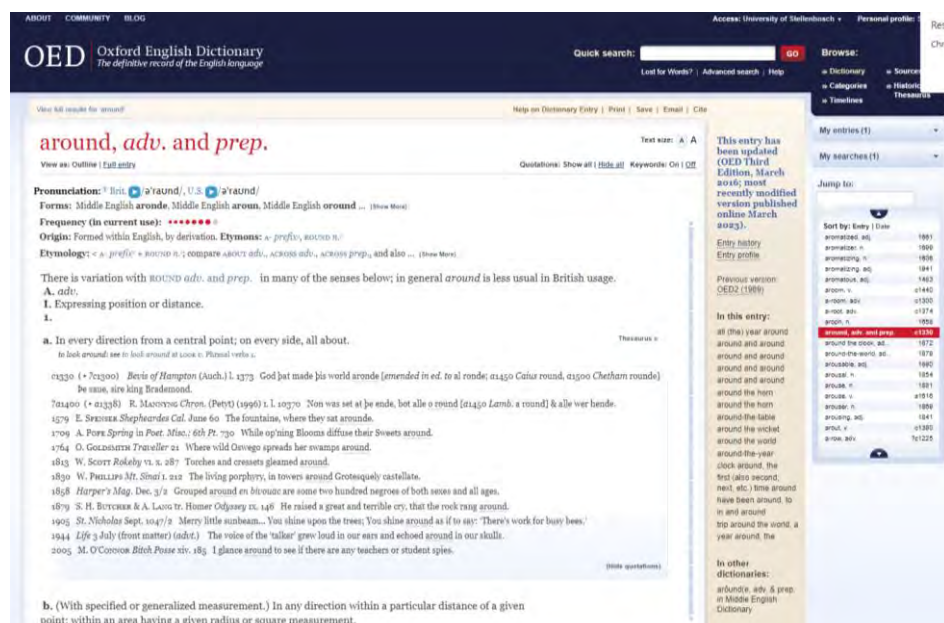
Die blokvorm waarin die buitetekste in die voorafgaande voorbeelde aangebied is, moet nie tot die gevolgtrekking lei dat teksblokke wat op die rand van 'n artikel geplaas is noodwendig buitetekste is nie. Vergelyk in hierdie verband figuur 10 uit die *Algemeen Nederlands Woordenboek* (ANW) waar die skerm-skoot 'n deel van die artikel van die lemma *koe* weergee.



Figuur 10: Uit ANW

Hier is wel 'n halwe raam met die voorteksafdeling wat twee buitetekste, naamlik 'n vertikale alfabetbalk asook 'n deeltrajek bevat. Aan die regterflank van die artikel is 'n teksblok wat met 'n eerste oogopslag soos 'n agterteks lyk. Uit figuur 11, 'n nader weergawe van hierdie teksblok, is dit duidelik dat dit nie 'n buitetekste is nie, maar 'n artikel-interne teksblok wat as deel van die kern-artikel 'n aantal soeksones bevat.

oor die herbewerking van die spesifieke artikel asook 'n lysing van enkele meerwoordige leksikale items en kollokasies wat in die betrokke omvattende artikel behandel word sowel as 'n verwysing na 'n woordeboek-eksterne bron waar meer inligting oor die woord wat deur die betrokke lemma verteenwoordig word, gevind kan word. 'n Klik op 'n inskrywing in hierdie lys neem die gebruiker na die betrokke soeksone in die woordeboekartikel. Tweedens is daar ook 'n niegeïntegreerde buitetekst, naamlik die deeltrajek waarbinne die lemma *around* voorkom.



Figuur 12: Uit OED

Hierdie buitetekste vorm gesamentlik die agterste raam van die artikel.

3.1.2 Vertikale uitbreiding

Die artikelraamstruktuur maak voorsiening vir 'n halwe raam met óf voor- óf agtertekste of vir 'n volle raam wat een of meer buitetekste in sowel die voor- as agtertekste-afdelings bevat. In figuur 13, 'n skermkootdeel van die artikel van die lemma *Fluss* in *dict.cc* se tweetalige woordeboek Duits–Nederlands, word die bewerking in die kernartikel aangevul deur 'n niegeïntegreerde buitetekst in die agtersteraamposisie waarin bykomende data verskaf word, naamlik 'n lysie taalpare, met Duits as eerste lid, wat ook in die *dict.cc*-woordeboekportaal verskyn. In hierdie lysie is die taalpaar Duits–Nederlands in vet gedruk

omdat dit aandui uit watter woordeboek die betrokke artikel kom. Hierdie buitekste is die gevolg van horisontale artikeluitbreiding na regs.

Die horisontale artikeluitbreiding om buitekste as voor- of agtertekste in te sluit, kan ook aangevul word deur 'n vertikale uitbreiding wat tot gevolg het dat 'n kernartikel deur vertikale raamtekste begrens word, naamlik opwaarts en/of afwaarts uitgebreide raamtekste.



Figuur 13: Uit *dict.cc*

Die woordeboek *dict.cc* het 'n standaardartikeluitleg wat deurgaans gebruik word en waarvolgens die skermkootstruktuur aangepas word. Hier bevat die verstekskermkoot 'n kernartikel met 'n horisontale uitbreiding na regs (die lys taalkombinasies) en 'n opwaartse vertikale uitbreiding wat meer as een raamtekste insluit. Die boonste raamtekste in figuur 13 bevat 'n alfabetbalk en regs van die balk die taal wat in die betrokke artikel met die Duitse woord gekombineer word, in figuur 13 is dit "Niederländisch" (=Nederlands). Onder die gleuf met die alfabetbalk is 'n gleuf met 'n teks wat die taalpaar asook die lemmateken aandui en heel regs in hierdie gleuf is daar 'n datamerker, hier "Translation 1-8 of 8", wat die gebruiker na verskillende beperkte artikels in die bewerking van die betrokke lemmateken kan lei. Hierdie buitekste in die boonste raam is leksikografies relevant. Die boonste teks is 'n niegeïntegreerde raamtekste en die onderste een is geïntegreerd in die leksikografiese aanbod van die woordeboekartikel.

Die raamtekste in figuur 5-7, 9-10 en 12-13, hetsy geïntegreerd hetsy niegeïntegreerd, is telkens relevant vir die betrokke artikel of beperkte artikel. Daarom staan hulle as primêre artikelraamtekste bekend.

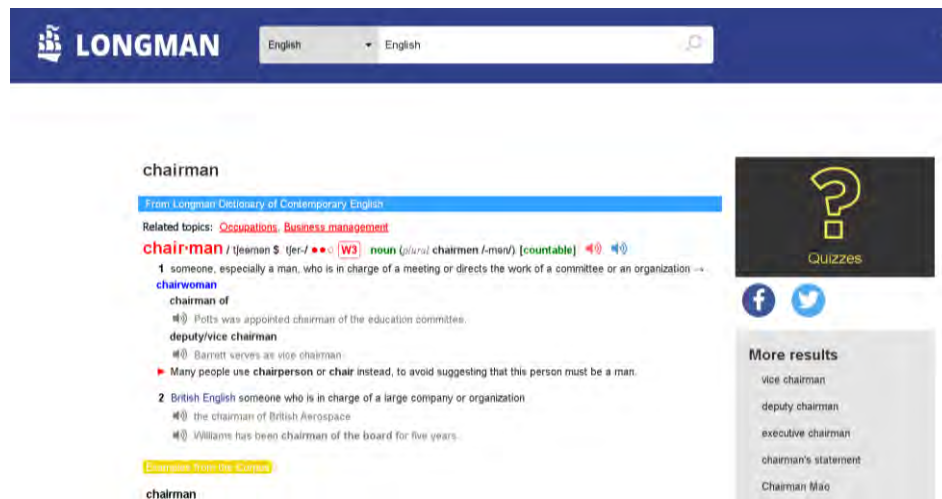
3.2 Verdere raamstrukture in die skermkootstruktuur

3.2.1 Sekondêre raam

Die skermkootstruktuur is 'n ordeningstruktuur. Dit vertoon die produk van die ordening en onderlinge posisionele verhouding tussen segmente in skermkote van woordeboekartikels. Dit kan onder meer die openingskerm, 'n woor-

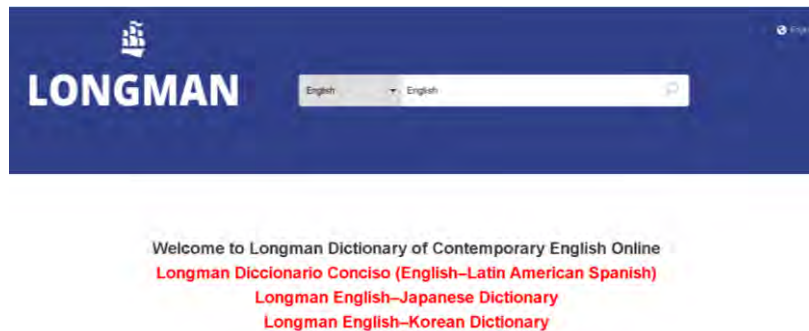
deboek- of artikelspyskaart, 'n alfabetbalk, 'n artikeltrajek of -deeltrajek of 'n artikel of beperkte artikel bevat.

Die dinamiese aard van aanlyn woordeboeke dwing leksikograwe om die skermkootstruktuur en die leksikografiese data wat daarin vertoon word vroegtydig te beplan. Die skermkootstruktuur vertoon al die verskillende segmente wat in die skermkoot verskyn en horisontaal en vertikaal georden word. Die kern van die skermkootstruktuur is die woordeboekartikel of artikelreeks saam met die buitetekste van die onderskeie artikels of beperkte artikels wat as primêre raamtekste optree. In die beplanning van die skermkootstruktuur moet leksikograwe besluit oor die opname en plasing van ander segmente in die skermkoot wat nie deel van die aanbod van die artikel is nie. Die verstekskermkoot in aanlyn woordeboeke vertoon nie noodwendig net die uitgebreide artikel, d.w.s. die kernartikel plus moontlike artikelraamtekste, nie. Vergelyk figuur 14, 'n skermkoot van 'n deel van die artikel van die lemma *chairman* uit *Longman Dictionary of Contemporary English* (LDOCE).



Figuur 14: Uit *Longman Dictionary of Contemporary English*

Naas die kernartikel met die niegeïntegreerde horisontaal na regs uitgebreide raamtekste is daar ook 'n teks bo die lemmateken wat die woordeboektitel bevat asook 'n aanduiding dat dit om die eentalige Engelse woordeboek (in hierdie woordeboekportaal) gaan. Daar is ook 'n blokkie waar die woord waarna gesoek word, ingetik kan word. Hierdie vertikaal opwaarts uitgebreide teks is nie 'n buitetekste van die artikel nie, maar die titelteks van elke skermkoot in hierdie woordeboek, soos ook te sien in figuur 15, wat 'n deel van die opening-skermkoot van hierdie woordeboek weergee, maar geen woordeboekartikel bevat nie.



Figuur 15: Uit *Longman Dictionary of Contemporary English*

Figuur 16 uit die *American Heritage Dictionary* (AHD) het ook 'n teks wat die woordeboek se titel bevat en wat in elke skermkoot verskyn. Daarbenewens is daar in hierdie skermkoot ook horisontaal geordende buitetekste wat eweneens in elke skermkoot verskyn. Hierdie tekste dra nie by tot die leksikografiese bewerking of kontekstualisering van die lemmateken of 'n aanduiding in die woordeboekartikel nie.



Figuur 16: Uit *American Heritage Dictionary*

Vanweë hulle posisie in die skermkoot is dit wel raamtekste wat ook die artikel of selfs die uitgebreide artikel kan omraam, maar hulle is nie deel van die uitgebreide artikel nie. Die inhoud van hierdie tekste is wel leksikografies van aard aangesien dit inligting oor sekere aspekte van die woordeboek verstrek. Dit is egter nie deel van die woordeboekartikel nie. Figuur 17 wys dat hierdie raamtekste ook verskyn in die openingskermskoot van die woordeboek wat geen woordeboekartikel bevat nie — waar hulle dus nie deel van 'n artikelraam is nie.



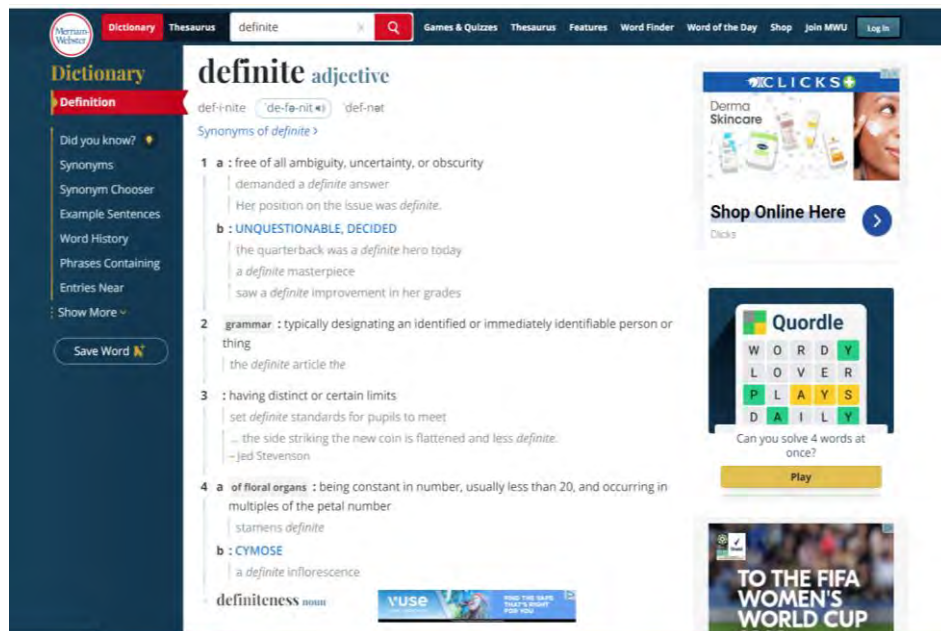
Figuur 17: Uit *American Heritage Dictionary*

Teenoor die primêre raamtekste wat die kernartikel omraam en die uitgebreide artikel vorm, vorm die buitetekste in figuur 15–17 'n sekondêre raam. Hierdie raamtekste is wel leksikografies van aard — die titel van 'n woordeboek en hulp met die gebruik daarvan — maar vorm nie deel van 'n uitgebreide artikel nie. In die beplanning van die skermkootstruktuur moet leksikograwe wel voorsiening maak vir 'n soekplek waar sulke tekste geplaas kan word. Hierdie data tree nóg in 'n soeksone nóg in 'n soekgebied op. Hulle beset 'n nuwe soekplek, te wete 'n soekraam. 'n Soekraam is die posisie in 'n skermkootstruktuur waarin raamtekste geplaas kan word wat leksikografies ter sake is, maar nie deel van 'n uitgebreide artikel is nie. Sulke tekste tree nie noodwendig net saam met 'n uitgebreide artikel op nie, maar kan ook saam met 'n kernartikel optree, soos blyk uit figuur 16 waar die artikel geen buitetekste het nie.

'n Skermkoot wat 'n soekraam met sy sekondêre raamtekste bevat, bied ruimte vir die artikel plus bykomende segmente wat vir die spesifieke woordeboek deur die leksikograaf en/of uitgewer verlang word. 'n Vergelykbare skermkootstruktuur is ook te sien in die voorafgaande figure 5 (Duden), 6 (MW), 9 (*lexiko*), 10 (ANW) en 12 (OED).

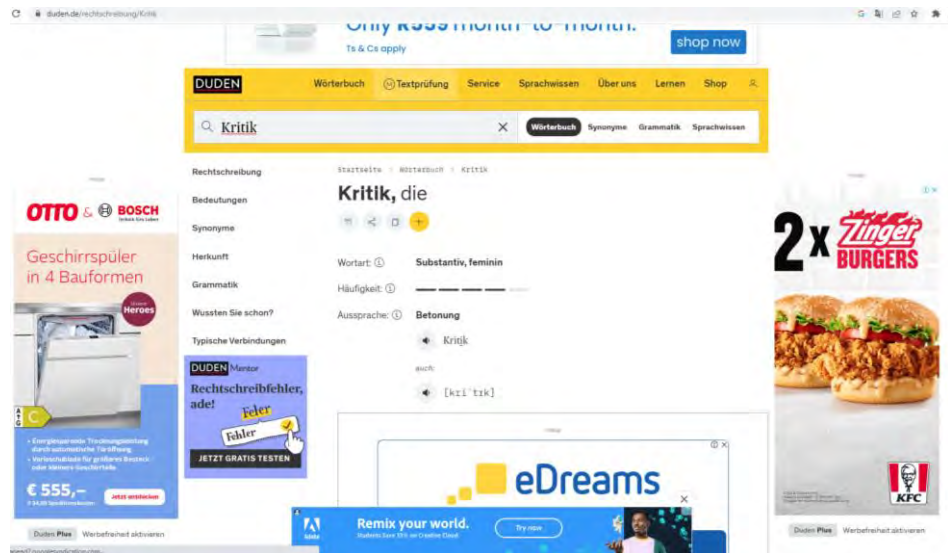
3.2.2 Skermkootraamstruktuur

Die artikelraamstruktuur met sy primêre en sekondêre raam is nie die enigste nuwe tipe raamstruktuur in aanlyn woordeboeke nie. Die skermkootstruktuur kan ook uitgebrei word om 'n eiesoortige raamstruktuur te vertoon, maar wat nie ter bevordering van leksikografiese inligtingsonttrekking benut word nie. Figuur 18, 'n skermkoot met 'n deel van die artikel van die lemma *definite* uit die *Merriam Webster*, vertoon 'n uitgebreide artikel wat uit die kernartikel en 'n voorste raam, die spyskaartteksteblok, bestaan. Daarbenewens bevat die skermkoot ook 'n titelteks, naamlik die gleuf bo die uitgebreide artikel waar die titel van die woordeboek verskyn, keuses van bronne waartoe die gebruiker toegang het asook die soekkassie waar 'n woord waarna gesoek word, ingetik kan word. Die uitgebreide artikel asook die titelteks is leksikografies relevante segmente van die skermkoot en vorm 'n skermkoot wat die kernartikel asook primêre en sekondêre raamtekste bevat.



Figuur 18: Uit Merriam Webster

Die skermkootstruktuur van aanlyn woordeboeke moet nie net voorsiening maak vir die plasing van leksikografiese data nie, maar moet ook 'n soekplek skep waar nieleksikografiese data geplaas kan word. Om nie die aandag van die leksikografiese data, die hoofokus van skermkote van aanlyn woordeboeke, weg te lei nie, is dit waarskynlik die beste dat die soekplek vir sulke nieleksikografiese data in die skermkootraam verskyn. Die skermkootstruktuur se verstekweergawe met 'n kernartikel, uitgebreide artikel en daarbenevens soms ook sekondêre raamtekste kan dus uitgebrei word deur die toevoeging van 'n ander tipe raamkomponent. In figuur 18 word die verstekskermkoot horisontaal en vertikaal uitgebrei deur die toevoeging van raamtekste. Naas die stukkie woordvermaak as middelste teks in die agterste raam, is daar in dié raam twee advertensies. Die onderkant van die skermkoot vertoon ook 'n advertensie as raamteks. In figuur 19 uit *Duden Online* word die leksikografies relevante data aangevul deur 'n volle horisontale sowel as 'n volle vertikale raam beset deur nieleksikografiese data. Hierdie weergawe van *Duden Online* wat gratis beskikbaar is, bevat talle advertensies en die skermkootstruktuur van die woordeboek is so beplan dat daar in die skermkootraam plek is vir hierdie advertensies. 'n Ander tipe dinamika is hier ter sprake deurdat die advertensies nie staties is nie, maar selfs tydens 'n enkele raadpleging van die betrokke artikel verander.

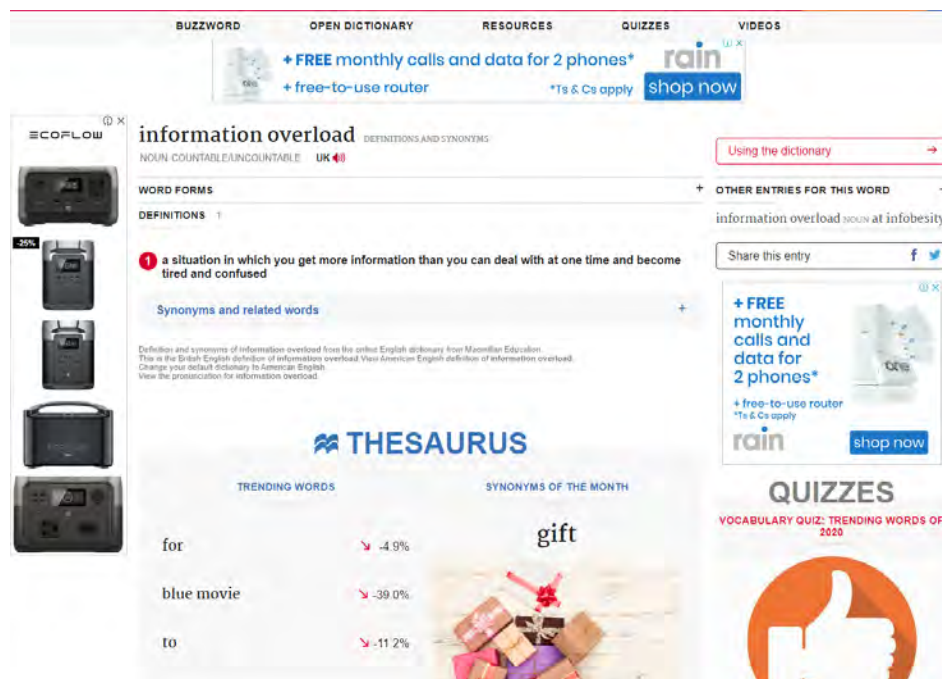


Figuur 19: Uit *Duden Online*

Aangesien dit nie leksikografiese data is wat in hierdie raam vertoon word nie kan hierdie raam nie as 'n leksikografiese soekplek beskou word nie. Dit word dus nie as sodanig benoem nie, maar staan as 'n skermkootraam bekend.

In gedrukte woordeboeke speel bladuitleg asook die uitleg en voorkoms van buitetekste 'n belangrike rol (Almind en Bergenholtz 2000). Aanlyn woordeboeke het ook 'n estetiese dimensie en in die kommersiële leksikografie moet die skermkootstruktuur so beplan word dat dit ook ruimte bied aan nieleksikografiese data. Dit mag nie lukraak gedoen word nie. Dit is deel van die leksikografiese proses dat die leksikograaf moet toesien dat die woordeboekspesifieke besetting van die skermkootstruktuur nie die gebruiker benadeel in die raadpleging van die woordeboek nie. Gevolglik is nie net die artikelstruktuur belangrik nie, maar ook die skermkootstruktuur en ook die funksionele benutting van 'n skermkootraam.

Data-oorlading kan maklik suksesvolle woordeboekraadpleging kortwiek — vergelyk Gouws en Tarp (2017). In figuur 20, die *Macmillan Dictionary* se openingskermkoot van die lemma *information overload* se artikel met sy uitgebreide artikelstruktuur en skermkootraamstruktuur is data-oorlading 'n wesenlike gevaar as die gebruiker slegs op soek is na die betekenis van *information overload*.



Figuur 20: Uit *Macmillan Dictionary*

Soos in die geval van artikelraamstrukture kan daar in die skermkootstruktuur ook 'n volle raam sowel as 'n halwe raam optree en dit geld op horisontale en vertikale vlak. Leksikograwe moet saam met uitgewers beplan aan die beste benutting van die nuwe tipes raamstrukture om te verseker dat geïntegreerde en niegeïntegreerde artikelrame asook die besetting van 'n skermkootraam tot voordeel van die woordeboek se teikengebruiker sal wees.

4. Ten slotte

Alhoewel die oorgang vanaf die gedrukte na die aanlyn omgewing reeds tot 'n groot verskeidenheid aanpassings in die leksikografie gelei het, is nog nie naastenby al die moontlikhede wat die nuwe medium bied, benut nie. Praktiese leksikograwe tree vernuwend op in hulle beplanning en produksie van aanlyn woordeboeke. Metaleksikograwe moet hierdie vernuwing in hulle teoretiese modelle insluit, maar self ook met vernuwendende voorstelle vorendag kom. Dan kan 'n grondige teoretiese benadering 'n vertrekpunt wees vir die verdere beplanning en produksie van goeie aanlyn woordeboeke. Sodoende kan die gebruiker baat by 'n omvattende data-aanbod en 'n optimale onttrekking van inligting aan hierdie data. In hierdie verband speel woordeboekstrukture 'n wesenlike rol. Leksikograwe moet saam met rekenaar- en inligtingswetenskaplikes werk om die gewenste strukture te ontwikkel wat sowel 'n omvattende dataverspreidingstruktuur as 'n soekstruktuur 'n werklikheid kan laat word. Nuwe strukture, soos die skermkootstruktuur, maar ook nuwe tipes raamstrukture kan 'n bydrae lewer om die gepaste leksikografiese data te akkommodeer en om 'n niesteurende posisie vir nieleksikografiese data te skep.

5. Bronnelys

5.1 Woordeboeke

5.1.1 Aanlyn woordeboeke (Almal geraadpleeg: Junie 2023)

AHD = *American Heritage Dictionary*.

<http://ahdictionary.com/>

ANW = *Algemeen Nederlands Woordenboek*.

<http://anw.inl.nl/>

dict.cc = <https://www.dict.cc/>

Duden = *Duden Online*.

<https://www.duden.de/suchen/dudenonline/>

elexiko = <http://www.owid.de/wb/elexiko/start.html>

LDOCE = *Longman Dictionary of Contemporary English*.

<http://www.ldoceonline.com>

Macmillan = *Macmillan Dictionary*.

<https://www.macmillandictionary.com/dictionary/british/>

(Woordeboek sluit op 30 Junie 2023.)

MW = *Merriam-Webster*.

<http://www.merriam-webster.com/dictionary>

OED = *Oxford English Dictionary*.

<https://www.oed.com/>

Pharos = *Pharos Afrikaans–Engels/English–Afrikaans Woordeboek*.

<https://www-pharosaanlyn-co-za.ez.sun.ac.za/soek/>

5.1.2 Gedrukte woordeboeke

Takebayashi, S. en Y. Kojima (Reds.). 1990. *Kenkyusha's Lighthouse English–Japanese Dictionary*. Tokio: Kenkyusha.

Wiegand, H.E., R.H. Gouws, M. Kammerer, M. Mann en W. Wolski (Reds.). 2020. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Deel 3*. Berlyn: De Gruyter.

5.2 Ander bronne

Almind, R. en H. Bergenholtz. 2000. Die ästhetische Dimension der Lexikographie. Fix, U. (Red.). 2000. *Bild im Text — Text und Bild*: 259-288. Heidelberg: Universitätsverlag C. Winther.

