http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Lexikos 35

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Lexikos 35

Hoofredakteur / Editor-in-Chief
André H. du Plessis

Mederedakteurs / Associate Editors
Dion Nkomo
Elsabé Taljard
Michele van der Merwe
Sonja Bosch (Duits/German)
Steve Ndinga-Koumba-Binza (Frans/French)

Resensieredakteur / Review Editor
Dané Claassen

African Association for Lexicography

AFRILEX-REEKS 35:2025
AFRILEX SERIES 35:2025

BURO VAN DIE WAT

STELLENBOSCH

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2025 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2025

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

ISBN 978-1-990998-54-6
ISSN 2224-0039

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

Licenced under Creative Commons License CC BY 4.0.

Menings wat in artikels en resensies uitgespreek word, is nie noodwendig dié van AFRILEX of die Buro van die WAT nie.
Opinions expressed in the articles and reviews are not necessarily those of AFRILEX or of the Bureau of the WAT.

Lexikos is elektronies beskikbaar by http://lexikos.journals.ac.za/
Lexikos is available online at http://lexikos.journals.ac.za/

Lexikos is elektronies beskikbaar by Sabinet, AJOL, Ebsco en Proquest
Lexikos is available online from Sabinet, AJOL, Ebsco and Proquest

Indekse Indexes
Arts and Humanities Citation Index® (Clarivate); Asian Digital Library; Current Contents®/Arts & Humanities (Clarivate); Current Contents®/Social and Behavioral Sciences (Clarivate); ERIH Plus; EuroPub Index; Index Copernicus Journals Master List; Journal Citation Reports/Social Sciences Edition (Clarivate); Linguistic Bibliography Online (Brill); Linguistics Abstracts Online; Linguistics and Language Behavior Abstracts; MLA International Bibliography; R.R.K. Hartmann's Bibliography of Lexicography; SciELO SA; Scopus (Elsevier); Social Sciences Citation Index® (Clarivate); Social Scisearch® (Clarivate)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Adviesraad / Advisory Board

Prof. H.L. Beyer (Namibië/Namibia)
Prof. G.-M. de Schryver (België/Belgium)
Prof. M.J. Domínguez Vázquez (Spanje/Spain)
Prof. P.A. Fuertes-Olivera (Spanje/Spain)
Prof. R.H. Gouws (RSA)
Prof. R. Lew (Pole/Poland)
Prof. T. Margalitadze (Georgië/Georgia)
Prof. P.A. Mavoungou (Gaboen/Gabon)
Prof. D.J. Prinsloo (RSA)
Prof. S. Tarp (Denemarke/Denmark)
Prof. H. Xu (China)

Redaksiekomitee / Editorial Committee

Prof. N. Bosman (RSA)
Dr. E.M. Ella (Gaboen/Gabon)
Prof. Y. Gao (China)
Prof. L. Khumalo (RSA)
Dr. P.A. Louw (RSA)
Dr. L. Mbindi Aninga (Gaboen/Gabon)
Dr. L. Morris (RSA)
Prof. E. Ndlovu (Zimbabwe)
Dr. G. Odendaal (RSA)
Dr. R.N. Ossai (Nigerië/Nigeria)
Dr. A. Ostroški Anić (Kroasië/Croatia)
Prof. T.J. Otlogetswe (Botswana)
Prof. A.N. Otto (RSA)
Dr. S. Wolfer (Duitsland/Germany)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Inhoud / Contents

Voorwoord x
Foreword xiii
André H. du Plessis

'n Woord van AFRILEX xvi
A Few Words from AFRILEX xvii
Langa Khumalo

Redaksionele doelstellings xviii
Editorial Objectives xix

Artikels / Articles

The Need for a National Terminology Policy for South Africa 1
Mariëtta Alberts

Analyse des Termes Émergents en Épidémiologie Post-COVID-19 68
Ma'moun Alshtaiwi

Digital Lexicography in ESP: Enhancing Vocabulary Acquisition and Reading Comprehension in Students of Commerce Degree 109
Beatriz Chaves-Yuste and Manuel Márquez-Cruz

A Comparative Study on the Effectiveness of AI Chatbots and Dictionary Apps for Lexical Tasks and Retention 157
Yuzhen Chen