Gouws, R.H. 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
<https://doi.org/10.5788/24-1-1256>

Gouws, R.H. 2018. Accessibility, Access Structures and Access Procedures. Jesenšek, V. en M. Enčeva (Reds.). 2018. *Wörterbuchstrukturen zwischen Theorie und Praxis*: 35-56. Berlyn: De Gruyter.
<https://doi.org/10.1515/9783110598650-003>

Gouws, R.H. 2022. Extended Article Comments in Online Dictionaries. *Lexicography. Journal of ASIALEX* 9(2): 111-138.

Gouws, R.H. en S. Tarp. 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

Hausmann, F.J. en H.E. Wiegand. 1989. Component Parts and Structures of General Monolingual Dictionaries. Hausmann, F.J. et al. (Reds.). 1989-1991: 328-360.

Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*. Berlyn: Walter de Gruyter.

Kammerer, M. en H.E. Wiegand. 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.

Klosa, A. en R.H. Gouws. 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.

Wiegand, H.E. 1989. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Reds.). 1989-1991: 462-501.

The Concept of Lexicographic Condensation: A Review of and Perspectives on Digital Lexicography

M.^a Teresa Fuentes Morán, *Department of Translation and Interpreting,
University of Salamanca, Salamanca, Spain*
(tfuentes@usal.es)

Abstract: This study focuses on the concept of lexicographic textual condensation in order to determine its role in digital lexicography. Current interpretations of this concept, which was originally coined for printed dictionaries, are analysed. Special emphasis is placed on the relevant differentiation between the various textual levels in which the content of these works are distributed and on the difference between primary and secondary condensation. Examples from English–Spanish digital dictionaries will be explored in order to answer the question of how condensation is presented in these works and what types of components or items are particularly affected by it. Finally, meaning of condensation forms in current dictionaries based on the reading approaches applied to reference works is discussed. We support the argument that these forms are intrinsic to the type of *text* that we call a *dictionary*, and that the reasons behind condensation mechanisms are not limited to mere spatial constraints, but that these, together with other formal aspects, can make it easier to find the desired answers through a nonlinear reading process that is generally and legitimately applied to dictionaries.

Keywords: DIGITAL LEXICOGRAPHY, PRINTED LEXICOGRAPHY, LEXICOGRAPHIC TEXTUAL CONDENSATION, THEORY OF THE FORM OF DICTIONARIES, PRIMARY LEXICOGRAPHIC CONDENSATION, SECONDARY LEXICOGRAPHIC CONDENSATION, BILINGUAL ENGLISH–SPANISH LEXICOGRAPHY, MICROSTRUCTURE, REFERENCING, NATURAL LANGUAGES

Opsomming: Die konsep van leksikografiese verdigting: 'n Oorsig van en perspektiewe op die digitale leksikografie. In hierdie artikel word gefokus op die konsep van leksikografiese tekstuele verdigting om sodoende die rol daarvan in die digitale leksikografie te bepaal. Huidige interpretasies van hierdie konsep wat oorspronklik vir gedrukte woordeboeke geskep is, word geanaliseer. Besondere klem word geplaas op die relevante onderskeiding tussen die verskillende tekstuele vlakke waarin die inhoud van hierdie werke versprei is en op die verskil tussen primêre en sekondêre verdigting. Voorbeelde uit Engels–Spaanse digitale woordeboeke word ondersoek om te bepaal hoe verdigting in hierdie werke aangebied word en watter soorte komponente of items spesifiek hierdeur beïnvloed word. Laastens word die betekenis van verdigtingsvorme in huidige woordeboeke wat gebaseer is op die leesbenaderings wat op naslaanwerke toegepas word, bespreek. Ons steun die argument dat hierdie vorme inherent is aan die tipe *teks*

wat 'n *woordeboek* genoem word, en dat die oorsake van verdigtingsmeganismes nie beperk is tot blote ruimtelike beperkings nie, maar dat hierdie meganismes, saam met ander formele aspekte, dit makliker kan maak om die verlangde antwoorde deur middel van 'n nieliniêre leesproses wat oor die algemeen en aanvaarbaar op woordeboeke toegepas word, te vind.

Sleutelwoorde: DIGITALE LEKSIKOGRAFIE, GEDRUKTE LEKSIKOGRAFIE, LEKSIKOGRAFIESE TEKSTUELE VERDIGTING, TEORIE VAN WOORDEBOEKVORM, PRIMÊRE LEKSIKOGRAFIESE VERDIGTING, SEKONDÊRE LEKSIKOGRAFIESE VERDIGTING, TWEETALIGE ENGELS–SPAANSE LEKSIKOGRAFIE, MIKROSTRUKTUUR, VERWYSING, NATUURLIKE TALE

1. The concept of lexicographic condensation

The framework of the *general theory of the form of dictionaries* (e.g. Wiegand 1984, 1996a, 1997, Wiegand and Fuentes Morán 2009) presents and develops the concept of *lexicographic textual condensation* (Wiegand 1996b, Bustos Plaza and Wiegand 2005–2006) as one of the main pillars for the formal description of its subject matter. This concept attempts to explain partially the particular, albeit not unique, form in which data are presented in standardized dictionaries. Indeed, one of the characteristics that identifies the *lexicographic text*¹ is the fact that, to a large extent, linguistic data are essentially expressed throughout the dictionary with forms of expression that are not part of natural language (Wiegand 2009). As a consequence, lexicographic condensation is described as the process that leads from a full text (a text showing *complete* cohesion and explicit syntax) to a condensed article text (with addressing as syntax, cf. Gouws 2015).

This can be illustrated with a very simplified example. If we look for information about the word *increase* (which, in an attempt to simplify the question, will be discussed here exclusively as a verb) in a bilingual English–Spanish dictionary, we do not generally find entries in modern dictionaries with the following information:

- (1) *increase* is a verb
- (2) *increase* means aumentar

but rather, e.g.

- (a) **increase** *verb* aumentar
- or
- (b) **increase** *v* aumentar

When comparing (1) and (2), which are presented in natural language, with (a) or (b), which present the characteristics of a lexicographic *text*, we can observe that (a) and (b) do not include "is a", that "verb" in (b) is abbreviated and presented as "v", and that "means [...]" is omitted.² (a) and (b) are condensed sequences in non-natural language that try to provide the same propositional



content than (1) and (2). Both (a) and (b) apply one of the most common condensing procedures: omitting; and abbreviating is also a technique applied in (b). Therefore, (b) is more condensed than (a). Together with omitting and abbreviating, other condensation mechanisms have been described: shortening, shifting, substituting, summarizing, and embedding.³


However, *condensation* is described as a *transformation* process. If interpreted literally, this would imply that the text starts as a sequence in natural language (for example: "*increase* is a verb") which is then subjected to different transformations until it reaches the form in which the result is presented in the dictionary. However, we must admit that this principle can only ever be applied from a theoretical perspective (Rascón Caballero 2021: 104-106). That is, lexicographers in a real context do not generally formulate the contents in natural language and then *transform* them through condensation procedures in order to include the data in the dictionary according to the required standard format. Therefore, an actual *condensation* process is not taking place. However, the purpose behind the application of this concept goes beyond this principle. The objective is to determine formally the propositional content of the sequences that are transformed — in other words, the data that need to be transmitted — and to contribute to the description, development, or assessment, among others, of the corresponding forms of representation of these data in the dictionary. This, in turn, provides a tool that can be applied, for example, to perception studies — in other words, to studies on how the transmitted data are interpreted and on what potential sources of error can be found in these interpretations. Therefore, the strengths in the application of this concept include its contribution to the detailed description of the differences between natural language and the forms of expression used in the dictionary and, with it, to an accurate description of the different levels of *lexicographic textualization* (e.g. Wiegand 1996a). In sum, this type of description makes it possible to establish the initial guidelines of a partial methodology for the planning, design, comparison, and assessment of dictionaries on a formal plane — which is intrinsically connected to their contents (Wiegand 1998). This provides a base to study the way in which users infer the information that is conveyed through the dictionary, as supported by Nielsen (2002: 597):



The lexicographical information can only be inferred from the article when the user successfully establishes the relation between two items, e.g. the item giving the meaning paraphrase and the item giving the form of the lemma sign. Therefore, the way in which the items have been written is of the utmost importance in establishing this relation, and this may be analysed in terms of textual condensation, which also has a direct bearing on the readability of an article.

The incidence of condensed components in a lexicographic article, as can be seen in (a) and (b) allows us to classify it as being more or less condensed, as can be also seen in the following two examples:


increase

INTRANSITIVE VERB [ɪnˈkriːs]  

[number, size, speed, pain] **augmentar** 

[prices, temperature, pressure] **subir**  **↑** **augmentar** 

I.increase *v. intr.* [ ɪnˈkriːs,  ɪnˈkriːs]

 Mostrar la tabla de conjugación

 **increase** number/size: **augmentar**

 **increase** prices: **augmentar**

Figure 1: Partial article of *increase*
from *Collins E–S*

Figure 2: Partial article of *increase*
from *Pons/Oxford E–S*

The example in Figure 1 presents "intransitive verb" without an abbreviation (and therefore, without condensation), whereas the example in Figure 2 shows "V. intr.", which is an equivalent propositional content to the previous one, but which is presented here in its abbreviated (condensed) form. In any case, it is necessary to consider that all the dictionary articles, as they are currently known and according to the basic conventions of the *object* that we call a *dictionary*, always present a certain degree of condensation. Even if all the components in the lexicographic article were presented in their uncondensed form — which is not at all common — we always find at least some omissions: the components are not linked to each other through the procedures of natural languages, as can be seen in Figures 1 and 2. There are no sequences such as "an equivalent for *increase* in Spanish is ...", "the meaning of *increase* is ...", "*increase* is an intransitive verb ..." or "*increase*, as an intransitive verb, means ...". No, as observed, the corresponding data are formalized through different procedures that are typical of dictionaries and that allow us to identify them as such. Moreover, the process that we call *lemmatization* is a type of condensation that is typical of lexicography and characterizes it. For example, in all the dictionaries in which the word is included, *increase* is a lemma sign that, through a process of selection — and, consequently, the omission of all other potential elements — represents the entire paradigm of the verb.⁴

Therefore, it is essential to establish a category for a type of condensation that will be described as *primary* and that encompasses (1) the procedures that determine the form of the lemma sign, and (2) the omissions of expressions in natural language that join the different components of the article, as seen before (expressions such as "the meaning of ... is ..." or "... is an intransitive verb ...").

These omissions can also be seen in instances of non-lemmatic addressing. For example, bilingual dictionaries may show the following:

(a₂) **increase** *verb* **augmentar** *verbo*

or

(b₂) **increase** *v* **augmentar** *v*

Both "augmentar *verbo*" and "augmentar *v*" show the omission — a characteristic of modern dictionaries, as has been pointed out — of the element of natural language that would join syntactically "augmentar" with "*verbo*" (a₂) or with "*v*" (b₂).

Therefore, this is what we call a *primary* type of condensation that has been conducted through omission.

The other varied forms of lexicographic condensation which are generally decided when planning the design of each specific dictionary and which may be motivated by the linguistic conventions of each language — such as the abbreviation systems — may be referred to as *secondary* condensation.

Currently, primary condensation identifies the modern lexicographic text as such. Secondary condensation is just a possibility. It seems that those who insist that condensation is no longer necessary in digital dictionaries refer to this secondary form (Wolski 2020).⁵

The result of the condensation does not necessarily appear through linguistic signs, but also, for example, through icons and symbols, such as asterisks, arrows, etc., or through a combination of linguistic and non-linguistic signs. That is the case, for example, in the printed dictionary *Oxford Study*, targeted at Spanish-speaking English students:

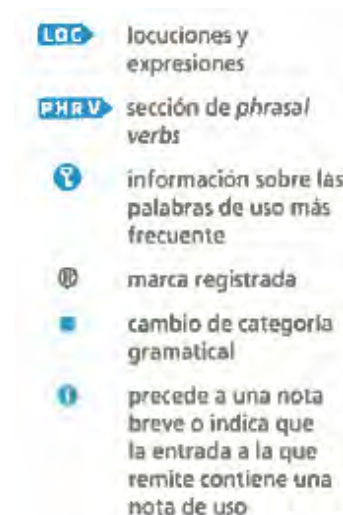


Figure 3: Fragment of the instructions in *Oxford Study*

Here, the symbol of an arrow with the letters "LOC" inside it is revealed as "locuciones y expresiones" [phrases and expressions], while the same symbol with the letters "PHR V" inside means "sección de phrasal verbs" [phrasal verb section]. The symbol of a key inside a circle means "información sobre uso de palabras de uso más frecuente" [information on the usage of the most frequent words]. In the example in Figure 2 there are symbols that represent the flags of the United States and the United Kingdom, which correspond to "American English" and "British English", respectively, as can be seen in Figure 4 by hovering the cursor over the symbols.



Figure 4: Pons/Oxford E-S explanation of the flag symbols

In contrast with the limitations of the two-dimensional surface of printed dictionaries, digital dictionaries may present up to three characteristic textual levels:

- (a) An initial level (extended by means of a *more-less* folding system)
- (b) A second level that is accessed by hovering the cursor over a component in an article
- (c) A third level that is accessed by clicking a component in an article

Let us briefly discuss some examples of these three levels.

- (a) The initial level is sometimes expanded or collapsed through the *more-less* feature (or a similar procedure), as can be seen in *Wordreference E-S*:

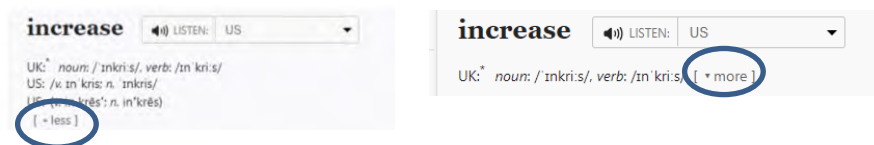


Figure 5: Partial article of *increase* in *Wordreference E-S*

- (b) In the example in Figure 4, the explanations "American English" and "British English" are visible by hovering the cursor over the flags, as previously shown. The flag symbols are, therefore, items that are presented in a condensed form, and which become uncondensed through a textual level that is accessed by hovering the cursor over them. Another example of un-condensation that is visible by hovering the cursor over a condensed form in the initial level can be found in *Wordreference E-S*:

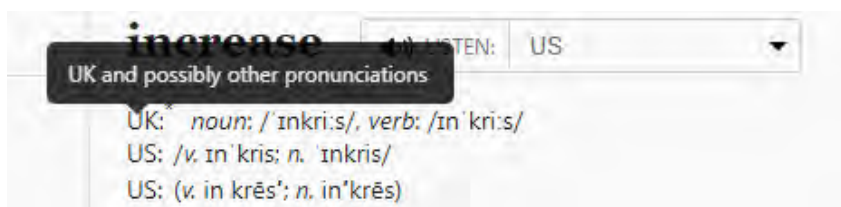


Figure 6: Partial article of *increase* in *Wordreference E-S* with information that is presented in a non-condensed form for UK through hovering

(c) Examples of the textual level that can be accessed by clicking are found, for example, in *Linguee E-S*:

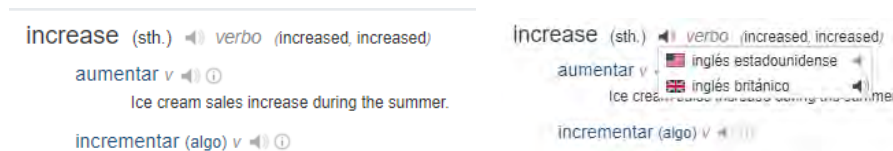


Figure 7: Partial article of *increase* in *Linguee E-S*

In *Linguee E-S* (in which the textual level is not developed by hovering), the information regarding pronunciation (a recording) is accessed by clicking on the loudspeaker symbol. This is a way to expand information through a reference: the initial level, through the symbol of the loudspeaker, refers to a different textual level in which, through another reference, the user can access the oral reproduction of the word.

As we can see, changes in the textual level are particularly used to expand the information (references) or to un-condense some of the elements that are initially condensed. The example in Figure 7 is a very common type of reference in digital dictionaries.

Although it is generally possible to establish a clear difference between the expansion of information achieved through *references* and *un-condensation*, we can also find some less precise cases in which both concepts are entwined. The following example can be found in *Cambridge E-S*:



Figure 8: Partial article of *increase* in *Cambridge E-S*

By clicking on [I or T] — hovering the cursor does not have any effect — we can access the section "Labels & Codes", where we can find the full list of these elements and, in several cases, a short explanation of their meaning (Figure 9). This is, therefore, an un-condensation procedure that presents the full form of items. In addition, the expanded information included in the original lexicographic article can be found in that same location.

Etiquetas y códigos

[Ayuda](#) = Etiquetas y códigos

Ir a:

[Adjetivos](#)

[Sustantivos](#)

[Verbos](#)

[Otras etiquetas](#)

[...]

Verbos

<i>verb</i>	Una palabra que describe una acción, estado o experiencia.
[T]	Verbo transitivo: un verbo que tiene objeto directo.
[I]	Verbo intransitivo: un verbo que no tiene objeto directo.

Figure 9: Partial article of *increase* in *Cambridge E–S*

2. Trends in digital lexicography (English–Spanish)

This section discusses the most characteristic types of items that include condensation and the type of procedures they follow in the selected dictionaries. Indeed, in some of these dictionaries we can find, apart from primary condensation forms, instances of secondary condensation, particularly in (1) morpho-syntactic items, such as *v* (verb) or *pl.* (*plural*) — probably the most classical and visible type — and (2) the way in which disambiguators are presented, and occasionally the definitions, equivalents and examples. The pertinence of including these items or the forms in which they are presented are not discussed here, but the fact is that we can find a range of different decisions in the examples from the selected dictionaries. An overview of the most significant cases will be presented and the way in which condensation has been applied in the articles corresponding to the verb *increase* will be analysed.

2.1 Morphosyntactic items

Included here are all the items that refer to different classes of words, syntax (such as *transitive/intransitive*), and morphological items (e.g. conjugation models). In *Collins E–S*, *Cambridge G. E–S*, and *Longman E–S*, we can find information presented in a non-condensed form regarding the fact that *increase* is a verb, and also regarding its transitive or intransitive nature. Figure 1 showed the example from *Collins E–S* ("intransitive verb"). Figure 10 shows two more examples.



Figure 10: Partial articles of *increase* from *Cambridge G. E-S* and *Longman E-S*

The dictionary *Nglish E-S* does not condense the word *verb*, but in this case there is no information on the transitivity or intransitivity of the word (Figure 11).

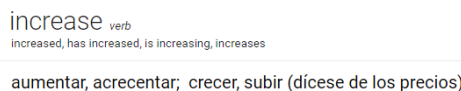


Figure 11: Partial article of *increase* from *Nglish E-S*

Examples where the form *verbo* or *verb* is not condensed but the information on its transitivity or intransitivity can be found in a condensed form in *Cambridge E-S* (Figure 8) and in *Linguee E-S* (Figure 12).

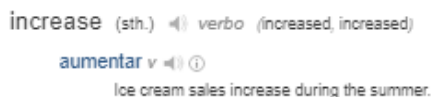


Figure 12: Partial article of *increase* from *Linguee E-S*

The information on the transitivity or intransitivity of the verb is indicated through the condensed form "(sth.)". In this case, the mention to the object is abbreviated (*sth.*, as an abbreviation of *something*), and the parentheses indicate that this is an optional element. The item is not linked here to another textual level and, therefore, it is a condensed element that does not give access to its explanation or un-condensation.

Both *Wordreference E-S* and *Pons/Oxford E-S* include condensed items on the category of the verb and its transitivity or intransitivity. Figure 13 shows an example from *Wordreference E-S*, which is discussed below, together with its conjugation forms (Figures 17–20).

Principal Translations		
Inglés		Español
increase	⇒ vi (go up)	aumentar ⇒ vtr subir ⇒ vtr

Figure 13: Partial article of *increase* from *Wordreference E-S*

In *Pons/Oxford E-S*, hovering the cursor over the condensed part of speech shows the following pop-up information (Figure 14):

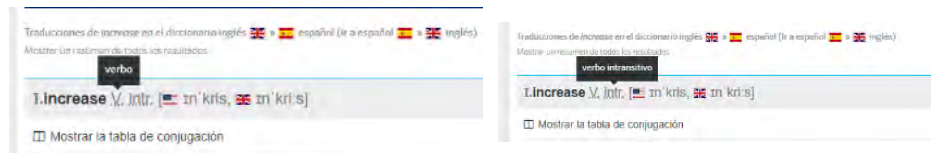


Figure 14: Partial article of *increase* from *Pons/Oxford E-S* with an explanation for "V." and "intr.", respectively

Although in *Pons/Oxford E-S* the item "intr." Always appears next to "V.", "intr." is uncondensed as "verbo intransitivo" [intransitive verb].

The forms of inclusion on the verb conjugation also show that different decisions have been made. *Pons/Oxford E-S* also presents a hyperlink to a conjugation table that provides extensive information related to the article through a process similar to referencing (Figure 15).

Conjugación de Increase									
Simple tenses Continuous tenses Conditional Imperative Impersonal									
SIMPLE TENSES									
Present			Past			Present Perfect			
I	increase		I	increased		I	have	increased	
you	increase		you	increased		you	have	increased	
he/she/it	increases		he/she/it	increased		he/she/it	has	increased	
we	increase		we	increased		we	have	increased	
you	increase		you	increased		you	have	increased	
they	increase		they	increased		they	have	increased	
Past Perfect									
I	had	increased							
you	had	increased							
he/she/it	had	increased							
we	had	increased							
you	had	increased							
they	had	increased							
Will-Future			Going-to-Future			Future Perfect			

Figure 15: Fragment from *Pons/Oxford E-S* with part of a conjugation table

In *Linguee E-S* (Figures 7 and 12), next to the word "verbo" can be found "(increased, increased)", which, as users may deduce, are the forms of the past simple and the participle (in no specific order). By clicking on "increased" more information is available (Figure 16):

^ Diccionario inglés-español

increased *adjetivo / participio pasado*

aumentado *adj m* (*aumentada f sing, aumentados m pl, aumentadas f pl*)

mayor *adj m/f*

The candidate strove for increased citizen engagement in his campaign. El candidato se esforzó por una mayor participación ciudadana en su campaña.

menos frecuente:

incrementado *adj m* · incrementada *adj f* · ampliado *adj m* · elevado *adj m* · subido *adj m* · acrecentado *adj m* · intensificado *adj m* · realizado *adj m* · encarecido *adj m* · acrecido *adj m*

Figure 16: Partial article of *increased* from *Linguee E–S*

The classical verb conjugation forms (in this case *increased*, *has increased*, *is increasing* and *increases*, in that order) can also be found in *NGLISH E–S* (Figure 11), although without any hyperlink, explanation or additional information.

The case of the tool *Wordreference E–S* is unique and will be discussed separately.

increase LISTEN: US

UK: ^{*} *noun*: /ˈɪnkriːs/, *verb*: /ɪnˈkriːs/
US: /v. ɪnˈkriːs; n. ˈɪnkriːs/
US: (v. ɪnˈkrɛsː; n. ɪnˈkrɛs)
[• less]

[definition](#) | [Conjugación \[ES\]](#) | [English synonyms](#) | [English collocations](#) | [Conjugator \[EN\]](#) | [in context](#) | [images](#)

Inflections of 'increase' (v): (= conjugate)
increases: v 3rd person singular
increasing: v pres p
increased: v past
increased: v past p

[WordReference](#) | [Collins](#) | [WR Reverse \(66\)](#)

WordReference English-Spanish Dictionary © 2022:

Principal Translations

Inglés	Español
increase ⇒ <i>vi</i> (go up) House prices have increased by 5%. El precio de la vivienda ha aumentado un 5%. El precio de la vivienda ha subido un 5%.	aumentar ⇒ <i>vtr</i> subir ⇒ <i>vtr</i>
increase <i>n</i> (rise) The increase in the number of models for sale failed to raise profits. El aumento en el número de modelos a la venta no dio como resultado más beneficios.	aumento <i>nm</i>

[Is something important missing? Report an error or suggest an improvement.](#)

Additional Translations

Inglés	Español
--------	---------

Figure 17: Partial article of *increase* from *Wordreference E–S*

The items on the word category and conjugation are condensed and uncondensed in the same textual level. This example presents the uncondensed items *noun* and *verb* in the first line, and the corresponding condensed items (*v.* and *n.*) in the second and third lines.⁶ These are, therefore, items with the same propositional content and presented at the same level, but in two different forms. The abbreviated items *v.* and *n.* are not explained by hovering the cursor over them or by clicking.

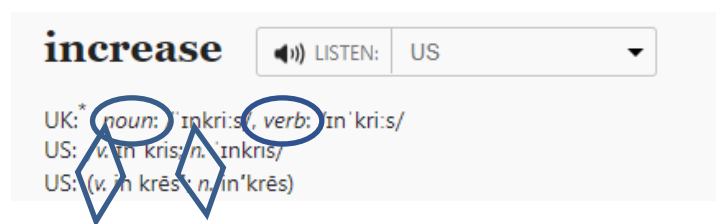


Figure 18: Partial article of *increase* from *Wordreference E-S*

The section "Inflections of 'increase'" also shows different items, related to inflection, and presented in both condensed and uncondensed form in the same textual level:

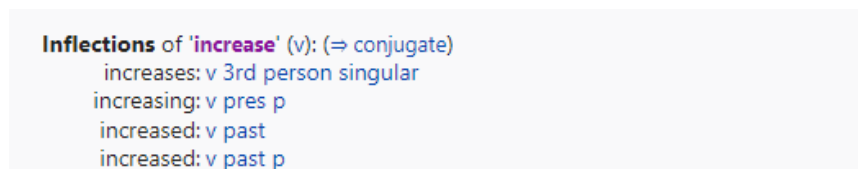


Figure 19: Partial article of *increase* from *Wordreference E-S*

The classical forms from the conjugation model are included (*increases*, *increasing*, *increased* and *increased*), as in other dictionaries, but in this case, they add an indication of the form (or person) for each of them. Once again, "person singular" is not condensed, but the forms "pres p", "past" and "past p" are partially condensed. In this case, condensed forms are presented in full in a second textual level that can be accessed by hovering the cursor over them:

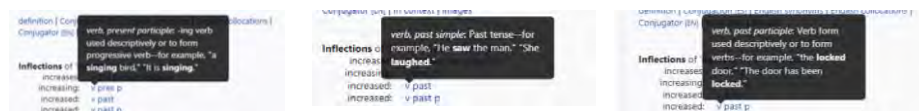


Figure 20: Partial article of *increase* from *Wordreference E-S*

Here, apart from the full form (1), we can find an explanation (2) of the meaning of "present participle", "past simple" and "past participle", respectively, and examples (3) for each of them. These are different types of information, with two different forms of addressing, in the same pop-up. In summary, as we can see, there is a wide range of decisions that have been taken regarding the presentation of morphosyntactic items.

2.2 Disambiguators and other condensed components

Semantic annotations act as disambiguators and establish differences between the meanings and equivalent alternatives. In *Collins E-S* (Figure 21), the annotation [number, size, speed, pain] is partially condensed through the omission of the characteristic linking elements from natural language, and through a linking procedure with commas; and it may be interpreted as a simple group of generic items of a cotext.



Figure 21: Disambiguators in *Collins E-S*

The same procedure can be seen in *Pons/Oxford E-S* (Figure 2) and in *Longman E-S*. In this last case, this does not appear in the English-Spanish section, but in the Spanish-English one (Figure 22), which may be due to the didactic approach of this dictionary.

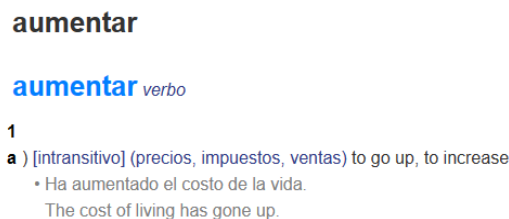


Figure 22: Disambiguators in *Longman E-S*

As is known, some bilingual dictionaries also contain definitions.⁷ The following example from *Cambridge E-S* (Figure 23) includes a case of condensation through embedding.

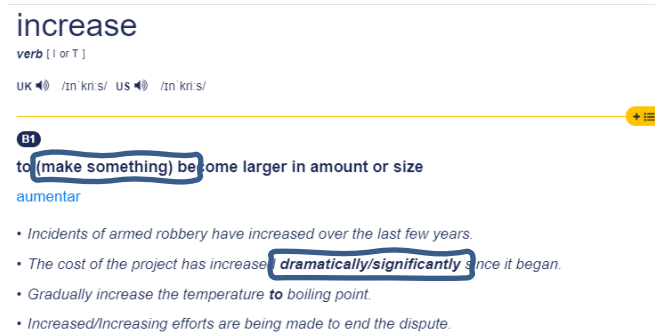


Figure 23: Embedding in a definition in *Cambridge E-S*

This example shows an embedding mechanism which is not uncommon in definitions, but which is less frequent within examples. The sentence "to (make something) become larger in amount or size" embeds two definitions into one: "to become larger in amount or size" and "to make something become larger in amount or size". A similar procedure can be found in the second example: "The cost of the project has increased dramatically/significantly since it began". As we know, the slash replaces "or" in natural language. The same type of formalization can be found in *Collins E-S* (Figure 21: "to increase in weight/volume/size/value ...") or *Cambridge G. E-S* (Figure 24).

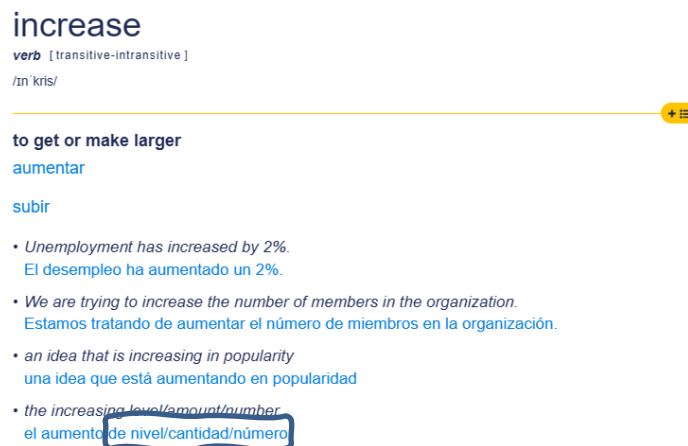


Figure 24: Embedding in a definition in *Cambridge G. E-S*

This example shows that the definition is not included in this procedure, because it is presented in natural language at all times, using "or". We can also find slashes in the presentation of translations of examples into Spanish, as can be seen in *Longman E-S* (Figure 25)



Figure 25: Translation with slashes in *Longman E-S*

Finally, we must mention another significant factor: the presence of identical condensed items ("US", second and third lines), with a difference that can only be observed by hovering the cursor over them:

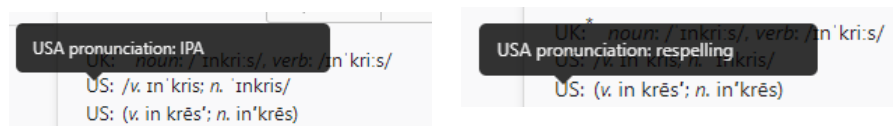


Figure 26: Partial article of *increase* from *Wordreference E-S*

3. Condensation in digital dictionaries?

The presence of condensed forms in the printed format was markedly determined by the need to save space and provide as much information as possible, and it was considered necessary in order to optimize the printing surface. This may be a plausible explanation for the use of abbreviations and acronyms, as well as other condensation forms such as omission or embedding. The lack of space is a reasonable explanation for the fact that data in printed dictionaries are presented in formats that diverge from natural language.

Over the last decades of the 19th century and the first decades of the 20th century, many social and commercial changes have expanded the scope of interlinguistic communication, and there was a growing need for dictionaries that were easier to transport and use. These changes paved the way for the

arrival of more reduced formats that joined large dictionaries, particularly those known as "pocket dictionaries".⁸ This decrease in size did involve including fewer entries and less information, but also led to the systematic development of different methods to save space and, as a consequence, of the forms that we now know as *condensation*. Dictionaries evolved to become increasingly standardized texts, and only rarely can we find, in the last decade of the 20th century, some examples of a slight change of trend in pedagogical lexicography.

However, in modern printed dictionaries, the question of space is combined with other aspects related to the forms in which the book is consulted, the ways in which it is *read*. In fact, textual density in dictionaries is not only motivated by the need to save space — which is more evident in printed dictionaries —, but also, as many other current means of communication, by the search for greater dynamism and effectiveness in the transmission of the message (Torres del Rey and Fuentes Morán 2013). In fact a dictionary is not a *text* meant to be read from left to right and from top to bottom (or in the direction required by each language). The same applies to each of the lexicographic articles it contains. This does not occur now, in digital dictionaries, just as it did not occur earlier, in printed formats.

Let us remember, in line with Rowe (2013), the difference between at least two reading modes — linear and tabular reading:⁹

The act of reading has numerous and highly distinct functions and attitudes [...]. Two basic types of engagement with a text may be identified, however, which are generally termed linear and tabular reading. Linear or intensive reading characterises the way we consume narrative fiction, [...] the reader of such works is typically highly absorbed in their storylines, borne along almost automatically by the temporal and causal narrative connections between sentences, paragraphs, and chapters. The tabular mode of reading, however, does not correspond to the storytelling form [...]. In this case, reading is interrogative, seeking information about a specific subject, and this is reflected in the formats of tabular texts — encyclopedias, dictionaries and other reference books. Such works employ a number of organisational strategies — including alphabetisation, block spacing of text, section headings, and so on [...], all of which are designed to facilitate the readers' search for specific information as well as to direct them towards the text's sources or related material if necessary.

In fact, the typical reading mode when consulting a dictionary is tabular, according to this classification, but it is also referred to as "scanning" (Rosenwald 2014) or hyper reading (Hayles 2010), for example, and its true purpose is to locate and obtain specific information or, in fewer cases, more general information about a lexical unit. We navigate across the dictionary and our eyes scan the lexicographic article until they have found (or they believe they have found) the information they were looking for. The process may continue if the previous results were not satisfactory, or it may end at that point, regardless of whether the article contained more information than what was actually assimilated by the reader. In dictionaries, as in many other *texts*, information is

structured with more or less advanced formal characteristics that enable this type of reading. These characteristics are position, size, font, style, color, and contents. With *contents* we refer to the components of the lexicographic article as they are presented in the dictionary, but regardless of their position, size, font, style, or color (because these elements, as structural indicators, are not meant to convey linguistic information and propositional contents, but to support and clarify the structural organization of the lexicographic article).

Finally, as stated by Krug (2014) when he tried to answer the question "Why do we scan?":

- [1] We're usually on a mission. Most Web use involves trying to get something done, and usually done quickly. As a result, Web users tend to act like sharks: They have to keep moving, or they'll die. We just don't have the time to read any more than necessary.
- [2] We know we don't need to read everything. On most pages, we're really only interested in a fraction of what's on the page. We're just looking for the bits that match our interests or the task at hand, and the rest of it is irrelevant. Scanning is how we find the relevant bits.
- [3] We're good at it. It's a basic skill: When you learn to read, you also learn to scan. We've been scanning newspapers, magazines, and books — or if you're under 25, probably reddit, Tumblr, or Facebook — all our lives to find the parts we're interested in, and we know that it works.

4. Conclusions and perspectives

The practices that were originally justified in printed formats because of the spatial constraints acquire a new dimension and different characteristics in digital lexicography. The revision of the essential aspects of the concept of lexicographic condensation allows us to reconsider its relevance and to qualify and modify it in the field of digital lexicography. This is not due to the inertia of classical procedures, but to the need to adapt the works to the changing characteristics of the surface — which can now be visualized in all sorts of screens, from the smallest mobile phone to the largest monitors — and to the different textual levels that can be defined. However, beyond these considerations, current formats promote or must promote and enable reading modes that were already characteristic of dictionaries and are now inherent to digital tools as a whole.

Acknowledgements

I want to thank the two anonymous reviewers and the editor of *Lexikos* for their detailed and valuable suggestions and comments. This research has been conducted thanks to a grant from the Spanish Ministry of Science and Innovation, as part of the project Parameters for more accessible multilingual lexical resources (PID2022-137210OB-I00).

Endnotes

1. We refer here to dictionaries and lexicographic articles as *text*, as usual, in a generic sense. We are aware, however, that in a strict linguistic sense, these works do not completely meet the requisites for textuality (cf. De Beaugrande and Dressler 1981, Bernárdez 1982).
2. We will not discuss other aspects that are intrinsically related to the entries, such as the characteristics of the format (bold, italic) that act as a support for information and, therefore, for legibility and readability (see Sections 3 and 4).
3. A summarized and exemplified account of lexicographic condensation can be found in Wolski 1989, on monolingual dictionaries; and in Wolski 1991, on bilingual dictionaries. The form in which each of these condensation types and methods can be described may be consulted in Wiegand 1996a, 1996b and Wiegand 1998, particularly. The works that analyze the condensation procedures of specific dictionaries include Nielsen 2002, Gouws 2005, and Fuentes Morán and Pradas Macías 2009.
4. This work will not establish a difference between *inner textual condensation* — which concerns all those lexicographic partial texts containing a carrier of the guiding element, in particular dictionary articles and index entries — and *outer textual condensation* — pertaining to the carriers of the guiding element such as lemmata and sub-lemmata, for example. This difference is particularly relevant for printed dictionaries. In addition, a qualitative approach, rather than quantitative, will be applied here, therefore, the different degrees of condensation are not described here.
5. All the dictionaries used here to provide examples show instances of primary condensation. As we shall see, they also present secondary condensation to different extents.
6. Other possible incoherences regarding position and structure can also be seen, but they will not be discussed here.
7. This study will not discuss whether these dictionaries are strictly bilingual, bilingualized, semi-bilingual, etc. They will be treated as bilingual dictionaries because they contain two languages which are analyzed, and our teaching experience indicates that students do treat them as bilingual dictionaries.
8. Observing and interpreting the evolution of dictionary formats through history can only be done with a broad perspective. This evolution is influenced by a wide range of factors, among which commercial, ideological and socio-cultural ones should not be overlooked. Some interesting notes on this can be found e.g. in Cowie 1998 and 1999, Hartmann 2000 or Collison 1982.
9. There are several works that cast light on this topic and present different advances. We particularly recommend Durant 2017 because of its analytical literature review on reading forms in the digital era within the framework of the widely known *Charleston Briefings*.

References

A. Dictionaries

Cambridge Dictionary (<https://dictionary.cambridge.org/>), español–inglés (*Cambridge S–E*), inglés–español (*Cambridge E–S*).

Cambridge Global Dictionary (<https://dictionary.cambridge.org/>), español–inglés (*Cambridge G. S–E*), inglés–español (*Cambridge G. E–S*).
Collins Dictionary (<https://www.collinsdictionary.com/>), inglés (Collins-E), inglés–español (*Collins E–S*) español–inglés (*Collins S–E*).
Linguee (<https://www.linguee.es/>), español–inglés (*Linguee S–E*), inglés–español (*Linguee E–S*).
Longman (<https://www.ldoceonline.com/>), español–inglés (*Longman S–E*), inglés–español (*Longman E–S*).
Nglsh by Britannica (<https://www.nglsh.com/>), español–inglés (*Nglsh S–E*), inglés–español (*Nglsh E–S*).
Pons/Oxford Dictionary (<https://es.pons.com/traducci%C3%B3n>), español–inglés (*Pons/Oxford S–E*), inglés–español (*Pons/Oxford E–S*).
Wordreference (<https://www.wordreference.com/>), español–inglés (*Wordreference S–E*) inglés–español (*Wordreference E–S*).

B. Other literature

- Bernárdez, E.** 1982. *Introducción a la Lingüística del Texto*. Madrid: Espasa-Calpe.
- Bustos Plaza, A. and H.E. Wiegand.** 2005–2006. Condensación textual lexicográfica: esbozo de una concepción integral. *Revista de Lexicografía* 12: 7-46.
- Collison, R.L.** 1982. *A History of Foreign-language Dictionaries*. London: André Deutsch.
- Cowie, A.** 1998. A.S. Hornby: A Centenary Tribute. Fontenelle, T. et al. (Eds.). 1998. *EURALEX '98 Proceedings*: 3-16. Liège: English and Dutch Departments, University of Liège.
- Cowie, A.** 1999. Learners' Dictionaries in a Historical and a Theoretical Perspective. Herbst, T. and K. Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?)*: 3-13. *Lexicographica*. Series Maior 95. Tübingen: Niemeyer.
- De Beaugrande, R. and W. Dressler.** 1981. *Introduction to Text Linguistics*. London: Longman.
- Durant, D.M.** 2017. *Reading in a Digital Age*. ATG LLC (Media).
- Fuentes Morán, M.T. and E.M. Pradas Macías.** 2009. Zur Textverdichtung in spanischen Lernerwörterbüchern. *Lexicographica: International Annual for Lexicography* 25: 3-20.
- Gouws, R.H.** 2005. Textual Condensation in the DUDEN. DAS GROSSE WÖRTERBUCH DER DEUTSCHEN SPRACHE IN ZEHN BÄNDEN. Herbert Ernst Wiegand (Ed.). 2005. *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden. Print- und CD-ROM-Version. Vol. 2*: 279-294. Berlin/Boston: Max Niemeyer.
- Gouws, R.H.** 2015. Expanding the Notion of Addressing Relations. *Lexicography. Journal of ASIALEX* 1(2): 159-184.
- Hartmann, R.** 2000. European Lexicography: Perspectives on Dictionary Research, with Special Reference to the Countries of the European Union. *Dictionaries: Journal of the Dictionary Society of North America* 21: 1-21.
- Hayles, N.K.** 2010. How We Read: Close, Hyper, Machine. *ADE Bulletin* 150: 62-78.
<https://www.maps.mla.org/content/download/155872/file/ade.150.62.pdf>
- Krug, S.** 2014. *Don't Make Me Think, Revisited. A Common Sense Approach to Web Usability*. Third edition. San Francisco: New Riders.
- Nielsen, S.** 2002. Textual Condensation in the Articles of *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Wiegand, H.E. (Ed.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II: Untersuchungen anhand des »de Gruyter Wörterbuchs Deutsch als Fremdsprache«*: 597-608. Tübingen: Max Niemeyer.

- Rascón Caballero, A.** 2021. *Teoría y práctica del ejemplo lexicográfico bilingüe*. Barcelona: Universidad Autónoma de Barcelona.
- Rosenwald, M.S.** 2014. Serious Reading Takes a Hit from Online Scanning and Skimming, Researchers Say. *Washington Post*, April 6, 2014. Available at:
https://www.washingtonpost.com/local/serious-reading-takes-a-hit-from-online-scanning-and-skimming-researchers-say/2014/04/06/088028d2-b5d2-11e3-b899-20667de76985_story.html.
- Rowe, C.** 2013. The New Library of Babel? Borges, Digitisation and the Myth of a Universal Library. *First Monday* 18(2).
<https://doi.org/10.5210/fm.v18i2.3237>.
- Torres del Rey, J. and M.T. Fuentes Morán.** 2013. Diccionarios electrónicos: (re) estructuras de acceso. Cordon García, J.A., R. Gómez Díaz and J. Alonso Arévalo. 2013. *Documentos electrónicos y textualidades digitales*: 237-245. Salamanca: Universidad de Salamanca.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter Proceedings. Papers from the International Conference on Lexicography at Exeter 9.-12. September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Niemeyer.
- Wiegand, H.E.** 1996a. A Theory of Lexicographic Texts: An Overview. *South African Journal of Linguistics* 14(4): 134-149.
- Wiegand, H.E.** 1996b. Textual Condensation in Printed Dictionaries. A Theoretical Draft. *Lexikos* 6: 133-158.
- Wiegand, H.E.** 1997. Printed Language Dictionaries and their Standardization: Notes on the Progress toward a General Theory of Lexicography. Hock, H.H. (Ed.). 1997. *Historical, Indo-European, and Lexicographical Studies. A Festschrift for Ladislav Zgusta on the Occasion of his 70th Birthday*: 319-380. Trends in Linguistics. Studies and Monographs 90. Berlin/New York: De Gruyter Mouton.
- Wiegand, H.E.** 1998. Lexikographische Textverdichtung. Entwurf zu einer vollständigen Konzeption. Zettersten, A., V.H. Pedersen and J.E. Mogensen (Eds.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth Symposium on Lexicography May 2-4, 1996 at the University of Copenhagen*: 1-35. Lexicographica. Series Maior 90. Tübingen: Niemeyer.
- Wiegand, H.E.** 2009. Nichtnatürlich über natürliche Sprache schreiben. Zu einigen formalen Aspekten von Wörterbuchartikeln. Felder, E. (Ed.). 2009. *Sprache*: 247-285. Heidelberger Jahrbücher 53. Berlin: Springer.
https://doi.org/10.1007/978-3-642-00342-4_10
- Wiegand, H.E. and M.T. Fuentes Morán.** 2009. *Estructuras lexicográficas: aspectos centrales de una teoría de la forma del diccionario*. Granada: Tragacanto.
- Wolski, W.** 1989. Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Vol. 1: 956-967. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1. Berlin/New York: De Gruyter.
- Wolski, W.** 1991. Formen der Textverdichtung im zweisprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Vol. 1: 2837-2854. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1. Berlin/New York: De Gruyter.

Wolski, W. 2020. Textverdichtung. Wiegand, H.E. et al. 2020. *Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research*. Vol 3, I–U: 934-937. Berlin/Boston: De Gruyter.

C. Additional literature

Béjoint, H. 2000. *Modern Lexicography. An Introduction*. Oxford: Oxford University Press.

Gorodischer, H. and J. Scaglione. 2020. *Legibilidad y tipografía: La composición de los textos*. Valencia: Campgràfic Editors, S.L.

Gouws, R.H. 1999. Mediostructural Representation, Textual Condensation and User-orientation in the WAT X. *Lexicographica: International Annual for Lexicography* 15: 4-37.

Ising, G. 1978. Textverdichtung und Redundanz in der Lexikographie. *Wissenschaftliche Zeitschrift der Wilhelm-Pieck-Universität Rostock* 27. *Gesellschafts- und sprachwissenschaftliche Reihe* (1/2): 9-13.

Jardí, E. 2021. *Cincuenta y tantos consejos sobre tipografía*. New revised and expanded edition. Barcelona: Editorial GG.

Rahnenführer, I. and G. Böhme. 1995. Orthographiewörterbuch und Textverdichtung. Ewald, P. und K.-E. Sommerfeldt (Eds.). 1995. *Beiträge zur Schriftlinguistik. Festschrift zum 60. Geburtstag von Dieter Nerijs*: 257-268. *Sprache, System und Tätigkeit* 15. Frankfurt a.M./Berlin/Bern/New York/Paris/Vienna, Lang.

Torres del Rey, J. 2009. Dicionarios electrónicos bilingües: nuevas posibilidades de futuro. Fuentes Morán, M.T. and B.A. Model (Eds.). 2009. *Investigaciones sobre lexicografía bilingüe*: 29-79. Granada: Tragacanto.

A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and Implications for the Future

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

and

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

Abstract: This paper deals with developments in the field of lexicography that resulted in the need for a new definition of this term. The paper offers a brief look at the origin, use and development of the term *lexicography* and at the use of glosses in former and current times. It is shown how the digital era enables the use of certain lexicographical features in other sources than dictionaries. This leads to an expansion of the scope of the term lexicography. A strong focus is on the use of glosses, originally inserted by scribes as snippets into manuscripts to help with the understanding of difficult words and expressions. The use of glosses has increased, and they are currently commonly used in a variety of environments. In digital products glosses play an innovative, productive and significant role to present new types of lexicographical data. This demands the recognition of glosses as lexicographical entries and leads to a redefinition of the term *lexicography* that includes two major sub-fields, namely *dictionography* and *glossography*.

Keywords: ARTIFICIAL INTELLIGENCE, DIGITAL ERA, GLOSS, GLOSSOGRAPHY, LEXICOGRAPHICAL DATA, SCRIBE, DICTIONOGRAPHY, INTERDISCIPLINARY COLLABORATION, READING ASSISTANT, WRITING ASSISTANT

Opsomming: 'n Noodsaaklike herdefiniëring van leksikografie in die digitale era: Glossografie, woordegrafie en implikasies vir die toekoms. Hierdie artikel handel oor ontwikkelinge in die veld van leksikografie wat gelei het tot die behoefte aan 'n nuwe definisie vir hierdie term. Dit bied 'n blik op die herkoms, gebruik en ontwikkeling van die term *leksikografie* en die gebruik van glosse in die verlede en tans. Daar word gewys hoe die digitale era die gebruik van bepaalde leksikografiese kenmerke in ander bronne as woordeboeke moontlik maak. Dit lei tot 'n uitbreiding van die bestek van die term *leksikografie*. Daar is 'n sterk fokus op die gebruik van glosse wat oorspronklik deur skriptore as brokkies in manuskripte gevoeg is om te help met 'n beter begrip van moeilike woorde en uitdrukkings. Die gebruik van glosse het toeneem en word tans algemeen gebruik in 'n verskeidenheid omgewings. In digitale produkte speel

glosse 'n vernuwend, produktiewe en belangrike rol om nuwe tipes leksikografiese data aan te bied. Dit vereis die erkenning van glosse as leksikografiese inskrywings en lei tot 'n hedefiniëring van die term *leksikografie* wat twee hoofafdelings insluit, te wete *woordegrafie* en *glossografie*.

Sleutelwoorde: KUNSMATIGE INTELLIGENSIE, DIGITALE ERA, GLOS, GLOSSOGRAFIE, LEKSIKOGRAFIESE DATA, SKRIPTOR, WOORDEGRAFIE, INTERDISSIPLINÊRE SAMEWERKING, LEESHULP, SKRYFHULP

0. Introduction

Lexicography as a cultural practice dates back at least to the 24th century BCE, when the Sumerian–Akkadian glossary known as *Urra–Hubullu* was carved on clay tablets, believed to be the first work of its kind in the world. With the term *lexicography* coined much later, the discipline has developed in both breadth and depth over the past 4300 years, with dictionaries covering virtually all the world's languages and countless aspects of human life in terms of knowledge, arts, crafts and other pursuits, making them one of the most successful man-made products ever. At the same time, its various expressions have undergone profound changes, not only as a result of accumulated experience, but also due to periodic technological innovations that have affected the preparation, presentation and use of the finished product. Suffice it to mention the evolution from clay tablets, through bamboo strips, papyrus scrolls, handwritten and printed paper books, to today's digital platforms. And the same can be said of the evolution of its production tools, from styluses to pens, typewriters, computers and, most recently, artificial intelligence, to name but a few.

All this has influenced the way lexicographers work and, over time, has greatly changed the way they carry out their tasks and relate to the empirical material, the compilation tools, the presentation of the final product, as well as to its users. In particular, the digital revolution that began at the end of the last century has been a major driver of this development. It is therefore quite natural that Grefenstette (1998), looking forward another 1000 years, formulated the following question in the title of the paper he presented at the Euralex Congress in the same year:

Will there be lexicographers in the year 3000?

It was Rundell (2012: 18) who took up the gauntlet and gave an answer that was as optimistic as it was challenging:

There will still be lexicographers, but they will no longer do the same job.

What will they do then? That's the question! And it might be added that not only will they not be doing the same work, they will also be making very different products, although this does not mean that dictionaries of some kind will not continue to exist, at least for the time being.

This article will not speculate and try to guess what lexicography will look like in the year 3000. But its authors are convinced that current technological breakthroughs, particularly in the field of artificial intelligence, will have a huge impact on the discipline in the coming years, as can be seen both in the recent adoption of chatbots in the lexicographic compilation process and in the increasing integration of lexicographical data into AI-based language models and the creation of entirely new products.

This development requires lexicographers to adapt, as must the discipline itself. In this perspective, there is an urgent need to redefine the very term *lexicography* and to bring it into line with actual practice, taking into account both history and current challenges.

The following sections will deal exclusively with lexicography in a European context, from its origins in ancient Greece up to the present day. Section 1 gives a brief overview of how the subject matter of lexicography has been defined in dictionaries and scholarly literature until now, and Section 2 then reflects on the origin and development of this term. Section 3 traces the use and development of glosses from antiquity to the present day, with special focus on the period after the introduction of printing technology, especially on the more recent use. Section 4 then briefly discusses what seems to be the rebirth of glosses in the digital age, while the following section provides some examples of how lexicographical data are used in AI-driven writing assistants. Finally, Section 6 attempts a redefinition of lexicography based on the findings and arguments of the previous sections.

1. Definitions of the term *lexicography* until now

When attempting a redefinition of the term *lexicography* it is necessary first to take cognizance of the meaning traditionally associated with this term. This can be found in definitions in general dictionaries, but also in sources beyond these dictionaries. Bergenholtz and Gouws (2012: 32) indicate that the use of the term *lexicography* shows many differences in its interpretation and there is a variety of perspectives on the nature, extent and scope of this term within the broader lexicographical and metalexicographical field.

The typical definition of the term *lexicography* in general dictionaries includes a reference to the making, compiling or writing of dictionaries, as seen in the following definitions from the *Longman Dictionary of Contemporary English* and the *The New Oxford Dictionary of English*, respectively:

The writing and making of dictionaries

the practice of compiling dictionaries

Some dictionaries do present a second sense that provides for something more than the practice of writing dictionaries. This is seen in the *Webster's Ninth Collegiate Dictionary* that has two senses for the term *lexicography*:

1: The editing or making of a dictionary. 2: the principles and practices of dictionary making.

These principles referred to in the second sense remain undefined, the focus remains on dictionaries and no mention is made of a formal theoretical component. Even in dictionaries dealing with special fields, the theoretical component of lexicography is not always acknowledged. In the *Wörterbuch zur Lexikographie und Wörterbuchforschung*, the brief tautological definition of lexicography makes no reference to a theoretical component:

total of all activities directed at the preparation of a lexicographic reference work.

A significant aspect of this definition is that it does not specifically link lexicography to dictionaries, but rather to the broader concept of lexicographical reference work, which includes, among others, dictionaries, encyclopaedias, thesauri, lexicons and glossaries.