La Lexicographie Scolaire comme Fondement de la Lexicographie Moderne au Gabon 209
Edgard Maillard Ella

Terminologiebestuur binne die Suid-Afrikaanse konteks 439
Michelle Goosen en Elsabé Taljard

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Avoiding Information Overload — Ways of Exemplifying in Bilingual e-Dictionaries Depending on the Systematicity of the Linguistic Data 183
Alfonso Rascón Caballero

Endemann's *Wörterbuch der Sotho Sprache* (1911): A Worthy Candidate for Digitisation 462
Elsabé Taljard, Gertrud Faaß, Danie Prinsloo and Sonja Bosch

Traitement des Verbes Auxiliaires Modaux Polysémiques dans le Dictionnaire Chinois–Français Basé sur le Prototype et le Schéma 256
Zhijie Tang

Leksikofokus / Lexicofocus

Introduction to the Special Section of Papers from the 6th GLOBALEX Workshop on Lexicography and Neology 286
Annette Klosa-Kückelhaus and Ilan Kernerman

Lexicography and Neology: Building Bridges through a Multilingual Approach with ENEOLI 294
Giovanni Tallarico, Rute Costa, Ana Salgado, Judit Freixa, Kris Heylen, John Humbley, Ilan Kernerman, Annette Klosa-Kückelhaus, Ana Ostroški Anić, Federica Vezzani, Vincent Balnat, Petar Božović, Špela Arhar Holdt and Weronika Woźniak-Żak

Neologisms and Their Functions in Critical Discourse 318
Petra Storjohann

The Role of Combining Forms in Creating New English Compounds: Data-Driven and Construction Approaches 344
Jin-hong Huang and Yongwei Gao

Semantic Verbal Neology in Spanish Dictionaries: Case Study of the Lexicographic Representation of *anestesiar* and *coreografiar* 367
Marina Berri and Victoria de los Ángeles Boschiroli

Digging for *-ings*: A Survey of Selected (Pseudo)-Anglicisms in Dictionaries of Modern Greek 385
Anna Vacalopoulou

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Semi-Automatic Detection of New Words in Modern Georgian 399
Tamar Laluashvili and Tinatin Margalitadze

Detection and Description of Neologisms in Korean Lexicography: Methodological Issues in Corpus Balance, Word Unit Bias and LLM Assistance 414
Kilim Nam, Soojin Lee and Hae-Yun Jung

Projekte / Projects

Krajšavar — An Algorithm for Automatic Recognition of Abbreviations in English Dictionary Entries Compiled in a Dictionary of Abbreviations 129
Mojca Kompara Lukančič

LEXICC: The Design and Development of an Online Dictionary Writing System 83
Udiluz del Carmen Monsalve Muñoz, Johnatan E. Bonilla, Ruth Yanira Rubio López and Andrés Steban Luna Cortés

The S*erbian–Russian and Russian–Serbian Idiomatic Dictionary for Expressing Emotions*: Theoretical and Practical Aspects 233
Milena Mrkaić

Leksikonotas / Lexiconotes

The Effectiveness of the "Real-Life-Based School Dictionary" as an Electronic Dictionary 143
Ayça Çaglayan Özkan

An Analysis of Word-Formation Processes in Neologisms: A Case Study of the *Oxford Advanced Learner's Dictionary* 26
Aling Shi

Prompts for Language Learners: A Practical Guide to Using DeepSeek as a Dictionary 274
Ai Zhong

Leksikohuldeblyk / Lexicotribute

Reinhard R.K. Hartmann 1938 – 2024 44
Danie J. Prinsloo en/and Rufus H. Gouws

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Resensies / Reviews

Edward Finegan and Michael Adams. *The Cambridge Handbook of the Dictionary* 57
Huahua Jiang and Hai Xu

Yongwei Gao. *A Dictionary of Contemporary English Neologisms and Trendy Words* 199
Aling Shi

Du Ruiqing and Jiang Yajun. *New Century Chinese–English Dictionary* 48
Yongwei Gao

Stephen Turton. *Before the Word was Queer: Sexuality and the English Dictionary, 1600–1930.* 62
Xiaotong Zhao