Explicit reference to the theoretical component of lexicography can be found in the definition given in a specialized dictionary dealing with the discipline itself, namely the *Dictionary of Lexicography*. This source defines *lexicography* as:

The professional activity and academic field concerned with DICTIONARIES and other REFERENCE WORKS. It has two basic divisions: lexicographic practice, or DICTIONARY-MAKING, and lexicographic theory, or DICTIONARY RESEARCH. ...

Important here is the reference not only to dictionaries but also to other reference works in general.

Scholarly publications in the field of lexicography do not always acknowledge the scientific nature of theoretical lexicography, for example in the title of *Dictionaries: The Art and Craft of Lexicography* (Landau 1984), which does not indicate that lexicography is a scientific field or that the compilation of dictionaries may have a scientific or a theoretical basis. However, the recognition of a theoretical component of lexicography does come to the fore more strongly in other scientific publications from this field. Wiegand (1998: 41), when discussing general language dictionaries, distinguishes between a non-scientific and a scientific form of lexicography. He regards scientific lexicography as an independent cultural and scientific practice that is directed at the production of language reference works and these products should enable the establishment of another cultural practice, namely the use of these reference works. Although he only refers to "language dictionaries" the term *lexicography* also includes all other types of dictionaries — as shown in subsequent sections of this paper. In addition to the practice of using dictionaries Wiegand (1998: 46) makes provision for theoretical lexicography when stating that lexicography is the subject matter domain from which the different research areas of dictionary research develop.

It is worthwhile noting that for Wiegand lexicography has to do with reference works and not specifically with dictionaries.

From the different ways in which the term *lexicography* has been defined one can deduce that a default understanding of the term *lexicography* makes provision for two types, namely:

- (1) The planning and compilation of concrete dictionaries and other reference works. This part of lexicography is known as practical lexicography or the lexicographical practice.
- (2) The development of theories about and the conceptualisation of dictionaries and other reference works. This part of lexicography is known as meta-lexicography or theoretical lexicography.

The focus in the majority of the above-mentioned definitions has been on the making, compilation and editing of dictionaries, but not on the presentation of dictionaries or on dictionaries as such. A more directed focus on dictionaries (and other reference works) is found in Tarp (2018: 19) when he defines lexicography as:

the discipline that deals with dictionaries and other reference works designed to be consulted in order to retrieve information.

Lexicography deals with dictionaries but not only with dictionaries and not only with dictionaries that have been planned and compiled in a traditional way and, as we will see, something essential has been forgotten in previous definitions of this field. Consequently, a more comprehensive interpretation of this term is needed. Currently other products like some writing assistants, reading assistants and learning apps also display certain lexicographical features, cf. Tarp (2020), Huang and Tarp (2021) and Nomdedeu-Rull and Tarp (2024). This needs to be accounted for in a definition of *lexicography*. Once again lexicography faces a transition that can drastically influence the nature of the term.

Lexicography is increasingly dealing with products other than dictionaries — something omitted in the discussion of e.g. Bergenholtz and Gouws (2012) and Tarp (2018). The definition of the term *lexicography* needs to explicitly acknowledge these other products as being part of a lexicographic practice. When defining the term *lexicography*, it could be helpful to distinguish between a dictionary perspective and a non-dictionary perspective. The first perspective will lead to a focus on the traditional lexicographic products, in printed and digital format, like dictionaries, glossaries, thesauri, lexicons, encyclopaedias, and so on. This component of lexicography can be described with the term *dictionary*. The second component, focusing on other products, will be discussed in subsequent sections of this article.

2. Origin and development of the term *lexicography*

European lexicography was born and began to develop more than two thousand years ago in ancient Greece, as reported by McArthur (1986) and Stathi (2006),

among others. It is therefore appropriate that the term commonly used in the different European languages to describe the discipline defined in the previous section has its etymological roots in the classical Greek language. The term is composed of the two words *léxis* and *gráphein*, meaning "word" and "to write", respectively. This suggests that *lexicography* originally meant "writing about words".

The term *lexicography* was introduced into the Western European tradition in the seventeenth century, but it is an open question whether it was already used, at least orally, as early as Classical Greece, for example in the world-famous Library of Alexandria, whose scholars were the first in the European tradition to compile glossaries, the prototypes of later dictionaries.

In any case, the original meaning of a word, however important, does not necessarily reflect its meaning in later periods, since words, like everything else, are subject to the laws of evolution and change. According to current knowledge (see Hanks 2013), what we now call European lexicography dates back to the fifth century BCE, when Greek scribes began inserting glosses into manuscript copies of texts by Homer and other classical authors to explain rare or obsolete words to the readers of the time. This writing about the words was, of course, lexicography in the original sense of the word, even if the term was not yet in use. It is only later, when the scholars of the Library of Alexandria began to compile the glosses into glossaries, that the term can reasonably be associated with the compilation of word lists or glossaries.

Initially, glossaries were organised systematically, with words arranged in the order in which they appeared in a given book. In the second century BC, Dionysius Thrax, who worked in the library and wrote the famous *Tékhnē Grammatikē* (Art of Grammar), suggested that they should be arranged alphabetically instead. Although he did not use the term *lemma*, this scholar is also credited with inventing the *concept of a lemma* as a headword or form of citation, representing all the inflected forms of the respective words, to be followed by definitions and other data. In this way, he and other scholars of the Hellenistic period helped to standardise the content of glossaries, paving the way for their more sophisticated descendants, the dictionaries.

As a result, and without using modern lexicographical terminology, the ancient Greeks introduced five innovations that revolutionised the discipline: (1) the book form; (2) the lemma; (3) the macrostructure as the arrangement of the lemmas; (4) the article as the set of all the data attached to the lemma; and (5) the microstructure as the arrangement of these data. These five innovations are the key features of the uniquely successful product that evolved into the dictionary as we know it today, and took it from triumph to triumph for the next two thousand years.

Just as the meaning of the term *lexicography* was narrowed down to primarily refer only to dictionaries, so the meaning of the term *dictionary* itself changed over time, especially with the introduction of specialised dictionaries, spurred on by the European Renaissance, Enlightenment Age and industriali-

sation. The dictionary form was simply so successful that it was adopted by subject-field specialists who did not just write about words. From this perspective, the editors of the multi-volume Spanish *Diccionario enciclopédico hispano-americano de literatura, ciencias y artes* (1887–1910: 3–4) explained their choice of structure as follows:

In order for our work to be of truly practical use, it is essential that it should be easy to use, and to achieve this, there is no doubt that the most appropriate form is that of a dictionary, since the alphabetical order is the least likely to surprise people who are not well versed in the difficult matter of the classification of sciences, which is always prone to arbitrary errors.

The large number of dictionaries written in this spirit meant that the term *dictionary*, which until then had referred to something about language, now included "things and facts", as d'Alembert (1754: 958) astutely noted when he classified these works in a special article in the Great French Encyclopaedia. Although many lexicographers with a linguistic background resisted, this could only mean that the term *lexicography*, which had previously been narrowed down to refer specifically to dictionaries, was being broadened again to mean not only *writing about words*, but also *about things and facts*, i.e. what words refer to.

In all this development, which ended up equating lexicography with dictionaries and other reference works, one thing had been tacitly left out: the glosses. The insertion of glosses in manuscript copies was the starting point not only for classical Greek lexicography, but also for European lexicography in general. Moreover, history repeated itself in several European countries. Thus, the use of glosses in handwritten books was also the trigger for specific national lexicographies such as English and Spanish, as shown by Yong and Peng (2022), and Nomdedeu-Rull and Tarp (2024), respectively. At the same time, glosses, like dictionaries, continued to evolve and find new uses even after manuscript copying of books was abandoned with the introduction of printing in the fifteenth century, as we will discuss in the next section.

3. Use and development of glosses over time

The word *gloss* can be traced back to the Classical Greek word *glossa* that refers to a difficult word or a word that needs explanation. Where a glossa occurred in a text a brief explanation was added to the text. In the course of time the meaning of the word *gloss* shifted from referring to a difficult word to the brief explanation of such a word — like the snippets inserted into manuscripts by the scribes. Glosses were entered into texts where a reader of such a text realised that people might have difficulties with the understanding of a specific word or where they could benefit from some form of additional information. Therefore, not all words in a text received a gloss — too many glosses would have been redundant because the readers had no need for an explanation of the meaning of all the words in the text. Glosses were entered on a need to have basis.

Figure 1 is reproduced from Ruiz-Asencio et al. (2020), which stands out as one of the few systematic studies of glosses in specific medieval works. It shows 4 glosses out of the 369 glosses that appear in the mid-11th century Latin Penitential, known as *Glosas Silenses* (Silos Glosses). The four glosses (*non quisieret dare, vivere, culpaules* and *o sen tiestes testimonio*) are written in an early Romance language that was in the process of splitting off from vulgar Latin. It is worth noting the four different small signs used to refer the reader from a particular word in the Latin text to the corresponding gloss in the margin.

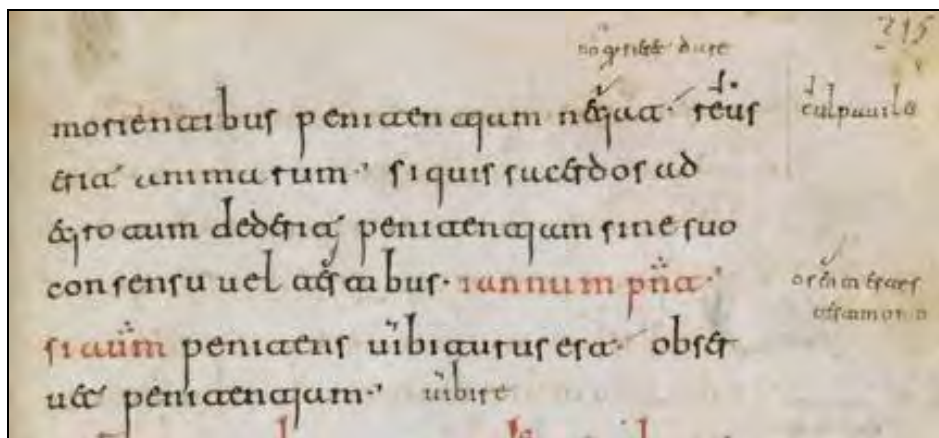


Figure 1: Extract from the *Glosas Silenses*

Glosses in books such as the *Glosas Silenses* were not only written by the scribes who would hand-copy these books from time to time due to wear and tear. They were also produced by readers, in most cases probably respected monks authorised to do so because of the extremely high value of books at the time. Just like the scribes, these new players responded to something in the text, for example to clarify some confusing concept. However, if too many glosses were inserted on a single page, it was difficult for the reader to link a specific gloss to the appropriate glossed word or expression. Consequently, the writers of glosses started to include signs, as can be seen in Figure 1, to help readers to connect a gloss to a specific word or expression. These were functional signs and nothing less than a prelude to the numbers, letters and other signs used much later to assign footnotes to particular words and phrases in a text.

As Tarp and Gouws (2019) have shown, the tradition of inserting glosses into written texts survived the introduction of printing in the late 15th century and found a number of new expressions. But the key players were different. Now it was either the authors themselves who introduced them into their books and articles, or it was the editors, especially when publishing later editions that needed additional clarification on meaning or other matters, thus performing a task

very similar to that of the classical Greek scribes more than two thousand years ago. By contrast, the readers of these texts now only added comments for their own personal use. In this respect, and according to Kennedy (2019), since the late 16th-century glosses were moved from the margins of texts to a designated place at the bottom of a page. Here they were entered as footnotes and assigned a number or sign corresponding to the number or sign entered next to a word or expression in the main text. This improved the organisation of the page layout and in the course of time such footnotes were often populated by glosses commenting on specific words and expressions in the text.

Today glosses are employed in different types of texts and different genres, including dictionaries. However, the most prolific use of glosses is found in non-lexicographic work where they are entered to provide assistance in, especially, a better comprehension of an unknown or foreign word. As Tarp and Gouws (2019) show, glosses are also presented in different forms, including their occurrence in footnotes.

Traditionally glosses were used in Greek and Latin texts, also texts of a religious nature. Even today in some editions of *The Bible*, glosses are still used to comment on a word or expression, or a thing referred to in the text. These glosses are usually not inserted as an annotation next to the word at which it is directed. Adhering to the principle applied in the *Glosas Silenses* and continued in the footnote approach, the user is directed from the glossed word or expression to a gloss accommodated elsewhere in the text — not in such close proximity as in the *Glosas Silenses* but rather in a footnote at the bottom of the page. Where a marker in the *Glosas Silenses* was a link to an entry in the margin that contained additional data, the footnote marker, usually a number, letter or symbol, in more recent publications makes the user aware of the fact that additional data are provided and can be found in the footnote introduced by that specific marker. The *Nuwe Testament en Psalms: 'n Direkte vertaling* (2014), one of the Afrikaans translations of the New Testament of *The Bible*, frequently enters glosses in footnotes. In this publication Psalm 63 commences as follows:

63 'n Psalm van Dawid. Toe hy in die woestyn^h van Juda was. ...
(63 A psalm of David when he was in the desert^h of Juda)

The word *woestyn* (desert) is followed by the superscript "h". At the bottom of the page a footnote, introduced by the letter "h", accommodates the following gloss:

h woestyn Dit verwys na 'n onbewoonde, onherbergsame gebied waar skape en bokke soms gewei het.
(**h desert** It refers to an uninhabited, barren area where sheep and goats sometimes graze.)

Here the glossed word from the main text is repeated and functions as "lemma" of the gloss. The gloss is provided so that the reader does not have to rely on

external sources for a better understanding of the meaning of the word. This enhances an uninterrupted reading of the text.

The system of footnotes is often used in modern-day publications to accommodate additional data. The user is referred by means of a footnote marker in the main text to the footnote section where the relevant gloss can be found.

Scientific textbooks utilise this form of glossing. Prins (2003), a book from the field of theoretical physics, frequently uses footnotes and often these footnotes contain glosses, as can be seen in Prins (2003: 132) where the text contains the sentence:

In terms of the latter concept, a light-quantum** collides 'like a particle' with an electron ...

The word *light-quantum* gets a footnote marker ("**") and in the footnote section the following gloss explains an aspect of the meaning of this expression:

** At present a light quantum is known as a photon.

This is additional data that the author considers relevant to the readers of the book in order to improve their knowledge of the subject field and to anchor the meaning of the term to the time of writing.

In Håkanson et al. (1977: 42-43), a textbook for Danish learners of Spanish, glosses are placed together in a box (without frames) somewhere on the page. As seen in Figure 2, instead of numbers, letters and signs, the marker introducing the gloss is the word to be explained itself.

Pero hay también mucha genta que no tiene trabajo o que sólo tiene trabajo una parte del año y gana muy poco dinero. Por esto muchos andaluces emigran. ¿Adónde van? Van a otras regiones de España o al extranjero.	<i>el extranjero</i> abroad <i>cada año</i> every year <i>la temporada</i> the period, the season <i>el país</i> the country <i>sobre todo</i> first and foremost <i>Suiza</i> Switzerland <i>Alemania Occidental</i> West Germany <i>e</i> (in front of words beginning with "i") and <i>Inglaterra</i> England
Cada año van unos cien mil españoles a trabajar una temporada a otros países. Van sobre todo a Francia, Suiza, Alemania Occidental e Inglaterra.	<i>varios</i> several <i>enviar</i> send <i>ahorrar</i> save up <i>poner un bar</i> open a bar <i>gasolinera</i> petrol station
Trabajan en fábricas, en hoteles, en restaurantes o en el campo. Muchos se quedan varios años en el extranjero.	
Muchos emigrantes envían dinero a su familia. Otros ahorran dinero para, más tarde, poner un bar, una gasolinera o comprar un piso en España.	

Figure 2: Extract from textbook for Danish learners of Spanish (translated)

Here the gloss box represents a mini-glossary, a kind of transitional form between "pure" glosses and glossaries structured according to the order in which the words to be explained appear in the text.

Glosses are also presented in other forms. *Sintaksis vir eerstejaars* (Du Plessis 1982), a textbook on Afrikaans syntax, uses interlinear glosses, presented between two horizontal parallel lines, to provide additional data in order to expand on a preceding word or statement or to emphasise some aspect of that word or statement. In a section discussing the verb phrase in Afrikaans (Du Plessis 1982: 66) the following is found (translated from the Afrikaans):

If we look at the similarity it is noticeable that just as the phrase NP cannot function without the core N, no VP can exist without the element *verb* (V)

A VP must in all circumstances contain a verb in the basic sentence.

Verbs display group-forming possibilities, this is where the similarities between the VP and the NP stop.

This gloss, given between two vertical lines to separate it from the surrounding text, does not explain the meaning but emphasises a statement made in the preceding sentence. Here the gloss has a didactic purpose, a tradition that can also be traced back to a class of medieval glosses that did not explain the meaning as such, but rather interpreted the text and helped the reader to grasp the message.

Supplementing the references made in Tarp and Gouws (2019) it is interesting to note a further widespread use of glosses. As indicated in Tarp and Gouws (2019) menus often contain glosses to provide a better understanding of a word referring to a specific dish. A dish on the menu of a restaurant is vaguely indicated as *The Chicken*. Because guests will not know what to expect, this entry is immediately followed by the gloss:

Oven-roasted free-range chicken supreme with chargrilled artichokes, blistered tomatoes, fennel, mangetout, stone fruit & chorizo.

Glosses also occur in popular publications like recipe books. *Top 500+ Wenre-septe 2* (Niehaus 2013), a well-known Afrikaans recipe book, uses glosses to complement the recipes. This information is not part of the recipe, and one does not have to follow the advice given there. These glosses are presented in a text box positioned at the end of a recipe and are introduced by the word *Wenk* (=Tip). These tips contain different types of data. The recipe for butternut soup has the tip "Die sop vries goed" (The soup deep-freezes well), whereas the recipe for rusks has the gloss "Jy kan die botter met margarien of olie vervang." (You can substitute the butter with margarine or oil.) The glossing is presented in such a way that the user can ignore the entry but can gain some additional guidance that could be of assistance when making the dish given in the recipe.

The preceding discussion shows that glosses are alive and well — and are living in different types of printed texts. Where dictionaries typically provide a variety of data types in their treatment of a word, a gloss by definition presents a limited treatment, in its origin mostly an explanation of the meaning of a word, but later also of "things and facts" needed in a particular context. This frequent occurrence of glosses demands a term to refer to this kind of activity.

While the term *dictionography*, defined in Section 1, was coined by the authors, *glossography*, the other term included in the title of this article, already exists. It usually refers both to the practice of preparing and inserting glosses into texts and to the compilation of glossaries. Here, however, it refers only to the former, while the compilation of glossaries is included under *dictionography*, since the development from glosses to glossaries to dictionaries can be seen as a continuum with no clear dividing lines.

4. Return of glosses in digital tools

It may seem paradoxical that something as old as the insertion of glosses into books, developed long before the advent of printing and largely ignored by lexicographers for centuries, could find new life and relevance at the time of the greatest technological paradigm shift in human history. Yet that is exactly what is happening today. An uneasy love affair between an ancient technique and modern software is unfolding before our eyes, fusing tradition and innovation. It is making its way into digital reading and writing assistants, where lexicographers and designers are constantly caught between doing business as usual and thinking out of the box to interpret user needs as they manifest in the new digital environment.

Digital reading assistants are defined here as all types of software designed to help readers who have difficulty understanding digital texts. As such, it includes both different types of digital dictionaries and other classes of lexicographical data provided for this purpose. While users may have different needs when reading, the predominant need in such a situation will undoubtedly be an explanation of unfamiliar words and phrases. The default solution in such cases should therefore be as short a definition as possible, or simply an equivalent in the users' native language, so as not to disrupt their reading flow and take their focus away from the text. Additional types of lexicographical data could only be disruptive in this particular situation and would constitute what Gouws and Tarp (2017: 408) describe as "relative data overload". For those users who need more information, they could easily access the relevant data if a "more" button was installed to enable this action.

As Tarp (2022) and others have shown, the problem with using current dictionaries, including those that are integrated into texts and can be activated with a single click, is that traditional dictionary articles are uploaded by default. These articles usually contain too many data that are irrelevant in the specific situation. Even if only the definitions of the senses are listed, there are often

five, ten or more that the user must go through to find the right meaning. Both solutions represent business as usual.

In reality, only one default definition is needed in each lookup, the one that explains the meaning in the concrete context. How to solve this requires thinking outside the box. One solution could be to create AI-based software that can determine the specific meaning of each word or expression, as suggested by Bothma and Gouws (2022). This is undoubtedly a solution of the future. However, as it is likely to take some time to develop such a product, it is interesting to see how some text producers and publishers are starting to take matters into their own hands and provide modern glosses to their users. One such example is the news section of the Danish National Television website.

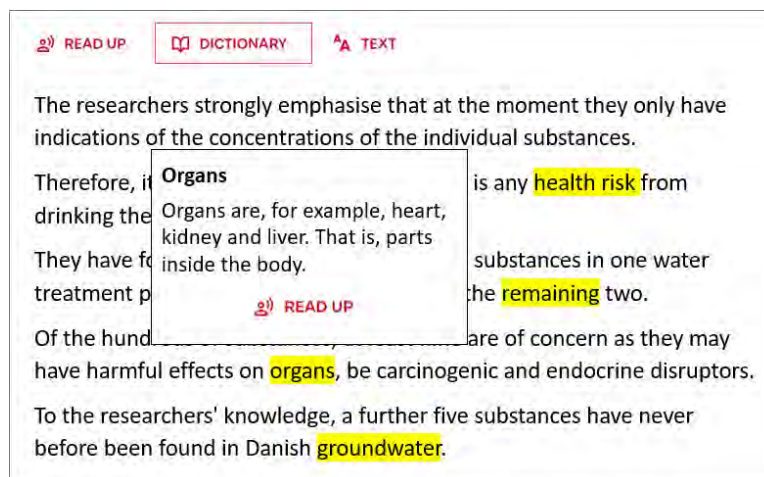
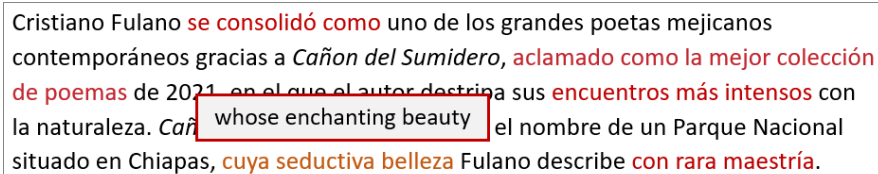


Figure 3: Extract from the Danish National Television website (translated)

Figure 3 shows the somewhat clumsy English machine translation of an article from the Danish website. When the reader activates what is called *Dictionary* (*Word explanation* would be more appropriate), a number of words in the text are highlighted. Clicking on one of these words immediately displays a box with a Collins Cobuild-like definition. The "lemma" *Organs* is completely redundant as the word is already highlighted in the text. If it is removed, all that is left is a traditional gloss, no more, no less. In this way, the five key features of traditional dictionaries mentioned in Section 2 have disappeared altogether. The method is also similar to that of the ancient scribes, in that only those words that are considered difficult for the reader are commented on. What is new under the sun, however, is that the latest technology allows the glosses to appear only when the reader needs them to understand a word, as opposed to the old days when they were always there, sometimes disturbing the reader as an early expression of data overload.

Other publishers, such as those producing learning applications and language courses, also seem to feel a spontaneous need to provide glosses to help their readers. One example is the online Spanish language course for English speakers *News in Slow Spanish*, which offers Spanish news articles to learners at three levels: beginner, intermediate and advanced. In this language course, words and word sequences that are considered to be difficult for users to understand are highlighted in the text and then explained, as in the case of the Danish website mentioned above. The publishers seem to have little knowledge of traditional lexicography, as they do not offer conventional definitions, but direct and presumably automatic translations, which are provided in a rectangular box immediately above the highlighted words and sequences (see Figure 4). Although some of these translations may seem awkward, they work — at least to some extent — and demonstrate the technological possibilities of breathing new life and content into the glosses to meet the perceived needs of learners and other potential users.



Cristiano Fulano **se consolidó como** uno de los grandes poetas mejicanos contemporáneos gracias a *Cañon del Sumidero*, **aclamado como la mejor colección de poemas** de 2021 **en el que el autor destri**na sus **encuentros más intensos** con la naturaleza. *Cañ* **whose enchanting beauty** el nombre de un Parque Nacional situado en Chiapas, **cuya seductiva belleza** Fulano describe **con rara maestría**.

Figure 4: Reproduction of the glossing technique used in *News in Slow Spanish*

Of course, the technique and presentation in Figure 4 could be more elegant, for example by avoiding arbitrary sequences and highlighting only single words and *extended units of meaning*, as defined by Rundell (2018). Since the target audience is learners who may want additional information about the respective words and units, it would also be appropriate to allow them to click through to more detailed data from a lexicographical database, as recommended by Huang and Tarp (2021). Something similar has been suggested by Fuertes-Olivera and Tarp (2020) for lexicography-assisted writing assistants. Be that as it may, Figures 3 and 4 are evidence that the need for glosses in the digital world is real and more or less spontaneously understood by different stakeholders.

The ball is now in the lexicographers' court. It is now up to them to adapt their databases if they want to follow the route of this new normal, while at the same time engaging in interdisciplinary collaboration with designers to achieve the most appropriate presentation. This could also involve the development of special software that allows text producers to highlight the words and extended units of meaning they want to explain in a text and then upload the necessary data from a lexicographical database, as suggested by Nomdedeu-Rull and Tarp (2024).

5. New role of lexicography

As Rundell (2012) has predicted, future lexicographers will not do the same as their colleagues have done until now. In the previous sections, we have already seen how recent technological developments suggest that they should shift their focus from dictionaries to databases containing both new and old types of lexicographical data that can serve various tools, including but not limited to digital dictionaries.

Among the tools already introduced and currently under rapid development are digital writing assistants such as *DeepL Write*, *Ginger*, *Grammarly*, *LanguageTool* and *ProWritingAid*, which are based on AI-powered language models. These writing assistants also use lexicographical data to a greater or lesser extent, both for internal training and communication purposes and for external presentation to their users. As we shall see, some of these represent an entirely new category of lexicographical data, related to specific types of grammatical, orthographic and stylistic problems, as opposed to the more conventional data associated with different lemmas. As such, they cannot be stored in traditional lemma-based databases, but must be accommodated separately in problem-oriented databases directly related to the language model. The methodology for their elaboration is also different. Figure 5 provides an example of this class of data.

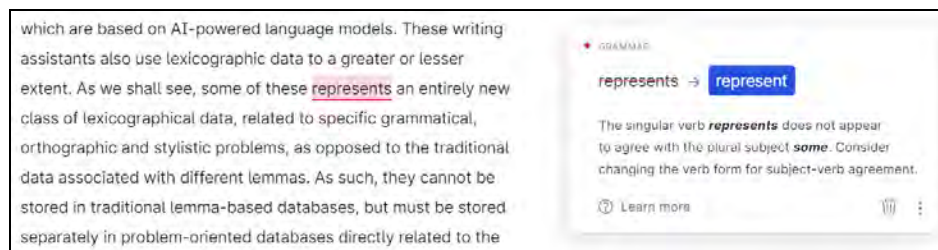


Figure 5: Pop-up window with suggestion and comment in *Grammarly*

When a writing assistant, in this case *Grammarly*, detects a potential problem in a piece of writing, it draws the user's attention to it by underlining it and giving it a colour that varies according to the severity of the problem. If the user clicks on the highlighted area, a pop-up window immediately opens with an alternative suggestion and an explanatory comment or annotation. With the exception of *DeepL Write* (at least for now), the other three writing assistants mentioned above employ similar techniques to serve their users.

The suggestions in all of these tools are automatically generated by the underlying AI-driven language model. The annotations, by contrast, are the result of human expertise and an innovative way of writing about the words or

vocabulary in the digital environment. As such, they represent a new category of lexicographical data that opens up a whole new field of activity for lexicographers, who, with their time-honoured user-centred approach, seem to be the most appropriate experts to add a communicative task of this sensitive nature to their classical repertoire. This observation is reinforced by the fact that the annotations vary considerably from one writing assistant to another and do not always seem to be of the necessary quality to serve the user group adequately.

Annotations like the one shown in Figure 5 have many similarities with the classical scribes' glosses and the way they prepared and inserted them into texts. As a term with Greek roots, *lexicography* does not mean *describing* vocabulary, but *writing* about it. Just like the ancient scribes, their modern colleagues do not aim to describe the whole vocabulary, but only to write about a part of it. Digital-age scribes, like their predecessors, write about only those words that they think might be a problem for their readers. In this sense, they produce modern glosses, adapted to the new reality, to be inserted into texts to assist their users.

The main difference between the traditional glosses discussed in the previous sections and glosses like the one in Figure 5 is that the latter aim to assist writers with text production problems and therefore contain a certain element of recommendation or instruction, whereas the former focus on readers with text reception problems, as well as learners of languages and specific subjects, and are therefore more explanatory. Another difference is that although modern glosses appear in texts related to specific words, in most cases they are not elaborated specifically in relation to these words, but to classes of problems involving several or even many different words. This is why they have to be stored differently from traditional lexicographical data.

In conventional lexicography, including digital lexicography, the work of lexicographers consists not only in selecting lemmas and writing dictionary articles, but also in preparing the necessary empirical material, such as special corpora. The same applies to problem-oriented digital glosses. Before the glosses can be written, the language model has to be trained to identify problems and suggest alternative solutions. This requires different types of empirical training material, the preparation of which can benefit from the input and active participation of lexicographers. For example, in addition to a "traditional" corpus, it may involve the compilation of a special set of parallel corpora from which the language model can learn to distinguish between right and wrong. It may also involve the preparation of validation material to test its performance and determine whether it should be further trained to serve the intended user group.

Once the language model has been trained and has reached an acceptable level of performance, it will automatically generate internal codes for each class of problems identified. These codes will number in the thousands and will be the starting point for producing explanatory glosses. Figure 6 shows five such codes taken from the AI-powered language model supporting a Spanish writing assistant for foreign learners currently under development.


```
$TRANSFORM_INFLEXION_verb-subjunctive-imperfect-third--person--plural_verb-indicative-present-first--person--singular
$TRANSFORM_INFLEXION_verb-subjunctive-imperfect_verb-indicative-present-first--person--singular
$TRANSFORM_INFLEXION_verb-subjunctive-imperfect-second--person--singular_verb-indicative-present-first--person--singular
$TRANSFORM_INFLEXION_verb-subjunctive-imperfect-first--person--plural_verb-indicative-present-first--person--singular
$TRANSFORM_INFLEXION_verb-subjunctive-imperfect-second--person--plural_verb-indicative-present-first--person--singular
```

Figure 6: Internally generated codes in AI-driven language model

It is now the task of the lexicographers to write short glosses and, if necessary, additional explanations, and then to link them to the relevant codes using specially designed software. One of these glosses will then appear in a pop-up window whenever a user clicks on a highlighted area corresponding to its particular type of error, regardless of the specific word to which it relates. Figure 7 shows an example of how this works when a learner has mistakenly used a Spanish verb in the indicative instead of the subjunctive. As can be seen, the writing assistant does not provide a general, one-size-fits-all gloss, but individualises it by inserting the two specific words in question — the appropriate one and the inappropriate one — into the gloss. The general gloss, which serves as a frame for the individualised gloss shown in the figure, can thus be used to comment on thousands of different Spanish verbs and explain the suggested corrections when the same class of grammatical problem occurs.

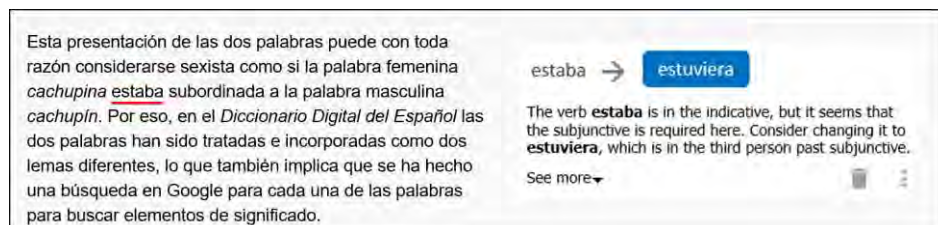


Figure 7: Pop-up window with suggestion and gloss

Working with high-tech tools such as writing assistants is certainly a new and relevant task that more lexicographers will have to take on. However, it requires an open mind and the ability to engage in interdisciplinary collaboration with computer scientists, programmers and designers, since these products are not lexicographical as such, but products with a lexicographical component. It also implies a willingness and ability to embrace the latest technologies, such as the much-discussed chatbots, which will undoubtedly revolutionise lexicographical production in the coming years.

In this respect, Huete-García and Tarp (2024) report that they use chatbots for three different tasks in the aforementioned Spanish writing assistant project. For the first task, which is for internal training purposes only, there is no need to revise the data produced by the chatbot. However, for the second task,

also for internal use, the expertise of the lexicographer is essential, as it requires 100% correct data, while for the third task, which involves external data to be presented directly to the end user, human knowledge and creativity are indispensable. The two researchers show how the three tasks represent three completely different types of relationship between the lexicographers and the chatbot and how this AI-based technology significantly increases productivity without reducing its human counterparts to irrelevant extras. On the contrary, it may require even more knowledge, expertise and creativity from lexicographers, both to give relevant and precise instructions that can guide the chatbot to produce data of the desired type and quality, and to evaluate and build on these data.

The quintessence of all this, together with the reflections in the previous sections, is the main reason why the concept of *lexicography* needs to be redefined, with a definition adapted to actual practice and brought up to date.

6. New definition of the term "lexicography"

As indicated in this paper the term *lexicography* was introduced much later than many products that today can rightfully be regarded as lexicographical work. In the course of time the term *lexicography* was predominantly used to refer to the practice of making dictionaries. Although the writing of dictionaries represents a form of writing about words and thus adheres to the original meaning of the word *lexicography*, dictionaries, as they are known today, are not the only products that fall within the scope of lexicography. The work done by the Greek scribes when they inserted glosses into texts was an early form of lexicography and although not identified as such at that time the current understanding of lexicography acknowledges and includes those products.

The introduction of the printing press was a breakthrough event in many ways — also in lexicography. Although the copying of manuscripts came to an end, the use of glosses to present contextualised data continued, cf. Tarp and Gouws (2019: 253). This use of glosses covered a spectrum of text types and introduced a significant lexicographic feature into these products. However, the insertion of glosses into texts was no longer a specialised craft carried out by scribes and literary monks, but a practice with a wide variety of expressions, in which a large number of different authors and editors of texts engaged. This dilution of the traditional craft is probably the reason why lexicographical research has for a long time overlooked this activity, or simply regarded it as a long-gone precursor of "proper" lexicography, which it associates exclusively with dictionaries and similar reference works.

The digital age, with its diverse text types, has radically changed this situation, allowing for an even more productive use of glosses, which are now also integrated into digital texts, where they can be activated by touching or clicking on the screen. The main players in this new era are publishers and editors of digital books, websites, learning courses, writing assistants, etc. These publishers and editors realise that the appropriate use of glosses increases the quality, use-

fulness and competitiveness of their products, so any qualified and up-to-date lexicographical support would most likely be more than welcome.

The current use of glosses is clearly comparable to the original way of glossing, where the gloss presents data that can help the user in understanding or using the word in an appropriate way. The gloss is used as an item with a lexicographical nature in an extra-lexicographical environment. Bearing in mind that glosses were originally also not used in lexicographic environments, but that the occurrence of glosses can be regarded as a major step in the beginning of lexicography, it is important that the current use of glosses should also be included within the scope of the term *lexicography*.

The proposal was made in Section 1 of this paper that the term *dictionography* should be introduced to refer to the sub-field of lexicography concerned with the planning, compilation and presentation of reference works like dictionaries, glossaries, thesauri, lexicons and encyclopaedias, and with the development of theories about them. *Dictionography* is the sub-field of lexicography concerned with the practice and theory of dictionaries, interpreted in the broad sense of the word.

Glosses did receive some attention in theoretical lexicography but with a focus only on their occurrence in dictionaries. This confirms the status of these glosses as lexicographically relevant, but it fails to include the more comprehensive use of glosses in the lexicographical discussion.

Section 3 of this article proposed the use of the term *glossography* to refer to the preparation and insertion of glosses into texts. Glosses resemble certain types of lexicographic items, and they are used to present specific types of data to readers to enhance their understanding of the meaning or use of a given word or expression. The extremely productive occurrences of glosses and their advanced and sophisticated use in the digital environment demand the recognition of glossography as a formal sub-field of lexicography. Glossography is not only the preparation and insertion of glosses into texts, but also the preceding preparation of empirical material to present an entirely new class of lexicographical data that can be used in, among other, writing assistants for both internal training and external presentation. Glossography is not only concerned with the practice of an expanded way of glossing, but also with the underlying theory. The lexicographical nature of glossing and glosses may never be underestimated. Consequently, glossography should also be regarded as a sub-field of lexicography.

The massive changes in lexicography during the digital era, the numerous innovative developments in the field of reference sources, the increased use of new kinds of lexicographical data in dictionary-external environments demand a new definition of *lexicography*. Such a definition has to be inclusive by negotiating the distinction between dictionography and glossography as two sub-fields of lexicography. It has to reflect the past, take cognizance of the present and make provision for future developments. Different types of dictionaries, other reference sources, glosses and other types of lexicographical data as well

as the underlying theories need to be covered by this definition.

The following is presented as a redefinition of *lexicography*:

Lexicography is the discipline that deals with dictionaries, other reference works, and glosses, all of which are designed to be consulted in order to retrieve information about words, things or facts.

It can be argued that as a discipline lexicography has a practical and a theoretical component, and that these two components have two main sub-fields, namely *dictionography*, with its focus on dictionaries and related products, and *glossography*, with its focus on glosses. From a historical and contemporary perspective, dictionography and glossography are concerned with handwritten, printed and digital products.

7. Conclusions

In this article we have argued that there is a need to redefine the term *lexicography*, taking into account both current and historical facts. The challenge lies in the largely overlooked glosses. The traditional focus on dictionaries and similar works has not been fair to these glosses, which have been orbiting dictionaries like small planets for more than two thousand years.

Originally, dictionaries evolved from glosses, which did not disappear but continued until the advent of printing, being inserted into existing texts by scribes and other learned people who wanted to explain difficult or obsolete words to readers. As we have seen, even during the long dominance of the printed book, authors and others kept glosses alive and gave them a myriad of new expressions.

The fact that the use of glosses in this new era was no longer a craft practised by a few scholars may, as has been suggested, be the reason why they were overlooked when seventeenth-century European lexicographers began to define their discipline as one concerned solely with dictionaries and related works. Be that as it may, the fact is that with the advent of digital technologies, and in particular artificial intelligence, glosses have been given a new lease of life that requires us to rethink history and include them in a redefinition of lexicography as a discipline.

In this article, therefore, we have redefined *lexicography* as a two-pronged discipline, concerned on the one hand with dictionaries and related works, and on the other with glosses, both traditional and new types that have emerged in the digital world. We have called these two sub-fields *dictionography* and *glossography* respectively.

We are convinced that all this is not an empty academic exercise, but a necessary foundation for a development that is already underway and is bound to accelerate in the near future. Michael Rundell's prediction that by the year 3000 lexicographers will not be doing the same thing as before is already becoming a reality.

As for dictionography, the authors of this article were recently involved in an experiment using chatbots and other digital techniques to write almost 3000 dictionary articles in a single day. This in itself not only increases productivity, but also radically changes the role of the lexicographer.

As for glossography, at least some of the glosses that can be used to advantage in digital texts, as we have seen in Figure 3, are largely the same type of data (definitions) that already exist in some lexicographical databases. This more than suggests that lexicographers should move away from primarily focusing on dictionaries when planning new projects, and instead focus on multi-purpose lexicographical databases that can both feed dictionaries and upload data to various digital software, such as writing and reading assistants. And even if it is not always the same type of data, lexicographers can plan from the outset to compile databases containing both types of data to cover multiple types of digital products. But we have also seen above how certain types of software, such as writing assistants, require entirely new types of glosses, which also require lexicographical expertise to be of high quality. These new types of glosses are problem-oriented and therefore cannot be stored in traditional lemma-based databases, but in a new class of problem-oriented databases. Today's lexicographers need to prepare themselves for these tasks.

The new technological breakthroughs will undoubtedly greatly increase the productivity of the traditional lexicographical compilation process. And while there will still be a need for highly skilled lexicographers to ensure quality, it is unlikely that there will be as many as there are today. However, with the redefinition of lexicography as a discipline that includes modern glosses, a development already observed in digital texts, a whole new field of work is opening up for well-trained lexicographers. This new field will consist not only of revision and routine, but also of creative tasks that require an open mind and a willingness to break new ground.

Rapid technological development does not mean the end of lexicography, as some have hastily suggested. It does mean, however, that the discipline cannot continue as before, but must adapt to new realities. Hopefully, a timely redefinition of the subject matter of lexicography, as proposed in this article, can contribute to this change in direction.

Bibliography

A. Dictionaries

Diccionario enciclopédico hispano-americano de literatura, ciencias y artes. 1887–1910. Barcelona: Montaner y Simón.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London: Routledge.

Pearsall, J. (Ed.). 1998. *The New Oxford Dictionary of English*. New York: Oxford University Press.

Mish, F.C. (Ed.). 1987. *Webster's Ninth New Collegiate Dictionary*. Springfield, Massachusetts: Merriam-Webster.

Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English*. Harlow: Longman.

Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer and W. Wolski (Eds.). 2010–2020. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. Berlin: De Gruyter.

B. Digital tools

DeepL Write (beta version): <https://www.deepl.com/write>

Ginger: <https://www.gingersoftware.com/>

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

LanguageTool: <https://languagetool.org>

News in Slow Spanish. 2023. <https://www.newsinslowspanish.com>

ProWritingAid: <https://prowritingaid.com/>

C. Other literature

d'Alembert, J.L.R. 1754. Dictionnaire. Diderot, D. and J.L.R. d'Alembert (Eds.). 1754. *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers*. Tome IV: 958-969. Paris: André le Breton, Michel-Antoine David, Laurent Durand and Antoine-Claude Briasson.

Bergenholtz, H. and R.H. Gouws. 2012. What is Lexicography? *Lexikos* 22: 31-42.

Bothma, T.J.D. and R.H. Gouws. 2022. Information Needs and Contextualization in the Consultation Process of Dictionaries that Are Linked to e-Texts. *Lexikos* 32(2): 53-81.

Bybelgenootskap van Suid-Afrika. 2014. *Nuwe Testament en Psalms. 'n Direkte vertaling*. Bellville: Bybelgenootskap van Suid-Afrika.

Du Plessis, H. 1982. *Sintaksis vir eerstejaars*. Pretoria: Academica.

Fuertes-Olivera, P.A. and S. Tarp. 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.

Gouws, R.H. and S. Tarp. 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

Grefenstette, G. 1998. The Future of Linguistics and Lexicographers: Will There Be Lexicographers in the Year 3000? Fontenelle, T., P. Hilgsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *Proceedings of the Eighth EURALEX International Congress in Liège, Belgium*: 25-41. Liège: English and Dutch Departments, University of Liège.

Hanks, P. 2013. Lexicography from Earliest Times to the Present. Allan, K. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.

Huang, F. and S. Tarp. 2021. Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvements. *Lexikos* 31(1): 68-92.

Huete-García, Á. and S. Tarp. 2024. Training an AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. (to appear)

Håkanson, U., J. Masoliver, H.L. Beeck and J. Jensen. 1977. *ESO ES 1. Spansk for begyndere. Tekstbog*. First Edition. Copenhagen: Grafisk Forlag. (Second Edition: PRAXIS).

Kennedy, A. 2019. A Brief History of the Footnote — From the Middle Ages to Today. Quetext blog: <https://www.quetext.com/blog/a-brief-history-of-the-footnote>

- Landau, S.** 1984. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer*. Cambridge: Cambridge University Press.
- Niehaus, C.** 2013. *Top 500+ Wenreseppe 2*. Cape Town: Human & Rousseau.
- Nomdedeu-Rull, A. and S. Tarp.** 2024. *Introducción a la lexicografía en español: funciones y aplicaciones*. London: Routledge.
- Prins, J.F.** 2003. *Superconduction at Room Temperature without Cooper Pairs*. Craighall: Mulberry MoonDesign.
- Ruiz-Asencio, J.M., I. Ruiz-Albi and M. Herrero-Jiménez.** 2020. *Las Glosas Silenses. Estudio crítico y edición facsímil. Versión castellana del Penitencial*. Burgos: Instituto Castellano y Leonés de la Lengua.
- Rundell, M.** 2012. The Road to Automated Lexicography: An Editor's Viewpoint. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 15-30. Oxford: Oxford University Press.
- Rundell, M.** 2018. Searching for Extended Units of Meaning — and What To Do When You Find Them. *Lexicography* 5: 5-21.
- Stathi, E.** 2006. Greek Lexicography, Classical. Brown, K. (Ed.). 2006. *Encyclopedia of Language and Linguistics*. Vol. 5: 145-146. Second Edition. Oxford: Elsevier.
- Tarp, S.** 2018. Lexicography as an Independent Science. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 19-33. London/New York: Routledge.
- Tarp, S.** 2020. Integrated Writing Assistants and their Possible Consequences for Foreign-Language Writing and Learning. Bocanegra-Valle, A. (Ed.). 2020. *Applied Linguistics and Knowledge Transfer: Employability, Internationalization and Social Challenges*: 53-76. Bern: Peter Lang.
- Tarp, S.** 2022. A Lexicographical Perspective to Intentional and Incidental Learning: Approaching an Old Question from a New Angle. *Lexikos* 32(2): 203-222.
- Tarp, S. and R.H. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.
- Wiegand, H.E.** 1998. *Wörterbuchforschung*. Berlin: De Gruyter.
- Yong, H. and J. Peng.** 2022. *A Sociolinguistic History of British English Lexicography*. London/New York: Routledge.

Verifying the Terminology of Arabic Rhetorical Tradition: The Case of *'Iltifāt*

Emad Abdul-Latif, *Arabic Department, Faculty of Arts and Sciences,
Qatar University, Doha, Qatar (emad.abdullatif@qu.edu.qa)*

Abstract: The terminology of Arabic *Balāgha* suffers various problems, such as having multiple terms that refer to one concept and multiple concepts that have one term. These problems cause vagueness and misunderstandings in many modern and contemporary studies. This article aims to decrease the possibility of both ambiguity and misunderstanding by suggesting a comprehensive methodology to verify Arabic rhetorical terminology. This methodology consists of two procedures: (1) identifying and analyzing the various concepts that a one-term refers to, and (2) identifying and analyzing the different terms that refer to one concept. These procedures are applied to the term *'iltifāt* and the concept of 'shifting among first, second, and third pronouns'. The Arab classical rhetoricians used *'iltifāt* to refer to thirteen different figures, styles, and features. Similarly, they used seven different terms for one concept. The reasons behind this terminology disorder are investigated and the necessity of a contextual dictionary of Arabic *Balāgha* is argued for.

Keywords: *'ILTIFĀT, BALĀGHA, GRAMMATICAL SHIFT, ARABIC TRADITION, TERMINOLOGY, RHETORIC*

Opsomming: Die verifikasie van die terminologie van die Arabiese retoriese tradisie: Die geval van *'Iltifāt*. Die terminologie van die Arabiese *Balāgha* openbaar verskeie probleme, soos om veelvoudige terme te bevat wat na een konsep verwys en om veelvoudige konsepte te bevat wat net een term het. Hierdie probleme veroorsaak vaagheid en wanopvattinge in baie moderne en kontemporêre studies. In hierdie artikel word gepoog om moontlike dubbelsinnigheid asook wanopvattinge te verminder deur middel van 'n omvattende metodologie om die Arabiese retoriese terminologie te verifieer. Hierdie metodologie bestaan uit twee stappe: (1) die identifisering en analisering van die verskillende konsepte waarna 'n enkelterm verwys, en (2) die identifisering en analisering van die verskillende terme wat na een konsep verwys. Hierdie stappe word op die term *'iltifāt* en op die konsep van 'wisseling tussen eerste, tweede, en derde voornaamwoorde' toegepas. Die Arabiese klassieke retorici het *'iltifāt* gebruik om na dertien verskillende figure, style, en kenmerke te verwys. Op 'n soortgelyke wyse, het hulle sewe verskillende terme vir een konsep gebruik. Die redes vir hierdie terminologiese wanorde word ondersoek en die noodsaaklikheid van 'n kontekstuele woordeboek vir die Arabiese *Balāgha* word bepleit.

Sleutelwoorde: *'ILTIFĀT, BALĀGHA, GRAMMATIKALE VERSKUIWING, ARABIESE TRADISIE, TERMINOLOGIE, RETORIEK*

1. Introduction

'iltifāt is an Arabic rhetorical term that commonly refers to moving from 1st, 2nd, and 3rd personal pronouns to another while referring to the same person. Within one decade, from 1988 to 1998, three researchers counted the number of times *'iltifāt* appears in the Qur'ān. al-Haysharī (1988) estimates the number of *'iltifāt* in the Qur'ān at 157, while Abdel-Haleem (1992) states that the number is 400. Hassan Tabl's (1998) raises that number to 790. This discrepancy in the number of *'iltifāt* in the Qur'ān results from the different concepts the three researchers have about that term. al-Haysharī (1988) confines the meaning of *'iltifāt* to the shifts between second and third-person pronouns. On the other hand, Tabl (1998) expands that definition to include the shifts among addressees, numbers, genders, and verb forms (past, present, and imperative). Abdel-Haleem (1991) adopts a middle ground where he defines *'iltifāt* as shifts between pronouns, numbers, genders and tenses.

The discrepancy in definitions of Arabic rhetorical terminology is not confined to *'iltifāt* alone; in fact, contemporary researchers in *balāgha*/Rhetoric¹ face many problems with most of its terms, such as *I'sti'āra*. This article explores the roots of terminological problems in Arabic rhetorical heritage and suggests a methodology to standardize terms to minimize the side effects of terminological inconsistency on contemporary research. A sample analysis will focus on the term *'iltifāt* and one of its concepts and attempt an explanation of the vast discrepancy among the concepts referred to by this term. There are religious, cultural, and social reasons for this discrepancy. To identify them, the Arabic writings from the second to the tenth centuries AH are surveyed. Every mention of *'iltifāt* in various disciplines is analysed, encompassing *balāgha*, literary studies, philology and linguistics, Qur'ānic studies, the science of jurisprudence, and theology. Five aspects which contribute to the problems of Arabic Rhetorical terms could be identified through the mentioned surveys and analyses, namely (1) the effect of religion; (2) replacing the term with another lexical equivalent; (3) diversity of disciplines using the term; (4) misreadings of classical writings; and (5) the reverse direction of establishing a terminology. These five aspects are discussed below.

1.1.1 The effect of religion

The relationship between Arabic *Balāgha* and Islam is very close and complex. Qur'ān studies provided influential motivations, concepts, and terminologies for Arabic *Balāgha*, but the influence of the Islamic religion has not always been positive, particularly on the terminology of *balāgha*. There are some examples where well-established rhetorical terms have been rejected for religious reasons. There are other examples where religion triggered rhetoricians to prefer one term to another or to drop some terms altogether.

For instance, Jamāl al-Dīn al-ʿAndalusī (1987: 166) called for the abandonment of the term *ʿiltifāt* for religious reasons. He (ibid.) says, "This point is illustrated in the Qurʾānic verse where God says: '... whom the Prophet wishes to wed- this is only for you [Prophet] and not the rest of the believers ...' [The Qurʾān 33:50.] This example should not be called *ʿiltifāt* because it [the term *ʿiltifāt*] cannot be used to talk about God" (al-ʿAndalusī 1987: 166). He builds this opinion on the assumption that the general and academic meanings of the word *ʿiltifāt* are identical in shaking the head to the right and left. Because God does not shake his head, he believes we must not use the term *ʿiltifāt* to describe God's Words. According to al-ʿAndalusī (1987), because God does not do this movement (*ʿiltifāt*), we must not attribute this description to His word.

The overlap between the general and academic meaning of words has resulted in the absence of a term denoting pronoun shifts in the Holy Qurʾān. Although this figure of speech is there and is recurrent in two Qurʾānic research areas: the first is Qurʾānic lections (different modes of reading or recitation) as in the works of Abū ʿAlī al-Fārisī, Ibn Khālawayh, al-ʿAzharī, al-ʿAkbarī, and others. The second is Qurʾānic exegesis, which is apparent in the commentaries of al-Ṭabarī, al-Ṭūsī, al-Fakhr al-Rāzī, Ibn al-Jawzī, al-Khāzin, etc. Confusing the academic and general meanings of the word *ʿiltifāt* reflects the lack of awareness that the academic and technical meaning diverges from the word's general meaning.

1.1.2 Replacing the term with another lexical equivalent

The terms that refer to pronoun shifts in Arabic include *ʿiltifāt*, *ʿinṣirāf*, and *ṣarf*. These three terms denote a 'change of direction'. Ibn Manẓūr (n.d.: 379-380), in his *Lisān al-ʿArab*, cites the usage of these verbs: "*lafata/ṣarafa his face from people (he turned it away); talaffata/ʿiltafata to something (he turned his face to it); laft and ṣarf are synonyms ... lafattu someone from his opinion (I dissuaded him/her; made him/her change his/her opinion)*". *Sarf* and *ʿinṣirāf* are two nouns derived from the same root (Ṣ-R-F). These three terms refer to the same concept based on their linguistic proximity. This interchangeability of terms hinders terminological unification achieved by having one specific referent for a particular concept.

1.1.3 The diversity of disciplines employing *ʿiltifāt*

Arabic *Balāgha* is a meeting point for different disciplines with their fields of study, methodologies, and goals, including linguistics, Quranic studies, Islamic theology, logic, literary studies, grammar, etc. Each discipline has left its impact on Arabic *Balāgha*, including its terminology; *ʿiltifāt* is not an exception.