Publikasieaankondigings / Publication Announcements 482

Voorskrifte aan Skrywers 483
Instructions to Authors 484

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Voorwoord

In 1989 het 'n formele leksikografiese gesprek op Stellenbosch tussen die redaksie van die *Woordeboek van die Afrikaanse Taal* (WAT) en 'n aantal gerekende taalkundiges plaasgevind. Hierdie sleutelfigure het besef dat daar 'n groot behoefte aan die uitruil van leksikografiese kennis en ondervinding in Afrika bestaan. Dit het uiteindelik gelei tot die totstandbrenging van *Lexikos*, 'n uitstekende medium en voertuig vir leksikografiese gesprekvoering. *Lexikos* 1, 'n stuk leksikografiese pionierswerk, verskyn in 1991 onder die stuur van wyle mnr. Pieter Harteveld. Vir die eerste keer het Afrika 'n vaktydskrif wat hoofsaaklik op die leksikografie gerig is. Hierdie tydskrif het 35 jaar later kennelik een van die vernaamste internasionale vaktydskrifte vir en oor die leksikografie geword. Die menigte toonaangewende *Lexikos*-bydraes oor die laaste drie en 'n half dekades het 'n blywende en beduidende impak op die leksikografie as vakrigting gehad. Die redaksie, uitgewer (die WAT) en AFRILEX is geweldig trots op die produk wat oor die afgelope 35 jaar uitgegee is. Mag *Lexikos* vir die volgende 35 jaar die leksikografie ter wille wees en mag daar nog vele volumes suksesvol verskyn!

Soos met die publikasie van *Lexikos* 1, het die 35ste volume ook vir uitdagings en moeilike besluite gesorg. Sedert 2022 het die aantal bydraes wat ingelewer is, stelselmatig vermeerder, maar in 2025 het die spreekwoordelike sluise behoorlik oopgetrek. Die redaksie het watergetrap terwyl daar 'n plan geformuleer moes word om al die bydraes te verwerk en te publiseer. Daar is uiteindelik, met verlof van die uitgewer, besluit om vanjaar twee uitgawes te publiseer. Hierdie geweldige toename in bydraes getuig van die reikwydte en aansien van *Lexikos*. Dit is verblydend dat so baie outeurs *Lexikos* as geskikte tuiste vir hul navorsingsuitsette ag. Dit plaas egter ongewone druk op die redaksie, redaksionele steundienste en uitgewer. Die toename in artikels gepaard met *Lexikos* se huidige redaksionele en finansiële model het aanleiding gegee tot sinvolle en nodige gesprekke tussen AFRILEX, die redaksie en die uitgewer oor die volhoubare bestuur van die tydskrif. Die besluit is solank geneem dat *Lexikos* 35 die laaste volume sal wees wat bladgeld slegs van Suid-Afrikaanse outeurs eis. Vanaf 2026 sal die eis van publikasiefooie ("article processing charges") van outeurs buite Suid-Afrika geïmplementeer word. *Lexikos* se missie is wel om steeds te verseker dat kennis gratis met die leksikografiese gemeenskap gedeel word. Derhalwe bly *Lexikos* 'n ooptoegangtydskrif vir alle lesers.

Vanjaar se eerste uitgawe bevat 'n spesiale afdeling wat toegewy is aan die *6th GLOBALEX Workshop on Lexicography and Neology*. Hierdie werkswinkel is met groot sukses by die *28th International Conference of the African Association for Lexicography* aangebied. Sewe uit die elf referate is in artikels omskep en bied interessante en unieke perspektiewe op die skep, gebruik en leksikografiese ontginning van neologismes die wêreld oor. Ons is dankbaar vir die samewerking

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

met GLOBALEX en waardeer die redaksionele en koördineringswerk wat die gasredakteurs van hierdie afdeling, prof. Annette Klosa-Kückelhaus en mnr. Ilan Kernerman, behartig het.