For example, the concept of *ʿiltifāt* in literary studies differs from that in Qurʾānic exegesis. Under the influence of Ibn al-Muʿtazz, literary critics dealt with *ʿiltifāt* as a two sided figure that includes shifts in topic and pronouns. On

the other hand, scholars in Islamic studies confined the concept to the syntactic shifts between pronouns in most of their works.

1.1.4 Misreadings of classical writings

Around the middle of the fifth century AH, 'Abd al-Qāhir al-Jurjānī (1984: 457) argued that some rhetoricians had misunderstood rhetorical writings that had been produced before their time. Although he did not cite examples to support his claim, we can find clear evidence that classical Arabic works on *'iltifāt* have been misunderstood. This misunderstanding has resulted in basic terminological dilemmas and can be proved by studying the effect of confusing the contents of three consecutive chapters in Ibn al-Mu'tazz's book *al-Badī'* (Figures of Speech, 1967).

At the turn of the fifth century AH, Ibn Rashīq (1972b: 45-46) wrote: "This section will talk about *'iltifāt* which is known as *'i'tirāḍ* (appositive) by some writers, while others call it *'istidrāk* (disjunctive) ...". Ibn Rashīq (ibid.) regards three completely different terms with distinctive concepts as synonyms, namely *'iltifāt*, *'i'tirāḍ* and *'istidrāk*. He unites them in one definition that is true of only one of them, i.e., *'i'tirāḍ*, which refers to embedding one statement in another. He cites examples from the chapter on *'i'tirāḍ* from *al-Badī'*. Ibn Rashīq (ibid.) defends this confusion of the terms (*'i'tirāḍ* and *'iltifāt*) by claiming, "All people consider them the same." Not only that, he illustrates *'iltifāt* with examples that describe a different figure, i.e., *tatmīm* (completion), saying, "These examples are closer to *'iltifāt* (than to *tatmīm*)." He goes on and illustrates *'iltifāt* with examples that belong to a fourth figure, *'istidrāk*.

This chaotic confusion of terms occurred even before Ibn Rashīq: al-Ḥātimī (1979: 45-46) equalizes *'iltifāt* and *'i'tirāḍ*. al-Bāqillānī (1978: 32) does not explain any concept or provide any definition for *'iltifāt*. Instead, he cites some examples from different extracts that illustrate four other figures of speech. Perhaps his confusion results from the overlapping comments on each concept. Most of the examples al-Ḥātimī, and al-Bāqillānī provide are taken from *al-Badī'*. They had probably read that book and confused the three confusing chapters. al-Sijlīmāsī (1980: 442) realized this confusion of terms made by Ibn Rashīq, and he did not hesitate to call it a mistake. We can picture what happened with those three rhetoricians who fell into that error: they, and others, saw three consecutive chapters in *al-Badī'* on *'iltifāt*, *'i'tirāḍ*, and *'istidrāk* with their examples. Then, the terms and their illustrations were confused, and the limits between their concepts vanished.

1.1.5 The reverse direction of establishing a terminology

There is a terminological rule indicating that formulating an idea precedes the creation of a term that refers to it (al-Trabelsī 1991). This rule assumes the

existence of a one-way process where a concept is discovered, followed by finding a name for it. It does not recognize the presence of a reverse movement that starts with creating a name for something unknown and then trying to create or discover that thing. This rule is based on another rule that reads: "The principal function of the terminology system is that it refers to the concepts system. For a terminology to be useful, it must reflect the emergence of the concepts" (Ibn Tālib 1989: 97). However, Arabic *Balāgha* contradicts this rule in some cases where a name (term) is coined before researchers define the concept.

This reversed process causes terminology inconsistency, which is what happened with the term *'iltifāt*. Classical *Balāgha* books repeatedly cite an anecdote involving al-'Aṣma'ī as the first occurrence of the term, which caused many terminological problems. al-Ḥatīmī narrates this story (1979: 388):

"Iṣḥāq ibn 'Ibrāhīm al-Mawsilī once said that al-'Aṣma'ī asked him: 'Do you know the *'iltifāt* of Jarīr?' I said: 'What are they?' He chanted:

Did you forget when Sulaima bade us farewell?

With a branch of Arabian balsam in her hand!

Oh, blessed is Arabian balsam!

Don't you see him talking to his companion before turning (*'iltafata*) to Arabian balsam and praying for it?!"

Ibn Rashīq (1972a) also cites this story in *al-'Umda* with a few changes in the text but many differences in the chain of transmitters: In one, al-'Aṣma'ī was talking to Iṣḥāq al-Mawsilī; in the other, it was al-Ṣūlī.

Moreover, al-Ḥatīmī (1979) says it was Muḥammad ibn Yaḥyā al-Ṣūlī narrating on the authority of Yaḥyā ibn 'Alī who was, in turn, narrating from his father on the power of Iṣḥāq al-Mawsilī. al-Ḥatīmī's version is the oldest, and al-Bāqillānī (1978) cites it in his *I'jāz al-Qur'ān* (Inimitability of the Qur'ān). With all its three versions, the story includes the term *'iltifāt* but without a clear concept or definition. al-'Aṣma'ī's comment on the lines of verse leaves it open to many interpretations, and this is what exactly happened later to some rhetoricians. For example, al-'Askarī (1984: 431) defines *'iltifāt* as follows: "It is when the speaker talks about something, and when you think he wants to talk about something else, he goes back to it and talks about it, adding new information or other words." al-Ḥatīmī understood the same comment about *'iltifāt* to mean *'i'tirāḍ* (the appositive). At the same time, Ibn Rashīq (1972a) mentions two definitions for *'iltifāt* from this story: One synonymous with *tatmīm* (completion) and the other *pronoun shift* as came to be the established concept later. However, he does not attach any of the two definitions to what is meant in this story. Interestingly, none of these two concepts is present in this anecdote, nor does it illustrate the use of *'i'tirāḍ*. All these opinions result from a confused reading of the cited examples. Perhaps Abū Hilāl (al-'Askarī) was more discerning when he attempted a concept that pertained to this story in particular

and said it was one of two concepts, the second of which he took from Qudamah ibn Ja'far. This illustrates that coining a term for an unspecified concept at an earlier stage in the history of scholarship has caused some terminological problems that continued into the present. These problems require some effort to standardize the terminology of Arabic *Balāgha*. Following is a suggested methodology to achieve that.

1.2 Suggested methodology to standardize the terminology of Arabic rhetoric

There is a preliminary stage before applying the proposed methodology, which includes the following steps:

Identifying the disciplines and research fields that use the term

The following disciplines use the term *'iltifāt*:

- Qur'ānic studies: Qur'ān meaning, inimitability and miraculousness, exegesis and lections (variant readings/recitations).
- Literary Studies.
- Arabic linguistics and philology.
- General *Balāgha*.
- Dictionaries of general and specialized terms.

Documentation

This involves the collection of written works on *'iltifāt* in Arabic rhetorical legacy, which is a challenging process due to the lengthy period over which these works have been compiled, the various disciplines that tackled it, and the unorganised fashion in which these books discuss this term (al-Qāsimī 1985).

Semantic identification of the terms used in terminological standardization

Establishing a term rests on three elements:

- A reality: which is the object that needs a name. It can be abstract or concrete, singular or plural.
- A concept: "an ideational representation of something (concrete or abstract) or a class of things that share some characteristics and it is referred to with a term or symbol" (al-Qāsimī 1985: 213).

- A term: is "any linguistic unit that refers to a meaning, and it can consist of one word (a simple term) or many words (a complex term), and it gives a name to one specific concept within a certain field or discipline" (al-Qāsimī 1985: 213).

The methodology that proposed and applied in this article to standardize the terminology of *Balāgha* consists of two procedures. The first is building a history of the term *'iltifāt* and the various concepts connected to it over time. The second is following the different terms used to refer to a particular rhetorical concept.

2.1 Terminological standardization of *'iltifāt*

2.1.1 Defining the concepts referred to by *'iltifāt*

Table 1 below shows the many concepts referred to as *'iltifāt*. The table consists of four columns: the first includes the name of the rhetorician who used the term, the year he died, and the book title and page number where the term is mentioned (where applicable). The second column has a number referring to the number of concepts attached to the term; these numbers are chronologically ordered. The third column defines the concept. Repetition of the number in column two means repeating the same concept even with different words. Some rhetoricians did not give definitions to their concepts, but I deduced them from their examples and comments on them.

Table 1: Concepts attached to the term *'iltifāt* in Arabic rhetoric

Author/date of his death/book/page	Number of the concept	Definition	Discipline
al-'Aṣma'ī (died in 831).	1	He mentioned the term and a line of verse without giving a definition. The best definition, given by Abū Hilāl al-'Askarī, 1984: 431, reads: "It [<i>'iltifāt</i>] is when the speaker talks about something, and when you think he wants to talk about something else, he goes back to it and talks about it, adding new information or other words."	Commentary on poetry

Ibn al-Muʿtazz (died in 908), <i>al-Badīʿ</i> , 1967: 59.	2 3	Shifting from using the 2nd person pronouns to the 3rd person pronouns and vice versa, and the like. Moving from one meaning/topic to another.	Criticism of poetry
Qudamah (died in 948).	4	Leaving one meaning and going back to emphasize or justify a previous one.	Criticism of poetry
al-Khuwārazmī (died in 998), <i>Mafātīḥ al-ʿUlūm</i> (<i>Key to Sciences</i>), 1924: 60.	2		A dictionary of terms
al-Ḥātimī (died in 998) <i>Ḥilyat al-Muḥaḍḍarah</i> , 1979: 157.	5	Embedding a phrase inside another.	Rhetoric
Ibn Jinnī (died in 1002), <i>al-Faṣr</i> , 2004, vol. 1: 210-211.	1		Commentary on poetry
Abū Hilāl al-ʿAskarī (died in 1005), <i>al-Ṣināʿatayn</i> (<i>The Two Genres</i>), 1984: 438.	4+1		Criticism of poetry and prose
al-Tannīsī (died in 1003), <i>al-Munīf</i> , 1994: 166.	2		Criticism of poetry
al-Bāqillānī (died in 1013), <i>ʿIjāz al-Qurʾān</i> (<i>Inimitability of the Qurʾān</i>), 1978: 30.	1+5		Inimitability of the Qurʾān

al-Marzūqī (died in 1030), <i>Sharḥ al-Ḥamasah</i> , 1991: 349.	1		Commentary on poetry
al-Tha‘labī (died in 1038), <i>Fiqh al-Lughah (Philology)</i> , 1885: 434.	1		Linguistics
‘Alī ibn Khalaf (died in 1040), <i>Mawādd al-Bayān</i> , 1982, vol. 2: 132.	2		Rhetoric
Ibn Rashīq (died in 1064), <i>al-‘Umdah</i> , 1972a: 46.	5+1 +6	Correcting a previous statement (by using a disjunctive article).	Criticism of poetry and prose
al-Tibrīzī (died in 1109), <i>Sharḥ al-Mufaḍḍalīyyāt</i> , 1977: 947.	1+5		Commentary on poetry
Abū Ṭāhir al-Baghdādī (died in 1124), <i>Qānūn al-balāgha</i> (the Canon of Rhetoric), 1981: 110.	5		Rhetoric
al-Zamakhsharī (died in 1144), <i>al-Kashshāf</i> , 1998, vol. 1: 118.	2		Qur’ān Exegesis
al-Rāzī (died in 1210), <i>Nihāyat al-‘Ijāz</i> (Utmost Conciseness), 1899: 203.	1+2		Qur’ān Exegesis

al-Sakkākī (died in 1229), <i>Miftāḥ al-ʿUlūm (Key to Sciences)</i> , 1990: 112-113.	2+ 7	Not abiding by the default mode of address (3rd person, 2nd person) and talking to a different addressee.	Rhetoric
Ibn al-ʿAthīr (died in 1239), <i>al-Mathal al-Sāʿir</i> , n.d., vol. 2: 167-181.	6+2 +8	Shifting from the past to the future and vice versa, shifting from the past and future to the imperative and vice versa.	Criticism of poetry and prose.
Ibn Abī al-ʿIṣḥāq al-Miṣrī (died in 654 AH), <i>Badīʿ al-Qurʾān</i> , n.d.: 42.	2+4 9	When a speaker presents two things/persons for discussion. Then he tackles the first and moves to the second. After that, he moves back to discuss the first again.	Rhetoric
al-Muʿayyad al-ʿAlawī (died in 656 AH), <i>al-Ṭirāz</i> , 1980: 132.	2+5		
Ḥāzīm al-Qartājannī (died in 1284), <i>al-Minhāj</i> , 1966: 316.	4+ 10	Shifting from one thesis/topic to another without any intermediary or purpose.	Rhetoric
al-Sijlimāsī (died in 704 AH), <i>al-Manzīʿ al-Badīʿ</i> , 1980: 442.	2+4		Rhetoric

Najm al-Dīn ibn al-'Athīr (died in 1336), <i>Jawhar al-Kanz (Jewel of the Treasure)</i> , 1974: 119-121.	5+8 11	Shifting from the dual to the plural and from the plural to the singular.	Rhetoric
al-Qazwīnī (died in 1338), <i>al-'Īdāh</i> , 1993: 72.			Rhetoric
al-'Alawī (died in 1348), <i>al-Ṭirāz</i> , 1980: 131-132.	8+11 12	Shifting from one style/technique to another.	
al-Subkī (died in 1372), <i>Shurūḥ al-Talkhīṣ</i> , 2003: 463.	2+7		Rhetoric
al-Bābirtī (1384), <i>Sharḥ al-Talkhīṣ</i> .	2+7		Rhetoric
al-Zarkashī (died in 1392), <i>al-Burhān Fī 'Ulūm al-Qur'ān</i> .	2+7 13	Shifting from the active voice to the passive.	Qur'ānic sciences

Table 1 demonstrates that *'iltifāt* has been used in Arabic rhetorical legacy to refer to 13 figures of speech. Analyzing the interrelationships among them, we can group them under one overarching general category divided into two sub-categories: each sub-category has a set of standard features that set it apart from the other.

The part that links all figures together is "shifting from one style of discourse to another" (al-'Alawī 1980: 132). al-Subkī (2003: 464) reported the same meaning from an anonymous: "shifting discourse from one mode to another". This generalized, unrestricted definition equalizes *'iltifāt* with *'udūl* (deviation). Perhaps the various concepts of the term and al-'Alawī's desire to set a comprehensive definition are the reasons for formulating so stretched a concept for *'iltifāt*.

Choosing this concept to encompass the other concepts attached to *'iltifāt* is justified by the lexical and semantic spaciousness of the word itself. al-'Alawī (1980) did not plan to gather all the forms of deviation under *'iltifāt*; otherwise,

he would have listed them, but he did not. He did not refer to any forms he had not mentioned in his book except for shifts among pronouns, verb tenses, and numbers. However, his writings imply that *'iltifāt* encompasses all forms of *'udūl* (deviation), which is divided into two categories of figures: the first includes shifts in the addressee pronouns, their tense and number. In contrast, the second category contains shifts from one meaning or topic to another. The first category has the following figures:

- Shifting from first, second, or third-person pronouns to another.
- Shifting from what is supposed to be in the second, third, or first person to another form.
- Shifting from the past to the future and vice versa, shifting from the past and future to the imperative.
- Shifting from the dual form to the plural and from the plural to the singular.
- Shifting from the active voice to the passive.

These five figures of speech have one thing in common: deviating from the grammatical rule to change the addressee pronoun, tense, or number. I believe that the reason for attaching the term *'iltifāt* to pronoun shift is the same reason for connecting it to the rest of the figures: it is a result of the similarity among them. al-'Akhfash al-'Awsat (1990: 275) grouped these figures under one type: "that whose beginning is changed". Moreover, these figures are also grouped by Ibn Qutaybah (1973: 275) under one heading: "The discrepancy between surface structure and meaning".

al-Subkī (2003: 464) realized one dimension of this relationship in his explanation of the view that shifting the number of addressees is one form of *'iltifāt*. He states:

Other scholars regard the shift from talking to one person to talking to two or more as one form, and this is the closest thing to the famous term *'iltifāt* due to its similarity to it in that it marks a shift between three types (singular, dual and plural) and its application to six cases.

Although al-Subkī does not go beyond this surface similarity, his notice of resemblance is worth appreciation. Concept no. 7 has been developed from no. 2: the latter marks a shift between actual pronouns inside the text, while the former marks a shift between the default pronouns used in a specific context and the actual one (which can violate this rule). Shifts among tenses always accompany pronoun shifts because they have never been considered one form of *'iltifāt* when they come alone; they come with pronoun shifts just to expand the concept. There is a close relationship between tense shift and pronoun shift in that they both form the addressee pronoun and tense. There are also unique relationships between some pronouns and some tenses. For instance, first and second-

person pronouns are linked to the present tense, while the third-person pronoun mode is usually related to the past. Therefore, shifts among pronouns may lead to shifts in the tense and vice versa.

Similarly, considering the shifts between the numbers of addressees (second person singular, dual, and plural), one form of *'iltifāt* can be explained in light of a similar link between this shift and pronoun shift. In addition to the surface similarity mentioned by al-Subkī, these two kinds of shifts come together in many texts, like in the Qur'ān, 35:9 (Abdel Haleem's translation, 2005), which reads: "It is God who [sent] forth the winds; they raise the clouds; We [drove] them to a dead land and with them [revived] the earth after its death ...". This verse has three shifts: (1) pronoun shift from third person to first person; (2) tense shift from past to present and then to the past again; (3) subject number shift from third-person singular to first-person plural. The three shifts collaborate to enrich the effect of the verse. However, the link between pronoun shifts and active-passive shifts is the strongest of these relationships because the latter is one form of the former. Formulating a sentence in the passive after addressing the subject in an active voice sentence entails a shift from the second person to the third person pronoun.

Viewing the development of the term over time makes it clear that including this figure under the umbrella of *'iltifāt* took place at a relatively later stage, starting with Ibn al-'Athīr. Perhaps the reason behind this is the desire to expand the concept of *'iltifāt* to include other techniques that did not have established names or terms and had some similarities with pronoun shifts. It is perhaps safe to say that these techniques or figures of speech were linked to *'iltifāt* because one of them was similar to *'iltifāt*, and they had something in common; therefore, the term was expanded to refer to them all due to this resemblance. Thus, it can be concluded that one of the procedures used by Arab rhetoricians to broaden the concept of a specific term was to apply it to concepts similar to the original one in one way or another.

The second subcategory of figures connected with *'iltifāt* revolves around shifting from one meaning/topic to another (concept no. 3). This concept was narrowed down by Qudamah ibn Ja'far (concept no. 4) by defining the reason for shifting and the relationship between the two meanings/topics. According to Qudamah, shifting between topics/theses happens when a doubt occurs in the mind of a poet concerning the first topic/thesis; or when he anticipates another person to present an antithesis to his thesis or asks him to justify it (Ibn Ja'far 1948: 147). The motivation for shifting, in Qudamah's opinion, can be intrinsic (e.g., when the poet thinks he did not give the best expression to the topic) or extrinsic (e.g., when the poet takes into consideration the context in which his poem will be received and tries to anticipate possible objections and respond to them before the recipients express them). Consequently, the new topic/thesis (to which the shift happens) does a service to the first one (from which the shift happens) in that "it emphasizes the other topic, justifies it, or dispels any doubts about it" (Ibn Ja'far, 1948: 147). In this way, Qudamah restricts

the concept of *'iltifāt* "shifting from one meaning /topic to another," because the shift fills a gap in the first meaning/topic or emphasizes or justifies it.

Similarly, concept no. 5, "embedding one statement/phrase in another", is one case of shifting from one topic/meaning to another. *'Istidrāk* using a disjunctive article (concept no. 6) puts a constraint on shifting in that it is stipulated that the two meanings/statements/topic are contradictory and that they are connected with a disjunctive article. In contrast, al-Qarṭājannī (1966) defines the absence of any intermediary (verbal or nonverbal) or purpose for shifting from one meaning/statement/topic to another. For him, *'iltifāt* is "shifting from one thesis/statement to another without any intermediary or purpose" (concept no. 10). The last concept in this subcategory further restricts the general definition of *'iltifāt*, "shifting from one meaning/statement/topic to another". Ibn Abī al-ʿIṣbaʿ al-Misrī stipulates for the figure to be called *'iltifāt* to have two shifts, not simply one: shifting from the first statement/meaning before its completion to the second and shifting from the second before its completion to the first.

This analysis shows that some concepts came to be attached to *'iltifāt* through two contradictory processes: the first expands the concept by including more figures, highlighting their similarities, and ignoring their peculiarities. The second process narrows down the concept by dividing it into smaller components, highlighting the differences at the expense of the similarities. In the first process, this attachment between the concept and the term occurs by merging similar figures, while the second process separates the similar figures.

The analysis also reveals that these two processes have been primarily employed in two different disciplines: the first has been used in the fields of general rhetoric and inimitability/miraculousness (*'i ʿjāz*) of the Qurʾān, and it has been applied to figures of speech that are similar in that they are used to disregard a typical grammatical rule. The second process is standard in literary studies and is related to figures of speech that mark semantic shifts. In short, it can be stated that attaching different concepts to the term *'iltifāt* in Arabic rhetorical heritage resulted from two processes: merging similar figures of speech or splitting them. Another finding from this analysis is related to Ibn al-Muʿtazz's concept of *'iltifāt*, which played a central role in crystalizing the terminological structure of this term. This concept was the point of departure for all other attempts that tried a conceptual identification of *'iltifāt*, highlighting the centrality of his book *al-Badīʿ* in Arabic rhetorical legacy.

2.2 Listing terms that refer to one concept

The purpose of this procedure is to define the concepts that pertain to a specific field/discipline and to study the words that refer to them in context and the inter-relationships among them (al-Qāsimī 1985: 223). The concept analyzed here is shifts among first, second, and third-person pronouns. The table below consists of three columns; the first shows the year the author who mentioned the con-

cept died; the second gives the term he used to refer to this concept, and the third quotes him verbatim.

Table 2: The connection between concept and term

Year of the author's death (AD)	Term	Quote
822.	<i>Majāz</i>	"One form of <i>Majāz</i> is when one is spoken to as if he were present (second person) then he is referred to as if he were absent (in the third person), or vice versa ...". <i>Majāz al-Qur'ān</i> by Abū 'Ubaydah (1954: 19).
822.	0 ²	"The Qur'ān sometimes addresses people to whom it was revealed (in the second person), and sometimes it talks about them as if they were absent (in the third person)." <i>Ma'ānī al-Qur'ān</i> (Meanings of the Qur'ān) by al-Farrā' (1980: 211).
830.	0	"It (the verse of Qur'ān) talks to them after talking about them ... so it uses the third person after the second person." <i>Ma'ānī al-Qur'ān</i> (Meanings of the Qur'ān) by al-'Akhfash (1990: 137-138).
889.	The discrepancy between the surface structure and meaning	"(There) is a section (of the book) on the discrepancy between the surface structure and meaning. An example is when you talk about something to the addressee and then talk to him as if he were a third person ...". <i>Ta'wīl Mushkil al-Qur'ān</i> by Ibn Qutaybah (1983: 289-290).
898.	0	"The Arabs sometimes use the third person pronoun in addressing present people and the second person pronoun to talk about absent people." <i>al-Kāmil</i> by al-Mubarrad (1956: 30).

908.	'iltifāt	"A section (of the book is) on 'iltifāt: when the speaker shifts from the second person to the third person and from the third person to the second person and the like. Another example of 'iltifāt is changing the topic ...". <i>al-Badī'</i> by Ibn al-Mu'tazz (1967: 58).
923.	0	"The Arabs use the second person, then shift to the third person, and then go back to use the second person again." <i>al-Ṭabarī's Commentary on the Qur'ān</i> (n.d.: 153)
923.	0	The Qur'ānic verse, "... when you are in ships, and they sail with them by a good wind" (10:22). It started as addressing people in the second person and then shifted to use the third person pronoun. 'I'rāb <i>al-Qur'ān</i> (attributed to the same author) (1982: 923).
940.	0	"The Arabs switch between the second and third-person pronouns in their discourse ...". <i>Sharḥ al-Qaṣā'id al-Sab'</i> (Commentary on the Seven Poems) by Ibn al-'Anbārī (1969: 300).
946.	ṣarf	"ṣarf, they <i>yaṣrefun</i> (change/shift) discourse from the direct addressee (second person) to talking about the absent (third person) and from the singular to the plural." <i>al-Burhān</i> by Ibn Wahb al-Kātib (1969: 152).
338 AH.	0	"The Arabs switch between the third and second-person pronouns in their discourse." <i>Sharḥ al-Qaṣā'id al-Tis'</i> by al-Naḥḥās, part 2: 463.
980.	0	"The Arabs switch between the second and third-person pronouns in their discourse ...". <i>al-I'rāb fī al-Qirā'āt al-Sab'</i> , Ibn Khālawayh (1992: 121).
980.	0	"The Arabs use the second person pronoun then switch to the third person pronoun." <i>Ma'ānī al-Qirā'āt</i> by al-'Azharī (1993: 96).

987.	0	"It is possible in Arabic to use the second person pronoun after the third person to talk to the same person like in verse 'It is You we worship' after using the third person pronoun." Abū 'Alī al-Fārisī, <i>al-Hujjah</i> (2000: 383).
998.	'iltifāt	"'iltifāt means changing between the second to the third person pronouns." <i>Maḥāṭib al-'Ulūm</i> by al-Khuwārazmī (1924: 61).
1002.	<i>Shajā'at al-'Arabiyya</i> (Courage of the Arabic Language)	"We call 'leaving the second person for the third person pronoun' <i>Shajā'at al-'Arabiyya</i> (Courage of the Arabic Language)." Ibn Jinnī: <i>al-Muhtasab</i> (1999: 139, 145); <i>al-Fasr</i> (2004, vol. 1: 210) and <i>al-Khāṭiriyyāt</i> (1988: 69).
1002.	0	"He (the poet) shifted from the second person pronoun to the third person pronoun." <i>al-Wasāṭah</i> by al-Jurjānī (1992: 344).
1003.	'iltifāt	"'iltifāt is used in poetry, and it means a switch between the third person and the second person pronouns and the other way round." <i>al-Munisf</i> by Ibn Wakī' al-Tannīsī (1994: 166).
1005.	0	"The shift from the third person to the second person pronoun, and vice versa." <i>al-Ṣāḥibī</i> by Ibn Fāris (1977: 356-357).
1030.	0	"Changing the mode of discourse from the third person to the second person pronoun", <i>Sharḥ al-Ḥamasah</i> by al-Marzūqī (1991: 349 & 280); and "he (the author) used the second person, then switched to the third person when talking about himself." al-Marzūqī (1991: 271).
1038.	0	"The switch between the third person and the second person pronouns and the other way round." <i>Fiqh al-Lughah</i> by al-Tha'ālibī (1885: 349).