Die ander bydraes in vanjaar se uitgawes is, soos verwag, verteenwoordigend van die verskeidenheid navorsingmoontlikhede in die leksikografie en terminologie. Daar is insiggewende en waardevolle bydraes oor onder andere terminologiebestuur, kunsmatige intelligensie, aanlyn woordeboeke, die metaleksikografie en inklusiewe leksikografie. Naas ons outeurs uit verskillende dele van Afrika, soos Botswana, Gaboen en Tunisië, is daar stemme uit, onder andere, Australië, China, Litaue, Montenegro, Pole en Spanje. Baie dankie aan elke outeur wat deur middel van hul publikasies in *Lexikos* aan die leksikografiese diskoers deelneem. Die redaksie is ewig dankbaar vir al die outeurs se samewerking en positiewe gesindhede.

Die publisering van *Lexikos* is altyd 'n spanpoging en die span wat 'n aandeel in vanjaar se uitgawes gehad het, is uiters bekwaam. 'n Spesiale woord van dank aan proff. Elsabé Taljard, Dion Nkomo en Michele van der Merwe vir die uitstekende werk wat hulle as mederedakteurs gedoen het, asook vir die waardevolle raad en bystand wat hulle gebied het. Baie dankie ook aan dr. Steve Ndinga-Koumba-Binza en prof. Sonja Bosch vir hulle volgehoue en uiters knap werk as Franse en Duitse redakteurs. Baie dankie aan dr. Dané Claassen vir die uitsonderlike werk met die resensies. Dit is 'n voorreg om op so 'n behendige redaksie te kan staatmaak. Aan mee. Tanja Harteveld en Hermien van der Westhuizen van die WAT, wil ek namens die redaksie opreg dankie sê vir hulle toewyding aan die stiptelike en professionele set en afronding van vanjaar se uitgawes, die administratiewe bystand, asook vir hulle bekwaamheid met al die ad hoc-take wat deur die loop van die jaar opgeduik het. So ook bedank ek namens die redaksie ons nuutste spanlid, me. Maryna Adshade. As redaksionele assistent het me. Adshade haar ontsettend goed van haar taak gekwyt en ons is innig dankbaar vir haar bereidwilligheid, flinkheid, toewyding en ondersteuning. 'n Hartlike bedanking ook aan dr. Phillip Louw, Uitvoerende Direkteur van die WAT, vir sy toegemoetkomendheid en toewyding aan *Lexikos*.

Lexikos is groot dank verskuldig aan die kranige portuurbeoordelaars wat verseker het dat vanjaar se bydraes die hoë wetenskaplike standaard van hierdie tydskrif handhaaf.

Ons het vanjaar met groot leedwese verneem van die heengaan van dr. Frikkie Lombard, voormalige eindredakteur van die WAT. Dr. Lombard het diep spore in die Suid-Afrikaanse, veral die Afrikaanse leksikografie gelaat. Dr. Lombard het 42 jaar die WAT met onderskeiding gedien, en 25 jaar in die Taalkommissie. Hy het in daardie tyd 'n blywende merk op die Afrikaanse taallandskap gemaak, veral deur die 10 WAT-dele waarby hy betrokke was en die betekenisvolle bydrae wat hy gemaak het om die Afrikaanse spelreëls te help standaardiseer. Dr. Lombard was ook betrokke by leksikografiese aktiwiteite buite die WAT, onder andere, as medewerker aan woordeboeke uitgegee deur OUPSA en gasdosent aan die Universiteit Stellenbosch. Hy het ook vier *Lexikos*-bydraes gelewer.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Dr. Lombard se toewyding aan die leksikografie, veral die praktyk, sowel as aan Afrikaans, is byna ongeëwenaard en sy heengaan is 'n enorme verlies vir ons vakgebied en vir die Afrikaanse taalkunde. Ons wil graag ons meegevoel aan elkeen van sy geliefdes, vriende, kollegas en medewerkers oordra.

Laastens wil ek graag my opregte waardering aan die AFRILEX-raad, die WAT en my mederedakteurs oordra vir hulle deurlopende vertroue en ondersteuning. Baie dankie ook aan AFRILEX vir die finansiële bystand wat die assosiasie aan *Lexikos* verleen. Ons is ook opreg dankbaar vir die goedgunstige en stewige finansiële bydrae van die Fonds Neerlandistiek.

André H. du Plessis
Hoofredakteur

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

Foreword

In 1989, a formal lexicographic discussion took place in Stellenbosch between the editorial staff of the *Woordeboek van die Afrikaanse Taal* (WAT — Dictionary of the Afrikaans Language) and a group of prominent linguists. These key figures realised that there existed a great need for the exchange of lexicographical information and experience in Africa. This ultimately led to the establishment of *Lexikos,* an excellent medium and vehicle for lexicographic discussion. *Lexikos* 1, a work of lexicographic pioneering, was published in 1991 under the editorship of the late Mr Pieter Harteveld. For the first time, Africa had a journal that was primarily focused on lexicography. This journal has, 35 years later, undoubtedly become one of the most prominent journals for and about lexicography. The multitude of trendsetting *Lexikos* contributions over the last three and a half decades have had a lasting and significant impact on lexicography as a discipline. The editorial board, publisher (the WAT) and AFRILEX are immensely proud of the product that has been published over the past 35 years. May *Lexikos* benefit lexicography for the next 35 years and may many more volumes be successfully published!

As with the publication of *Lexikos* 1, the 35th volume also brought along challenges and tough decisions. Since 2022, the number of contributions submitted has systematically increased, but in 2025 the proverbial floodgates opened. The editorial team treaded water while trying to devise a plan to process and publish all the contributions. It was ultimately decided, with permission from the publisher, to publish two issues this year. This tremendous increase in contributions speaks to the reach and prestige of *Lexikos*. It is gratifying that so many authors consider *Lexikos* to be a suitable venue for their research outputs. However, this places unusual pressure on the editorial team, editorial support services and publisher. The increase in articles coupled with *Lexikos'* current editorial and financial model led to meaningful and necessary discussions between AFRILEX, the editorial team and the publisher about the sustainable management of the journal. The decision has been made that *Lexikos* 35 will be the last volume to charge page fees only to South African authors. From 2026 article processing charges will be implemented for authors outside South Africa. *Lexikos'* mission is to continue to ensure that knowledge is freely shared with the lexicographic community. Therefore, *Lexikos* remains an open access journal for all readers.

This year's first issue contains a special section dedicated to the *6th GLOBALEX Workshop on Lexicography and Neology*. This workshop was presented with great success at the *28th International Conference of the African Association for Lexicography*. Seven of the eleven papers were converted into articles and offer interesting and unique perspectives on the creation, use and lexicographic extraction of neologisms around the world. We are grateful for the collaboration with

GLOBALEX and appreciate the editorial and coordination work carried out by the guest editors of this section, Prof Annette Klosa-Kückelhaus and Mr Ilan Kernerman.

The other contributions in this year's issues are, as expected, representative of the variety of research avenues in lexicography and terminology. There are insightful and valuable contributions on, among others, terminology management, artificial intelligence, online dictionaries, metalexicography and inclusive lexicography. In addition to our authors from different parts of Africa, such as Botswana, Gabon and Tunisia, there are voices from, among others, Australia, China, Lithuania, Montenegro, Poland and Spain. A big thank-you to the authors who participate in the lexicographic discourse via their publications in *Lexikos*. The editorial team is eternally grateful for all the authors' cooperation and positive attitudes.

The publication of *Lexikos* is always a team effort, and the team that had a part in this year's issues is an extremely competent one. A word of thanks to Profs Elsabé Taljard, Dion Nkomo and Michele van der Merwe for the outstanding work delivered as associate editors, as well as for all their valuable guidance and assistance. Many thanks also to Dr Steve Ndinga-Koumba-Binza and Prof Sonja Bosch for their continued and extremely deft work as French and German editors. A sincere thank you to Dr Dané Claassen for the exceptional work with the reviews. It is a privilege to be able to rely on such a highly capable editorial team. On behalf of the editors, I would like to sincerely thank Mss Tanja Harteveld and Hermien van der Westhuizen from the WAT, for their devotion in promptly and professionally typesetting and finalising this year's issues, for their proficiency with the overall production of the journal, their administrative support and for their efficiency with all the ad hoc tasks that popped up during the year. I would also like to thank our newest team member, Ms Maryna Adshade. As editorial assistant, Ms Adshade has done a fantastic job and we are deeply grateful for her willingness, efficiency, dedication and support. A sincere thank you to Dr Phillip Louw, the Executive Director of the WAT, for his cooperativeness and commitment to *Lexikos*.