1040.	'iltifāt	"'iltifāt occurs when the speaker switches from the third person to the second person pronoun, and vice versa." <i>Mawādd al-Bayān</i> by 'Alī ibn Khalaf (1982: 288).
1064.	'iltifāt	"Ibn al-Mu'tazz chose the right words when he said that 'iltifāt is 'when the speaker shifts from the second person to the third person, and from the third person to the second person pronoun ...". <i>al-'Umdah</i> by Ibn Rashīq (1972a: 46).
1064.	0	"He (the author) shifted from the direct second-person pronoun to the indirect third-person pronoun." <i>al-tibyān</i> by al-Tūsī (1957: 35).
1073.	0	"The change of adjectives and using pronouns outside their typical context." <i>Sirr al-Faṣāḥa</i> (the Secret of Eloquence) by Ibn Sinān al-Khafājī (1982: 109).
1109.	0	"He (the author) shifted from the second person to the third person pronoun." <i>Sharḥ al-Mufaḍḍaliyyāt</i> by al-Tibrīzī (1977: 445-448) "shifting the discourse from speaking to the addressee to speaking about another person" (the same author, vol. 1: 41) and <i>Sharḥ al-Qaṣā'id al-'Ashr</i> (1934: 268).
1122.	0	"Moving between the second-person and third-person pronouns." <i>Ma'ālīm al-Tanzīl</i> by al-Baghawī (1912: 149).
1144.	'iltifāt	"'iltifāt in the <i>balāgha</i> refers to the switch from third person to second-person pronouns and vice versa, and then to the first person ...". <i>al-Kashshāf</i> by al-Zamakhsharī (1998: 118).
1144.	<i>Khiṭāb al-Talawwun</i> (changing discourse)	" <i>Khiṭāb al-Talawwun</i> has three forms: to use the second person and then change to the third person pronoun, the other way round." <i>al-Durr al-Dā'ir</i> by al-Zamakhsharī (1968: 23).
1147.	0	"Using the second person, then the third person pronoun, and the other way round." <i>al-Muḥarrar al-Wajīz</i> by Ibn 'Aṭīyah (1974: 107).

1153.	0	"Moving from the third person pronoun to the second person pronoun." <i>Tafsīr at- Ṭabarūsī: Majma' al-Bayān</i> (1958: 40).
1188.	<i>'inṣirāf</i>	"A section (of the book is) on <i>'inṣirāf</i> which is shifting between the use of second and third person pronouns and vice versa." <i>al-Badr</i> by Ibn Munqidh (1987: 287).
1201.	0	"Arabs use the second person pronoun then switch to the third person pronoun and vice versa." <i>Zād al-Masīr fī 'ilm al-Tafsīr</i> Ibn al-Jawzī (1964: 14).
1210.	<i>'iltifāt</i>	"It is said that <i>'iltifāt</i> is the shift between the second-person and the third-person pronoun." <i>Nihāyat al- 'Ijāz fī Dirāyat al- 'Ijāz</i> by al-Rāzī (1899: 203).
1219.	0	"It is conventional for Arabs to use the second person pronoun then switch to the third person pronoun and the other way round." <i>'Imlā' mā Manna Bihi al-Raḥmān</i> by al- 'Akbarī (1969: 6).
1228.	<i>'inṣirāf</i>	"This is a chapter (of the book) on <i>'inṣirāf</i> which is to start a discourse with the third person pronoun and then shift to the second person pronoun." <i>Ma'ālīm al-Kitābah</i> by Ibn Shi' th al-Qurashī (1988: 106).
1229.	<i>'iltifāt</i>	"Moving among the first, second and third-person pronouns are called <i>'iltifāt</i> ." <i>al-Miftāḥ</i> by al-Sakkākī (1990: 112).
1239.	<i>'iltifāt</i> + <i>Shajā'at al- 'Arabiyya</i> (Courage of the Arabic Language)	"(A chapter) on <i>'iltifāt</i> ... and it is also known as <i>Shajā'at al- 'Arabiyya</i> (Courage of the Arabic Language), and it has three types: the first one is to shift between third to second-person pronouns and vice versa ... etc." <i>al-Mathal al-Sā'ir</i> by Ibn al- 'Athīr (n.d.: 168).
1258.	0	"It is a shift between a third to second-person pronouns and vice versa" <i>Nuḍrat al- 'Ighrīd fī Nuṣrat al-Qarīd</i> by al-Muzaffar ibn al-Faḍl al- 'Alawī (1976: 105-7).

1257.	'iltifāt	"'Ilṭifāt is the switch between third, second, and first-person pronouns and vice versa." <i>Mi 'yār al-Nuẓẓār fī 'Ulūm al- 'Ash 'ār</i> by al-Zanjānī (1991: 103).
1262.	'iltifāt + <i>Talwīn al-Khiṭāb</i> (Changing/colouring discourse)	"Moving from the third person to the second person is known as 'iltifāt" <i>al-Fawā'id fī Mushkil al-Qur'ān</i> by 'Ezz al-Dīn ibn 'Abd al-Salām (1967: 52). "It is also called <i>Talween</i> (changing/colouring)" (1967: 53). "Changing and using a variety of modes, namely, moving from the second person to the third, and vice versa" (ibid.: 101).
1284.	0	"They get bored of using only the second or third person pronouns, so they shift between them." <i>Minhaj Al-Bulaghāh</i> by al-Qarṭājannī (1966: 348).
1287.	'iltifāt	"First, second and third-person pronouns can be used in place of each other, and this is called 'iltifāt" <i>al-Miṣbāḥ fī al-Ma'ānī wa al-Bayān wa al-Badī'</i> by Ibn al-Nāẓim (1989: 30).
7th century AH.	<i>Tawassu'āt</i>	"Among the licenses the Arabic language allows is the freedom of movement between the third and second-person pronouns" <i>al-'Aqṣā al-Qarīb</i> by al-Tanūkhī (1909: 44).
1305 ³ .	'iltifāt	"'Ilṭifāt which is called by some people the varied discourse (<i>khiṭāb al-talawwun</i>), which means the movement between different modes of discourse." <i>al-Manzī' al-Badī'</i> by al-Sijlimāsī (1980: 442).
1321.	'iltifāt + <i>khiṭāb al-talawwun</i>	"(The) switch from the second to the third person pronouns, and vice versa is called 'iltifāt or <i>khiṭāb al-talawwun</i> ." <i>al-Rawḍ al-Murī'</i> Ibn al-Bannā' al-'Adadī (1985: 98).
1336.	'iltifāt	"There are different types of 'iltifāt, like moving from the third person to the second person pronoun and the opposite." <i>Jawhar al-Kanz</i> by Ibn al-'Athīr, (1974: 120).

1338.	<i>'iltifāt</i>	"The free movement of the first, second, and third-person pronouns, is called <i>'iltifāt</i> ." <i>al-'Idāḥ ft 'Ulūm al-balāgha</i> by al-Qazwīnī (1993: 72).
1348.	0	"In the Qur'ān, the verse 'It is You we worship; it is You we ask for help' marks a shift from the third person to the second person pronoun" <i>al-Baḥr al-Muḥīt</i> by Abū Ḥayyān al-'Andalusī (1987: 252).
1348.	<i>'iltifāt</i> + <i>Shajā 'at al-'Arabiyya</i> (Courage of the Arabic Language)	"Talking about <i>'iltifāt</i> ... it is also called <i>Shajā 'at al-'Arabiyya</i> (Courage of the Arabic Language) ... and the first type means the shift between the first, second and third-person pronouns" <i>al-Ṭirāz</i> by al-Mu'ayyad al-'Alawī (1980: 131).
1372.	<i>'iltifāt</i>	"Changing the mode of discourse from first, second and third-person pronouns are called <i>'iltifāt</i> " <i>Sharḥ al-Talkhīṣ</i> by al-Subkī (2003: 463).
1384.	<i>'iltifāt</i>	"Expressing a topic with a shift between first, second and third-person pronouns" <i>Sharḥ al-Talkhīṣ</i> by Akmal Ad-Din al-Bābirtī (1983: 257).

Table 2 reveals the increasing connection between the concept and the term over time: there is this development from (1) not using a term to (2) using one term to refer to a concept, and then (3) using more than one term to refer to the concept in one book. For example, in the fourth century AH, 13 scholars talked about shifts between first, second, and third-person pronouns. Four used names to refer to them, while the other nine did not use any words or terms to denote them. However, eight scholars discussed this figure of speech in the eighth century AH. All of them used terms to name it, and two used more than one term to dub it. In this way, we can track the terminological development of *'iltifāt* in the Arabic rhetorical heritage. This process starts with recognizing the figure and formulating a definition for it. This is followed by classifying it into major categories and similar figures referred to by the same term. Then, each figure is given one term that becomes well-established. Eventually, the figure is dubbed with more than one term to distinguish it from other similar figures.

The development of the concept within different disciplines indicates that two of the disciplines that contributed to rhetorical heritage on *'iltifāt* did not use any terms to refer to this concept in most cases. The first discipline is Qur'ānic lectures (modes of reading/recitation), where scholars like Abū 'Alī al-Fārisī,

Ibn Khālawayh, al-ʿAzharī, and al-ʿAkbarī did not give any term to the figure of pronoun shift (see section 1.1.). The second discipline is literary studies, as shown in commentaries on poetry by al-ʿAnbārī, al-Naḥḥās, al-Tibrīzī, and al-Marzūqī. It is also reported that three critics of poetry, i.e., al-Qāḍī al-Jurjānī, Ibn Sinān al-Khafājī, and Ḥāzim al-Qarṭājannī, did not use any name to dub this figure.

Different disciplines produced the terms that refer to pronoun shifts: *ʿiltifāt* was used in poetry criticism; *Shajāʿat al-ʿArabiyya* (Courage of the Arabic Language) was used in Arabic linguistics; while *Talween al-Khetab*, *Talween*, and *khiṭāb talawwun* were produced in the discipline of Qurʿānic studies. It can be noticed that the last three terms are based on the freedom of building the genitive and the attributive noun (freedom of ordering the nouns) in *khiṭāb al-talawwun* and *Talwīn al-khiṭāb* or otherwise just dropping one and using only one noun *Talween* that reflects the freedom enjoyed by classical Arab rhetoricians in coining new terms.

3. Conclusion

This article discussed the challenges that encumber contemporary understanding of Arabic rhetorical terms. It has explored the roots of these challenges and suggested some procedures to alleviate them. Suppose these procedures are applied to the bulk of Arabic rhetorical heritage, with the participation of a group of researchers. In that case, they may help compile a historical, contextual dictionary that we need. These procedures are summarised by tracing the different concepts that revolve around one term and listing the additional terms that refer to one concept. Finally, to understand the process of producing rhetorical knowledge, the influence of historical and cultural factors on coining terms in Arabic *Balāgha* were examined.

The procedures suggested by this research to standardize Arabic rhetorical terminology seem necessary and urgent because they fill confusing gaps and flaws in the currently used dictionaries of Arabic *Balāgha*, e.g., the repetitions of some concepts with different terms to dub them and the absence of basic concepts due to the lack of terms denoting them.

Endnotes

1. The term *balāgha* has been used as a synonym to the word Rhetoric. However, the Arabic *Balagha* is not exactly the Western Rhetoric. For a recent work on the differences between al-Balagha and Rhetoric see Halldén 2022.
2. The (0) here means that the author did not use any term to refer to this phenomenon.
3. The year the book was written, not the year the author died.

References

A. Resources

- Abū 'Ubaydah, Ma'mar ibn al-Muthanná al-Taymī. 1954. *Majāz al-Qur'ān*. Sizkīn, M.F. (Ed.). Cairo: Maktabat al-Khānjī.
- al-Khuwārizmī, Muḥammad ibn Aḥmad. 1924. *Mafātīḥ al-'ulūm*. Cairo: Maṭba'at al-Sharq.
- al-Rāzī, Fakhr al-Dīn Muḥammad ibn 'Umar. 1899. *Nihāyat Al-Ījāz Fī Dirāyat al-I'jāz: Fī 'ulūm al-Balāghah Wa-Bayān I'jāz al-Qur'ān al-Sharīf*. Cairo: Maṭba'at al-Ādāb wa-al-Mu'ayyid.
- al-Sakkākī, Abū Yaḥyā Yusuf b. Abi Bakr Muḥammad b. 'Alī. 1990. *Miftāḥ al-'Ulūm*. Second edition. Cairo: al-Babi al-Halabi Library.
- al-Bābirtī, Akmal al-Dīn. 1983. *Sharḥ Al-Talkhīṣ*. Sufiyyah, M.M.R. (Ed.). Tripoli: al-Munsha'ah al-Ammah Li-Inashr.
- al-Baghawī, Abū Muḥammad al-Ḥusayn al-Farrā'. 1912. *Ma'ālīm al-Tanzīl*. Cairo: al-Takadum Scientific Press.
- al-Baghdādī, Abū Ṭāhir Muḥammad ibn Ḥaydar. 1981. *Qānūn al-Balāghah Fī Naqd al-Nathr Wa-al-Shi'r*. 'Ujayl, M.G. (Ed.). Beirut: Mu'assasat al-Risālah.
- al-Bāqillānī, Abū Bakr Muḥammad b. al-Ṭayyib. 1978. *I'jāz al-Qur'ān*. Cairo: al-Halabi Library.
- al-Farrā', Yaḥyā ibn Ziyād. 1980. *Ma'ānī al-Qur'ān*. Najātī, A. and M. 'Alī al-Najjār (Eds.). Cairo: Dār al-Kutub al-Misrīyah.
- al-Ḥatīmī, Abū 'Alī Muḥammad b. al-Ḥasan b. al-Muẓaffar. 1979. *Ḥilyat al-Muḥāḍara Fī Ṣinā'at al-Shi'r*. al-Katani, J. (Ed.). Baghdad: Dar al-Rasheed Publishing Ministry of Culture and Information.
- al-Jurjānī, Abdel-Qāhir. 1984. *Dalā'il Al-I'jāz*. Shaker, M.M. (Ed.). Cairo: al-Khanji Library.
- al-Jurjānī, 'Alī b. 'Abdul-'Azīz. 1992. *Al-Wasāṭa Bayn al-Mutanabbī Wa Khuṣūmih*. al-Zein, A.A. (Ed.). Tunis: Dar al-Maaref Printing and Publishing.
- al-Khafājī al-Ḥalabī, Ibn Sinān Muḥammad 'Abdullāh b. Muḥammad b. Saeed. 1982. *Sirr Al-Faṣāḥa*. Beirut: Dar al-kutub al-'ilmiyyah.
- al-Khāzin, 'Alā' Al-Dīn 'Alī b. Muḥammad b. Ibrāhīm Al-Baghdādī. n.d. *Lubāb al-Ta'wīl Fī Ma'ānī al-Tanzīl*. The Library of Hazrat Muḥammad Abdul-Wahed Bey al-Toubi and his brother.
- al-Marzūqī, Abū 'Alī Aḥmad ibn Muḥammad. 1991. *Sharḥ Dīwān al-Ḥamāsah*. Amin, A. and A.-S. Harun (Eds.). Beirut: Dar al-Jil.
- al-Mubarrad, Muḥammad ibn Yazīd. 1956. *al-Kāmil*. al-Fadl Ibrāhīm, M.A. and S. Shihātah (Ed.). Cairo: Maṭba'at Nahḍat Misr.
- al-Naḥḥās, Aḥmad ibn Muḥammad. 1973. *Sharḥ al-Qaṣā'id al-Tis' al-Mashhūrāt*. Khaṭṭāb, A. (Ed.). Baghdad: Mudīriyat al-Thaqāfah al-'Āmmah.
- al-Qarṭājannī, Ḥāzim. 1966. *Minḥāj Al-Bulaghā' Wa Sirāj al-'Udabā'*. Ibn al-Khuja, M. (Ed.). Tunisia: Dar al-Kutub al-Sharqiah.
- al-Qazwīnī, al-Khatīb. 1993. *al-Īdāḥ fī 'Ulūm al-Balāghah*. al-Khafājī, M.A. (Ed.). Cairo: al-Maktabah al-Azharīyah lil-Turāth.
- al-Qurashī, Ibn Shi'th 'Abd al-Raḥīm 'Alī. 1988. *Ma'ālīm al-Kitābah*. Shams al-Dīn, M.H. (Ed.). Beirut: Dar al-kutub al-'ilmiyyah.
- al-Sijlīmāsī, Abū Muḥammad al-Qāsim al-Anṣārī. 1980. *al-Manza' al-Badī' Fī Tajnīs 'Asālib al-Badī'*. al-Ghazi, A. (Ed.). Rabat: al-Maaref Library.

- al-Subkī, Bahā' al-Dīn. 2003. *'Arūs al-'Afrāḥ Fī Sharḥ Talkhiṣ al-Miftāḥ*. Hindawi, A.-H. (Ed.). Beirut: Dar al-kutub al-'ilmiyyah.
- al-Ṭabarī, Muḥammad b. Jarīr. n.d. *Jāmi' al-Bayān 'an Ta'wīl al-Qur'ān*. Shaker, M.M. (Ed.). Cairo: Dar al-Maaref.
- al-Ṭabarūsī, al-Fadl b. al-Hassan. 1958. *Majma' al-Bayān*. Cairo: Dar Al-Taqrīb Between Islamic Schools of Thought.
- al-Tanūkhī, Zayn al-Dīn Abī 'Abdillāh Muḥammad b. Muḥammad b. Muḥammad b. 'Amr. 1909. *al-'Aqṣā al-Qarīb Fī 'ilm al-Bayān*. Cairo: al-Khanjī Publishers.
- al-Tha'ālibī, 'Abd al-Malik ibn Muḥammad. 1885. *Kitāb Fiḥ al-Lughah*. Beirut: Maṭba'at al-Ābā' al-Yasū'iyyīn.
- al-Tibrizī, Yaḥyā ibn 'Alī. 1934. *Sharḥ al-Qaṣā'id al-'ashr*. Cairo: Maṭba'at Muḥammad 'Alī Ṣubayḥ wa-Awlādiḥ.
- al-Tibrizī, Yaḥyā ibn 'Alī. 1977. *Harḥ al-Mufaḍḍaliyyāt*. al-Bajāwī, A.M. (Ed.). Cairo: Dār Nahḍat Miṣr.
- al-Ṭūsī, Abū Ja'far Muḥammad b. al-Ḥassan. 1957. *al-Tibyān Fī Tafsīr al-Qur'ān*. Najaf: Scientific Press.
- al-Zajjāj, Abū Ishāq Ibrāhīm ibn al-Sarī. 1982. *I'rāb al-Qur'ān*. al-Ibyārī, I. (Ed.). Cairo: Dār al-Kutub al-Islāmīyah.
- al-Zamakhsharī, Maḥmūd ibn 'Umar. 1968. *Āldār Ālmontakhab Mn Knāyāt Wā Ist'ārāt Wa Tšbyhāt Āl'rb*. al-Husni, B. (Ed.). Baghdad: Maṭbū'āt al-Majma' al-'Ilmī al-'Irāqī.
- al-Zamakhsharī, Maḥmūd ibn 'Umar. 1998. *al-Kashshāf 'an Ḥaqā'iq Ghawāmiḍ al-Tanzīl Wa-'uyūn al-Aqāwīl Fī Wujūh al-Ta'wīl*. 'Abd al-Mawjūd, A.A. and A.M. Mu'awwad (Eds.). Riyadh: Maktabat al-'Ubaykān.
- al-Zanjānī, Abdul-Wahhāb b. Ibrāhīm b. Abdul-Wahhāb al-Khazrajī. 1991. *Mi'yār al-Nuẓẓār Fī 'Ulūm al-'Ash'ār*. al-Khafājī, M.A. (Ed.). Cairo: Dar al-Maaref.
- al-Zarkashī, Abū 'Abdullāh Badr al-Dīn Muḥammad b. 'Abdullāh b. Bahādir. 1957. *al-Burhān Fī 'Ulūm al-Qur'ān*. Ibrahim, M.A. (Ed.). Cairo: Dār Iḥyā' al-Kutub al-Arabiyya.
- al-'Akhfash al-'Awsaṭ, Abū al-Ḥasan Sa'id b. Mas'ada. 1990. *Ma'ānī al-Qur'ān*. Qara'a, H. (Ed.). Cairo: al-Khanjī Library.
- al-'Andalusī, Abū 'Abdullāh Jamāl al-Dīn Muḥammad b. Aḥmad. 1987. *al-Mi'yār Fī Naqd al-'ash'ār*. Hindawi, A.M. (Ed.). Cairo: Al-Amanah Press.
- al-'Azharī, Abū Mansour Muḥammad b. Aḥmad. 1993. *Ma'ānī al-Qirā'āt*. Mustafa, E. and A. al-Qawzi (Eds.). Cairo: Dar al-Maaref.
- al-'Akbarī, 'Abdullāh b. al-Ḥusayn. 1969. *'Imlā' Mā Manna Bihi al-Raḥmān Min Wujūh al-'I'rāb Wa al-Qirā'āt Fī Jamī' al-Qur'ān*. 'Awad, A. (Ed.). Cairo: Mustafā al-Bābī al-Ḥalabī.
- al-'Alawī, al-Muẓaffar b. Faḍl. 1976. *Nuḍrat al-'Ighrīḍ Fī Nuṣrat al-Qarīḍ*. al-Hassan, N.A. (Ed.). Damascus: Maṭba'at Tirbin.
- al-'Alawī, Yaḥyā b. Ḥamzah b. 'Ibrāhīm. 1980. *Kitāb al-Ṭirāz al-Mutaḍammīn Li-'Asrār al-Balāgha Wa 'Ulūm Ḥaqā'iq al-'I'jāz*. Beirut: Dar al-kutub al-'ilmiyyah.
- al-'Askarī, Abū Hilāl. 1984. *Kitāb al-Ṣinā'atayn*. Second edition. Kamiha, M. (Ed.). Beirut: Dar al-kutub al-'ilmiyyah.
- al-Fārisī, Abū 'Alī al-Ḥasan ibn Aḥmad. 2000. *Al-Hujjah Fī 'Ilal al-Qirā'āt al-Sab'*. Nāsif, A.N., A. al-Najjār and A Shalabī (Eds.). Cairo: Maṭba'at Dār al-Kutub al-Miṣriyyah.

- Ibn Abī al-ʿIṣḥāq**, **Abd al-ʿAzīm ibn ʿAbd al-Wāḥid al-Misrī**. 1957. *Badīʿ al-Qurʿān*. Cairo: Maktabat Nahḍat Misr.
- Ibn al-Muʿtazz**, **ʿAbd Allāh**. 1967. *al-Badīʿ*. Krachkovsky, I. (Ed.). Baghdad: Al-muthana Publishers.
- Ibn al-Anbārī**, **Muḥammad ibn al-Qāsim**. 1969. *Sharḥ al-Qasāʿid al-Sabʿ al-Tiḥwāl al-Jāhiliyyāt*. Hārūn, A.M. (Ed.). Cairo: Dār al-Maʿārif.
- Ibn al-Bannāʾ**, **Aḥmad ibn Muḥammad**. 1985. *Al-Rawḍ al-Marīʿ Fī Ṣināʿat al-Badīʿ*. Binshaqrūn, R. (Ed.). Casablanca: Dār al-Nashr al-Maghribīyah.
- Ibn al-Jawzī**, **Abū al-Faraj Jamal al-Dīn Abd al-Rahman b. ʿAlī b. Muḥammad**. 1964. *Zād al-Masīr Fī ʿIlm al-Tafsīr*. Damascus: The Islamic Office for Printing and Publishing.
- Ibn al-Nāẓim**, **Badr al-Dīn ibn Mālik**. 1989. *al-Misbāḥ Fī al-Maʿānī Wa-al-Bayān Wa-al-Badīʿ*. ʿAbd al-Jalīl Yūsuf, H. (Ed.). Cairo: Maktabat al-ʿĀdāb.
- Ibn al-ʿAthīr**, **Abū al-Fath, Nasr Dīa al-Dīn**. n.d. *al-Mathal al-Sāʿir Fī Adab al-Kātib Wa al-Shāʿir*. al-Hofī, A. and B. Tabana (Eds.). Cairo: Nahdet Misr.
- Ibn al-Athīr**, **Najm al-Dīn Aḥmad ibn Ismāʿīl**. 1974. *Jawhar al-Kanz: Talkhīṣ Kanz al-Barāʾih fī Adawāt dhawī al-Yarāʾih*. Sallām, M.Z. (Ed.). al-Iskandarīyah: Munshaʿat al-Maʿārif bi-al-Iskandarīyah.
- Ibn Fāris**, **Abu al-Ḥusayn Aḥmad ibn Fāris ibn Zakarīyā**. 1977. *al-Sāḥibī*. Saqr, A. (Ed.). Cairo: Matḥaʿat ʿĪsā al-Bābī al-Ḥalabī.
- Ibn ʿAbd al-Salām**, **ʿIz al-Dīn Abū Muḥammad ʿAbd al-ʿAzīz**. 1967. *Āl-Fwāʿid Fī Mshkl al-Qurʿān*. Redwan, S. (Ed.). Cairo: Almtḥbʿt al-Mṣrīt.
- Ibn ʿAṭīyah**, **ʿAbd al-Ḥaqq ibn Ghālīb**. 1974. *al-Muḥqarrar al-Wajīz Fī Tafsīr al-Kitāb al-ʿazīz*. al-Mallāh, A.S. (Ed.). Cairo: al-Majlis al-Aʿlā lil-Shuʿūn al-Islāmīyah.
- Ibn Jaʿfar**, **Qudāmāh**. 1948. *Naqd al-Shiʿr*. Third edition. Mustafa, K. (Ed.). Cairo: al-Khanjī Library.
- Ibn Jinnī**, **Abū al-Fath ʿUthmān**. 1988. *al-Khatīrīyāt*. Shākir, D. (Ed.). Beirut: Dār al-Gharb al-Islāmī.
- Ibn Jinnī**, **Abū al-Fath ʿUthmān**. 2004. *al-Faṣr*. Ragab, R. (Ed.). Damascus: Dār al-Yanābiʿa.
- Ibn Jinnī**, **Abū al-Fath ʿUthmān**. 1999. *al-Muḥtasab Fī Tabyīn Wujūh Shawādh al-Qirāʾat Wa al-Idāḥ ʿanhā*. Nassef, A.N., A.H. al-Najjar and A.F.I. Shalaby (Eds.). Cairo: The Supreme Council for Islamic Affairs.
- Ibn Khālawayh**, **Abū ʿAbdullāh al-Ḥusayn b. Aḥmad**. 1992. *ʿIrāb al-Qirāʾat al-Sabʿ Wa ʿilʿAlīhā*. al-Uthaymeen, A.-R. (Ed.). Cairo: al-Khanjī Library.
- Ibn Manẓūr**, **Abū al-Faḍl Jamāl al-Dīn**. n.d. *Lisān al-ʿArab*. Khayat, Y. (Ed.). Beirut: Dar al-Lisan al-Arabi.
- Ibn Munqidh**, **Usāmāh**. 1987. *al-Badīʿ Fī al-Badīʿ Fī Naqd al-Shiʿr*. Muḥannā, A.A. (Ed.). Beirut: Dār al-Kutub al-ʿIlmiyyah.
- Ibn Qutaybah**, **ʿAbdullāh b. Muslim**. 1983. *Taʿwīl Mushkil al-Qurʿān*. Second edition. Saqr, A.-S. A. (Ed.). Cairo: Maktabat Dar At-Turath.
- Ibn Rashīq**, **Abū ʿAlī al-Hasan b. Rashīq al-Azdī**. 1972a. *al-Umdah Fī Maḥāsin al-Shiʿr Wa ʿĀdābiḥ*. Fourth edition. Abdul-Hamid, M. A.-D. (Ed.). Beirut: Dar al-Jeel.
- Ibn Rashīq**, **Abū ʿAlī al-Hasan b. Rashīq al-Azdī**. 1972b. *Qurādatu al-Dhahab Fī Naqd ʿAshʿār al-ʿArab*. Bou-Yahya, A.-S. (Ed.). Tunis: The Tunisian Company for Distribution.
- Ibn Wahb al-Kātib**, **Abī al-Ḥusayn Ishāq ibn Ibrāhīm ibn Sulaymān**. 1969. *al-Burhān Fī Wujūh al-Bayān*. Sharaf, H.M. (Ed.). Cairo: Maktabat al-Shabāb.
- Ibn Wakīʿ al-Tannīsī**, **Abū Muḥammad al-Hassan b. ʿAlī**. 1994. *al-Munṣif Li-al-Sāriq Wa al-Masrūq Minhu Fī ʿIzhār Sariqāt Abī al-Ṭayyib al-Mutanabbī*. bin Idris, O.K. (Ed.). Benghazi: Nashr Jamiʿat Qaryounis.
- Khalaf**, **Abū al-Ḥasan ʿAlī b. Khalaf al-Kātib**. 1982. *Maḥadd al-Bayān*. Libya: Manshurat Jamiʿat al-Fatih.