Lexikos is deeply indebted to each adjudicator that assisted with this year's peer review process, ensuring that our exacting scientific standards are maintained.

This year we learnt with heavy hearts about the passing of Dr Frikkie Lombard, former final editor of the WAT. Dr Lombard left a deep mark on the South African, especially Afrikaans, lexicography. Dr Lombard served the WAT with distinction for 42 years, and 25 years on the Taalkommissie (the South African Academy for Science and Arts' Language Commission). During his tenure, he made a lasting mark on the Afrikaans landscape, especially through the 10 WAT volumes on which he worked and the significant contribution he made to helping standardise the spelling rules of Afrikaans. Dr Lombard was also involved in lexicographic activities outside the WAT, among others, as a collaborator on dictionaries published by OUPSA and a guest lecturer at Stellenbosch University.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

He also contributed four articles to *Lexikos*. Dr Lombard's dedication to lexicography, especially its practice, as well as to Afrikaans, is almost unparalleled and his passing is an enormous loss for our field and for Afrikaans linguistics. We would like to extend our condolences to each of his loved ones, friends, colleagues and associates.

Finally, I would like to express my sincere appreciation to the AFRILEX board, the WAT and my fellow editors for their continued trust and support. A big thank you to AFRILEX for the financial aid provided to *Lexikos*. We are also sincerely grateful for the kind and substantial financial contribution of the Fonds Neerlandistiek.

André H. du Plessis
Editor-in-Chief

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

'n Woord van AFRILEX

Na aanleiding van 'n goed georganiseerde 29ste kongres van die *African Association of Lexicography* (AFRILEX) wat by die Universiteit van die Wes-Kaap gehou is, wens die uittredende AFRILEX-raad die *Lexikos*-span, onder leiding van die hoofredakteur, André du Plessis, en die uitgewer, die Buro van die WAT, geluk met die suksesvolle publikasie van hierdie bundel. Die Raad bly deurgaans bewus van die deeglike werk wat deur die *Lexikos*-redakteurs gedoen word. Hul volhardingsvermoë, stiptheid, noukeurigheid en deskundigheid in die veld sorg dat die hoë standaarde van die tydskrif gehandhaaf word.

Lexikos bly 'n belangrike mondstuk vir die toonaangewende navorsing van AFRILEX-lede, beide in Afrika en globaal. Die kongrestema vir die 29ste AFRILEX-kongres was **"Die rol van kunsmatige intelligensie en digitalisering in die toekoms van leksikografie"**, wat baie interessante aanbiedings ontlok het wat die kruispunt van kunsmatige intelligensie en digitale gereedskap met die leksikografiese praktyk ondersoek het. Dit is 'n tydige en toekomsgerigte kongrestema aangesien kunsmatige intelligensie steeds in heelwat dissiplines oor die akademiese spektrum ontwrig en geleenthede bied.

Dit is ook noemenswaardig dat sommige van die navorsing wat by die 29ste AFRILEX-kongres aangebied is, gefokus het op navorsing wat in talle Suid-Afrikaanse tale gedoen is. Die toenemende aantal samewerkings oor instellings en fokusareas heen is 'n ware weerspieëling van 'n gedeelde doelwit om positief en met trefkrag by te dra tot die bevordering van ons veld. Hierdie inter- en kruisdissiplinêre aard kan slegs die diepte en breedte van die leksikografie as 'n dissipline verbeter. Ek hoop dat daardie interessante aanbiedings in artikels omskep word en in toekomstige volumes van *Lexikos* gepubliseer word.

Ek wil weereens 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 'n sleutel- en strategiese vennoot van AFRILEX. Namens die AFRILEX-Raad, en ook die algemene AFRILEX-lede wêreldwyd, wil ek die *Lexikos*-span, die Buro van die WAT en die outeurs wat bygedra het tot hierdie uitgawe, opreg bedank. Dit is nog 'n noemenswaardige bydrae tot die leksikografie as wetenskap.