B. Other Sources

- Abdel-Haleem, Mohammed.** 1992. Grammatical Shift for Rhetorical Purposes, 'iltifāt and Related Features in the Qur'ān. *Bulletin of the School of Oriental and African Studies* 55(3): 407-432.
- Abdel-Haleem, Mohammed.** 2005. *The Qur'an*. Oxford: OUP.
- al-Haysharī, al-Shādhulī.** 1988. al-'iltifāt Fī al-Qur'ān al-Karīm. *Hawliyyat al-Jamī'ah al-Tunissiyyah* 23: 131-172.
- al-Qāsimī, 'Alī, Al-Mustalahiyyah.** 1985. *Muqaddima Fī 'ilm al-Muṣṭalaḥ*. Baghdad: Dar Ash-Shu'oun Ath-Thaqafiyya wa An-Nashr.
- al-Ṭrabelsī, Muḥammad al-Hadī.** 1991. Fī Mafhūm al-'Iqā'. *Annals of the Tunisian University* 32: 7-22.
- Halldén, Philip.** 2023. *Rhetoric (Islamic/Arabic)*. London: Routledge.
- Ibn Ṭālib, 'Uthmān.** 1989. 'Ilm al-Muṣṭalaḥ Bayna al-Mu'jamīyah Wa-'Ilm al-Dalālah: al-Ishkālāt al-Naẓariyyah Wa-al-Manhajīyah. *ta'sīs al-qaḍīyah al-iṣṭilāḥīyah*. Tunisia: Bayt al-Ḥikmah.
- Tabl, Hassan.** 1998. *'Uslūb al-'Iltifāt Fī al-Balāgha al-Qur'āniyya*. Cairo: Dar al-Fikr al-Arabi.

Bonny Sands en Kerry Jones (Hoofredakteurs). *N\uuki Namagowab Afrikaans English: ꞤXoakiꞤxanisi / Mîdi di ꞤKhanis / Woordeboek / Dictionary.* 2022, 313 pp. ISBN 978-0-6397-1245-1. Stellenbosch: African Sun Media. Gratis beskikbaar op aanvraag by die Departement van Sport, Kuns en Kultuur (DSKK) en African Tongue, Suid-Afrika.

Oor die meriete van die *N\uuki Namagowab, Afrikaans, English Woordeboek / Dictionary* (2022) bestaan daar geen onduidelikheid nie. Dié woordeboek is immers in 2023 met die ATKV-Woordwystoekenning vir woordeboeke en taalgidse bekroon. Die beoordelaarspaneel, bestaande uit die gevierde taalkundiges Rufus Gouws, Ernst Kotzé en Gerhard van Huyssteen, was eenparig van oordeel dat veral die innovasie, omvang en gemeenskapsgerigtheid van dié werk hiervan 'n merkwaardige naslaanbron maak.

In dié meertalige naslaanbron kom vier tale van die Noord-Kaap aan bod, naamlik N\uuki, Nama(gowab), Afrikaans en Engels. Sowel N\uuki as Nama word as bedreig beskou, maar die situasie is kritiek wat N\uuki betref en die taal is deur die United Nations Educational, Scientific and Cultural Organization (UNESCO) geklassifiseer as "op die rand van uitsterwing". Tans is Ꞥxuuleeki (mevrou) Katrina Esau van Upington die enigste oorlewende persoon wat N\uuki vlot kan praat. Die unieke variëteit van Afrikaans wat hier vir die eerste keer geboekstaaf word, staan bekend as "Onse Afrikaans" en die variëteit van Engels wat in die bron gebruik is, is Suid-Afrikaanse Engels.

Benewens die beperkte boekoplaag is dié bron ook in digitale formaat beskikbaar op die Suid-Afrikaanse Sentrum vir Digitale Taalhulpbronne (SADiLaR) se woordeboekportaal. Vir alle *N\uuki*-lemmas is daar vertaalekwivalente in Nama, Afrikaans en Engels, asook klankopnames wat help met die uitspraak en aanleer van die taal. Die app-weergawe van die woordeboek, Saasi Epsi, kan op Google Playstore afgelaai word en vir elke *N\uuki*-lemma word daar, benewens die vertaalekwivalente in Nama, Afrikaans en Engels, 'n fonetiese transkripsie gegee. YouTube-videos met duidelike, bondige gebruiksaanwysings is vir elk van die formate beskikbaar. Dié naslaanbronne is geskep met fondse wat deur die Departement van Sport, Kuns en Kultuur beskikbaar gestel is en SADiLaR en African Tongue — 'n taalkundige konsultasiediens wat met kontemporêre sprekers van die tale Ju, Tuu en Khoe in Suidelike Afrika saamwerk om kreatiewe en opvoedkundige hulpbronne te ontwikkel — het die projek gesamentlik onderneem om N\uuki en Nama te ontwikkel en te bewaar. Vir die doeleindes van hierdie resensie is slegs die gedrukte weergawe bestudeer.

In die digitale era waarin daar 'n klemverskuiwing is van gedrukte na aanlyn boeke is die publikasie van enige werk, veral 'n gratis naslaanbron van dié aard en omvang, betekenisvol. In die eerste plek is dit 'n tasbare erkenning aan Ꞥxuuleeki Katrina Esau vir haar onbaatsugtige ywer om haar moedertaal feitlik op eie houtjie te bewaar en te bevorder. Die groep akademici en navorsers wat dié naslaanbron die lig laat sien het, sou dit nie sonder haar steun kon doen nie. Terselfdertyd dra die boek daartoe by om die kulturele identiteit en waardigheid van 'n vergete groep mense te herstel en as sodanig is dit 'n nako-

ming van 'n wêreldwye onderneming, naamlik die Verenigde Nasies se Internasionale Dekade van Inheemse Tale (2022–2032), om taalverlies om te keer en nuwe taalgebruikers te bemagtig. Verder vervul die boek 'n opvoedkundige rol en oorbrug dit die gaping tussen akademici, die sprekers van die bedreigde tale en die algemene Suid-Afrikaanse publiek, om sodoende 'n omgewing te skep waarin ons historiese en kontemporêre konteks beter verstaan kan word.

Die voorwerk beslaan 'n indrukwekkende 35 bladsye en bied boeiende, insiggewende leesstof. Die materiaal word deurgaans in Afrikaans en Engels aangebied en wel in die genoemde volgorde, omdat Afrikaans (53,8%) eerder as Engels (3,4%) in die Noord-Kaap as lingua franca gebruik word.

In die Voorwoord verduidelik Nigel Crawhill, die Kanadese taalkundige en afdelingshoof van plaaslike en inheemse kennisstelsels van UNESCO, hoe die ontmoeting met David Kruijer en Elsie Vaalbooi daartoe gelei het dat die oorblywende sprekers van N\uuki opgespoor is en die geskiedkundige projek uiteindelik onderneem kon word.

Hierop volg die Vooraf-gedeelte deur die hoofredakteurs Bonny Sands en Kerry Jones. Interessant is die motivering vir die meertalige aanbod: dit weerspieël die ryk taalkundige en kultuurgeskiedenis van die Noord-Kaap en 'n konteks waarbinne die sprekers in hul daaglikse omgang met mekaar die tale vermeng.

In die gedeelte oor die vroeë dokumentasie van N\uuki word 'n bondige oorsig gegee van die navorsingsbydraes in die vorm van aantekeninge en klankopnames oor jare heen. Dié baanbrekerswerk oor meer as 'n honderd jaar word uiteindelik alles op een plek saamgevat en dit verleen gewigtigheid en geloofwaardigheid aan die naslaanbron.

Die afdeling oor die N\uu-herlewingsverhaal beskryf die proses wat in die negentigerjare van stapel gestuur is en waar taalkundiges en navorsers, veral van Namibië en Amerika, nou saamgewerk het met die oorblywende sprekers om oor 'n periode van meer as 25 jaar storieboeke en leerstof te produseer en uiteindelik die naslaanbron te publiseer. Die omvang en skaal van dié proses laat die verstand duisel en bring opnuut 'n bewondering vir die medewerkers wat die moed van hul oortuiging gehad het om die taak te voltooi. Uitgebreide digitale opnames van die Oostelike en Westelike dialekte van die N\uuki-taal is op 'n moeisame, tydrowende manier bekom en dié klankopnames vorm die basis van die woordeboek. Hiermee is die voortbestaan van N\uuki verseker, maar terselfdertyd het die taal ook 'n skrifbeeld of ortografie verkry wat dit binne die sfeer van geletterdheid plaas.

Kleurfoto's van ~~fxuuleeki~~ Katrina Esau en die nuwe geslag aanleerders van N\uuki, volwasse N\uuki-sprekers wat tot die woordeboek bygedra het en ander deelnemers aan die projek, naamlik verteenwoordigers van die N\uuki-taalowerheid, Nama-sprekers, plaaslike Afrikaanssprekendes, navorsers, taalaktiviste en taalkundiges wat in verskillende hoedanighede by die projek betrokke was, is ingesluit. Hiermee word die nodige erkenning verleen aan elkeen wat 'n opofferende bydrae tot dié grootse projek gelewer het en vir die gebruiker van die

naslaanbron kry die oeroue taal gestalte en word dit verlewendig deur middel van die fotobeelde.

Inligting oor die vier tale wat afwisselend as bron- en doeltale in die woordeboek verskyn, word vervolgens gegee. Dié interessante agtergrondfeite help die gebruiker om die plek van die betrokke tale binne die komplekse Suid-Afrikaanse taallandskap te begryp en onderlinge taalbeïnvloeding beter te verstaan. Die klem val nie op die standaardvorme van Afrikaans en Engels nie, maar op die variëteite daarvan soos dit deur die N\uuki- en Nama-sprekers in die alledaagse omgang gevorm en gebruik is. Veral van belang vir Afrikaanse akademici en leksikograwe is die dokumentasie van Onse Afrikaans, 'n variëteit van die Noord-Kaap. Vir die Woordeboek van die Afrikaanse Taal (WAT) wat saam met die Afrikaanse Taalraad (ATR) werk aan variëteite onder die ProVARIA-vaaandel, is dit byvoorbeeld waardevolle materiaal.

Die volgende afdeling is 'n gebruikersgids en dit bevat toeligting wat spesifiek op die woordeboek-gedeelte gerig is. Dit sluit toeligting in ten opsigte van die inhoud van elke woordeboekinskrywing, rededeel-etiket, afkortings en akronieme, die naslaan van woorde, spelling en uitspraak en simbole wat in die transkripsie van N\uuki-woorde gebruik word. Omdat N\uuki nie 'n skryftradisie het nie, vervul dié afdeling ook die rol van 'n minigrammatika. Dit is 'n waardevolle, onontbeerlike toevoeging tot die inhoud wat die taal vir 'n verskeidenheid gebruikers toeganklik maak.

Die gebruikstoeligting word aangevul met notas oor die N\uuki-grammatika wat aandui in watter opsigte dit van die grammatika van onderskeidelik Afrikaans en Engels afwyk. Ingewikkelde taalkonsepte word op 'n eenvoudige, helder manier verduidelik en bied 'n fassinerende blik op 'n baie komplekse taal.

Die laaste gedeelte van die voorwerk word in beslag geneem deur inligting oor waar en hoe om toegang tot die digitale weergawes van die woordeboek te verkry, asook 'n lys van addisionele hulpbronne wat publikasies, films en digitale versamelings insluit. Dié gegewens is uiters waardevol en nuttig en hiermee word die beskikbare navorsingsmateriaal wat oor die afgelope meer as honderd jaar versamel is, op een plek bymekaargebring. Interessante ensiklopediese inligting is verder vervat in drie lyste wat die volgende insluit: N\uuki-plekname, N\uuki-persoonsname en N\uuki-groetwoorde en ander algemene sinne. Hierdie lyste bevat, waar moontlik, vertaalekwivalente in Nama, Afrikaans en Engels, asook betekenisinligting. Dié gegewens dra daartoe by om, aanvullend tot die woordeboek, die ryke taal- en kulturele erfenis van die #Khomani San te weerspieël.

Die woordeboek-gedeelte beslaan 276 pp. en is verdeel in vier afdelings met opeenvolgend N\uuki (p. 37-92), Nama (p. 93-157), Afrikaans (p. 159-239) en Engels (p. 241-313) as brontale. Vir die doeleindes van hierdie resensie word vervolgens enkele opmerkings vanuit 'n leksikografiese perspektief gemaak en toegelig met Afrikaanse voorbeelde uit die artikels van N\uuki wat die fokus van dié naslaanbron is.

Die uitleg is gebruikersvriendelik met, in die middel bo-aan elke bladsy, 'n opskrif (die naam van die betrokke taal) wat aandui in watter afdeling die gebrui-

ker hom of haar bevind. Bo-aan die linker- en regterkant van elke bladsy word die eerste en laaste trefwoord wat daarop voorkom, aangegee. Hiermee word die vindbaarheid van woorde verhoog. Elke bladsy is in twee kolomme verdeel. Omdat die inligtingaanbod in die artikel van elke trefwoord so dig is, help die korter reëls met leesbaarheid en begrip.

Die trefwoorde is in vetdruk en die doeltale is onderstreep. Aan die reedele is eenvoudige kodes, bv. T1, T2, T3, ens. toegeken wat maklik in die voor- teks nageslaan kan word. Dié stelsel is funksioneel, want anders sou die naam van die betrokke reedele in elk van die vier tale herhaal moes word. In die N\uuki-afdeling word die fonetiese transkripsie vir elke trefwoord tussen vierkantige hakies aangedui. Dit is nie duidelik waarom fonetiese transkripsies vir trefwoorde in Nama, Afrikaans en Engels ontbreek nie. Waarskynlik is dit omdat die fokus van die woordeboek N\uuki is en omdat dit baie tydrowend sou wees om die transkripsies vir die ander tale, veral Nama, by te werk.

Indien daar vir 'n betrokke trefwoord verskillende vertaalekwivalente moontlik is, word hulle deur kommas van mekaar geskei. Die Afrikaans vir N\uuki **flaitsi** word byvoorbeeld aangedui as mondfloitjie, bekfluitjie, harmonika. Dié vertaalekwivalente is sinonieme. By die trefwoord **hogo** word vertaalekwivalente wat nie sinoniem is nie, ook met kommas van mekaar geskei, naamlik spook, duiwel, Satan, nagmerrie, gees, vyand. Dié werkwyse verskil van die konvensie in vertalende woordeboeke om vertaalekwivalente wat betekenismatig verskil, met kommapunte te skei. Dit sou beter wees om by die bestaande konvensie te hou.

Vertaalekwivalente wat in Standaardafrikaans nie dieselfde gevoelswaarde het nie, word ongemerk gelys. By N\uuki **kx'a'u** word *opdrink* en *opsuipe* naas mekaar gegee. Dit is nie duidelik of dit is omdat die betrokke woorde in die Noord-Kaapse variant van Afrikaans waarskynlik dieselfde sosiostilistiese waarde het nie. Hier sou 'n mens beter gebruikseleiding verwag.

Waar enkelwoordvertalings nie moontlik is nie, word verklarende frases gegee. By die N\uuki-trefwoord **tsyoo** staan daar byvoorbeeld: Afr.: sit met een been reguit en by **n'hobo**: Afr.: rituele verfmerke op die gesig, om die gesig soos 'n gemsbok s'n te laat lyk. Dié verklarende frases verskyn as meerwoordige trefwoorde in die Afrikaanse afdeling. Waar nodig word die verklarende frases aangevul met addisionele verklarende aantekeninge en dan is die onderstrepte taalnaam kursief. Vergelyk by dieselfde lemma: Afr.: Die verfmerke is soos dié van 'n gemsbok, bv. wanneer 'n jongmeisie haar mondigwording (hokmeisie) deurloop.

Betekeniswaardes word tussen enkelaanhalingstekens gegee. Die N\uuki-trefwoord **umsiloo** word vertaal met *sleutelbeenholte*. Ter verduideliking word bygevoeg dat die N\uuki-woord letterlik 'asem' + 'gat' beteken. Vir die gebruiker bied dit insig in die manier waarop woordvorming in N\uuki plaasvind.

Addisionele inligting rakende die vertaalekwivalent word tussen ronde hakies verstrek. In die N\uuki-afdeling by die trefwoord **a** word byvoorbeeld die volgende gegee: Afr.: hier (van iets wat naby is) en by die trefwoord **kina** Afr.: hulle (slegs vir mense gebruik). Dié inligting wat die funksie van etikette

vervul, help die gebruiker om die vertaalekwivalent reg te verstaan en te gebruik. Aan die Afrikaanse kant verskyn sowel die trefwoord **hulle** as die etiket (**gebruik vir iets wat nie menslik is nie**) in vetdruk. Dit is strydig met die algemene leksikografiese praktyk.

In sommige artikels is inligting oor die kulturele tradisie en leefwyse van die N\uuki-sprekers ingebed wat die gordyn effens lig op 'n vergete gemeenskap. Die woord **oro** word soos volg vertaal: Afr.: kewer wat 'maan' genoem word Afr.ons: lyfvretertjie, poepgogga. Die volgende inligting word ter agtergrond bygevoeg: Afr.: 'n ongeïdentifiseerde insek wat 'maan' genoem word omdat daar geglo word dat, wanneer 'n vrou daaroor loop, sy haar maandstonde sal kry. Dit is swart, die grootte van 'n R2 muntstuk. Dit het riffels op die rug, 'n klein koppie met voelers en 'n kort nek soos 'n toktokkie. Dit lyk na 'n liewenheersbesie met klein beentjies, is rond en het 'n ronde, klein koppie. Die insek lyk baie soos die *Omorgus asperulatus* en toktokkie.

Waar dit wel voorkom, word vertaalekwivalente in Afrikaans (Afr.) en Onse Afrikaans (Afr.ons) verstrek. By N\uuki **tsaus** word byvoorbeeld Afr.: oog en Afr.ons: puts gegee. Vir die navorser wat in die variëteite van Afrikaans belangstel, is dié dokumentasie van onskatbare waarde.

Homoniemnommers ontbreek deurgaans by inskrywings soos ||**xuri** (T1) Afr.: wildehond en ||**xuri** (T1) Afr.: wortel van die dawidjiesplant. Dit is waarskynlik omdat verhewe syfers in kombinasie met die N\uuki-klikletters verwarrend sou wees. Homoniemnommers ontbreek egter ook aan die Afrikaanse en Engelse kant waar dieselfde argument nie water hou nie.

Die trefwoorde wat in die N\uuki-afdeling nagegaan is, verskyn in die Nama-, Afrikaanse en Engelse afdelings as vertaalekwivalente, en omgekeerd. Dit is verantwoordelike leksikografiese praktyk.

Dié naslaanbron is 'n monumentale werk wat al die lof verdien wat in die media daaraan toegeswaai is. Dit is maklik leesbaar en toeganklik, nie net vir akademici en taalstudente nie, maar enige taaltoesias wat sy of haar horisonne wil verbreed. Die navorsers het met die hulp van gemeenskapsleiers en die oorblywende sprekers van N\uuki uitstekend daarin geslaag om 'n byna vergete taal in boekvorm te rekonstrueer en sodoende vir die nageslag te laat voortleef. In die laaste instansie dien dit as navolgenswaardige voorbeeld van hoe ander bedreigde tale, op eie bodem, maar ook internasionaal, gedokumenteer en bewaar kan word.

A.E. Cloete
Buro van die WAT
Stellenbosch
Suid-Afrika
(aec@sun.ac.za)

Publikasieaankondigings / Publication Announcements

Michael Adams. *Problems in Lexicography: A Critical/Historical Edition*. (Based on Fred W. Householder and Sol Saporta (Editors). *Problems in Lexicography*. 1967.) 2022, xii + 456 pages. ISBN 978-025-306-327-4 (Hardback), ISBN 978-025-306-328-1 (Paperback), ISBN 978-025-306-330-4 (eBook). Bloomington: Indiana University Press. Price \$90.00 (Hardback), Price \$40.00 (Paperback), Price \$39.99 (eBook).

Willem F. Botha en Phillip A. Louw (Hoofredakteurs). *Woordeboek van die Afrikaanse Taal, Sewentiende Deel: T*. 2023. www.wat.co.za. Stellenbosch: Buro van die WAT. R200 vir 12 maande (Aanlyn WAT).

Naomi Bruwer en Gerda Odendaal (Redakteurs). *Lexikos* 32(3) — 'n Man met 'n plan. *Huldeblyk aan / Tribute to Willem F. Botha*. 2022, xiv + 250 pp. ISBN 978-1-990998-50-8, ISSN 2224-0039. AFRILEX-REEKS 32B:2022. Stellenbosch: Buro van die WAT.

John Considine. *Sixteenth-Century English Dictionaries*. 2022, xiv + 482 pages. ISBN 978-0-1988-3228-7 (Hardback). Oxford: Oxford University Press. Price £90.00. (Review in this issue.)

Delana Fourie (Redakteur). *Senior tweetalige skoolwoordeboek / Senior Bilingual School Dictionary. Afrikaans–Engels/English–Afrikaans*. 2023, xii + 628 pp. ISBN 9781868902149 (Sagteband), ISBN 9781868902156 (ePUB). Kaapstad: Pharos Woordeboeke. Prys: R220 (Sagteband).

Annette Klossa-Kückelhaus and Ilan Kernerman (Editors). *Lexicography of Coronavirus-related Neologisms*. 2022, vi + 306 pages. ISBN 978-311-079-556-1 (Hardback), ISBN 978-311-079-808-1 (ePDF), ISBN 978-311-079-831-9 (ePUB). *Lexicographica*. Series Maior 163. Berlin/Boston: Walter de Gruyter. Price €109.95 (Hardback), eBook free. (Review in this issue.)

Bonny Sands en Kerry Jones (Hoofredakteurs). *N|uuki Namagowab Afrikaans English: ǀXoakiǀxanisi / Mǀdi di ǀKhanis / Woordeboek / Dictionary*. 2022, 313 pp. ISBN 978-0-6397-1245-1. Stellenbosch: African Sun Media. Gratis beskikbaar op aanvraag by die Department of Sports, Art and Culture (DSAC) en African Tongue, Suid-Afrika. (Resensie in hierdie nommer.)

Sóstenes Rego (Sisito Matete). *Gwanga la Mafala ya nciNyungwe. Dicionário Geral Monolingue da Língua Nyungwe*. 2023, LII + 795 pp. ISBN 978-989-704-553-0 (Hardback), ISBN 978-989-704-552-3 (eBook), ISSN 1388-8455. LINGUÍSTICA 28. Vila Real, Portugal: Centro de estudos em letras, universidade de trás-os-montes e alto douro. Price €99.00 (Hardback and eBook).

Heming Yong and Jing Peng. *A Sociolinguistic History of British English Lexicography*. 2022, xiii + 254 pages. ISBN 978-1-032-02467-7 (Hardback), ISBN 978-1-032-02468-4 (Paperback), ISBN 978-1-003-18347-1 (eBook). London/New York: Routledge. Price \$130.00 (Hardback), Price \$38.99 (Paperback), Price \$35.09 (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 evaluerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksiko-opname, Leksikobibliografie, Leksikonuus, Lexikofokus, Leksiko-eerbewys, Leksikohuldeblyk, Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)–(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. inkomstegenererende navorsingstydskrif goedgekeur. Dit verskyn ook op die *Institute of Science Index* (ISI).

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

Manuskripte word vir publikasie oorweeg met dien verstande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies. Indien van toepassing, moet besonderhede van die oorsprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

6. Gratis oordrukke en eksemplare

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 versoek kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting. 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 boskritef genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za)
or refer to our website: <http://lexikos.journals.ac.za/>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

- (1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected in an original, synoptic, interpretative, comparative or critically evaluative manner.
- (2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources. Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.
- (3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
- (4) **Projects:** Discussions of lexicographical projects.
- (5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
- (6) **Lexicovaria:** Any of a large variety of articles containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
- (7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index* (ISI).

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordboek van die Afrikaanse Taal. Authors are free, however, to use their

material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

6. Free offprints and copies

Lexikos is only available electronically on <http://lexikos.journals.ac.za> from volume 28 onward. No offprints or copies are available.

7. Invitation and editorial address

All interested authors are invited to submit contributions, preferably in electronic format, for publication in *Lexikos* to: lexikos@sun.ac.za, or

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH, Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Format and length of articles

Manuscript should preferably be submitted in electronic format by email or on a disk, in software compatible with MS Word. The typeface used should preferably be 10-point Palatino or Times Roman. Contributions should not exceed **8 000 words**.

Each article must be accompanied by **abstracts** of approximately 200 words and approximately 10 **keywords** in the language in which it is written, as well as in **English**. English articles of South African origin should carry an abstract and keywords in Afrikaans, whilst English articles of foreign origin should carry a second abstract and keywords in any of the other languages mentioned. In cases where this is not done, the editors will provide an Afrikaans version. Ensure that the abstract in the second language also contains a **translation of the original title**.

2. Graphics

Figures such as tables, graphs, diagrams and illustrations should be in an appropriate size to be well accommodated within the page size of *Lexikos*, namely 18 cm high by 12 cm wide. The locations of figures within the text must be clearly indicated. If orthographic marks or graphics used in the text prove problematic, a printout of the manuscript or an email in .pdf format may be requested.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details. 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**.