Langa Khumalo
President: AFRILEX

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

A Few Words from AFRILEX

Emerging from a successfully organised 29th African Association of Lexicography (AFRILEX) conference held at the University of the Western Cape, the outgoing AFRILEX Board congratulates the *Lexikos* team, under the stewardship of the Editor-in-Chief, André du Plessis, and the publisher, the Bureau of the WAT, for the successful publication of this volume. The Board remains mindful of the sterling work that is done by the *Lexikos* editors. Their tenacity, stringency, meticulous attention to detail, and expertise in the field has helped maintain the very high standards of the journal.

Lexikos remains an important academic mouthpiece of the cutting-edge research that is done by the AFRILEX members both in the continent and globally. The conference theme for the 29th AFRILEX conference was **'The Role of Artificial Intelligence and Digitalisation in the Future of Lexicography'**, which drew very interesting presentations that explored the intersection of artificial intelligence and digital tools with lexicographic praxis. The conference theme was timely and forward-looking, as artificial intelligence continues to disrupt and present opportunities in many disciplines across the academic spectrum.

It is also notable that some of the research presented at the 29th AFRILEX conference focused on research done in numerous South African languages. The increased number of collaborations across institutions and focus areas is a true reflection of a shared goal to contribute positively and impactfully to the advancement of our field. This inter- and cross-disciplinarity can only enhance the depth and breadth of lexicography as a discipline. I hope to see those interesting presentations developed into articles and published in future volumes of *Lexikos*.

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 world-wide, I sincerely thank the *Lexikos* team, the Bureau of the WAT, and the contributing authors to this *Lexikos* volume. It is another important contribution to the scholarship of lexicography.

Langa Khumalo
President: AFRILEX

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

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 interdissiplinêre 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.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

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/10.5788/35-1-2067 (Volle uitgawe / Full volume)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

The Need for a National Terminology Policy for South Africa

Mariëtta Alberts, *Former Director: Terminology Development and Standardisation, Pan-South African Language Board, South Africa (albertsmarietta@gmail.com)*

Abstract: The language policy of a country has an influence on terminology development while it also determines the status of a language and the number of languages to be developed. Although several international and national terminology policies are readily available terminology work is currently not practised in South Africa according to a national terminology policy. The point of departure in terminology work is always conceptual. It is, therefore, important to work according to a subject-oriented terminography approach to harvest terminology. This approach is usually used simultaneously, and in conjunction with other approaches, such as language planning-oriented, translation-oriented, linguistic community-oriented, process-oriented, functional, and interactive approaches to develop the official languages of South Africa. The hegemony of English, however, hampers terminology development in the other official languages. There is a lack of coordination and cooperation among terminology agencies which lead to duplication of projects. The national terminology office, i.e. the Terminology Coordination Section (TCS), National Language Service, Department of Sport, Arts and Culture, is the main beneficiary of a national terminology policy, and it should provide guidelines to other terminology agencies. Terminology policies and the need for a national terminology policy for South Africa are described and the eventual roadmap for the acceptance of the national terminology policy recently drafted, is supplied. A national terminology policy could assist in promoting the official, provincial, and local languages to become functional languages in all spheres of human activity.

Keywords: BILINGUAL, COPYRIGHT, LANGUAGE FOR SPECIAL PURPOSES, LANGUAGE PLANNING-ORIENTED TERMINOGRAPHY APPROACH, LINGUISTIC COMMUNITY-ORIENTED TERMINOGRAPHY APPROACH, MULTILINGUAL, POLICY, POLYTHEMATIC, REGISTER, SUBJECT-ORIENTED TERMINOGRAPHY APPROACH, TERM BANK, TERMINOGRAPHY, TERMINOLOGY, TRANSLATION-ORIENTED TERMINOGRAPHY APPROACH

Opsomming: Die behoefte aan 'n nasionale terminologiebeleid vir Suid-Afrika. Die taalbeleid van 'n land het 'n invloed op terminologieontwikkeling terwyl dit ook sowel die status van 'n taal as die aantal tale wat ontwikkel moet word, bepaal. Hoewel verskeie internasionale en nasionale terminologiebeleide geredelik beskikbaar is, word terminologiewerk nie tans in Suid-Afrika volgens 'n nasionale terminologiebeleid beoefen nie. Die vertrekpunt by terminologiewerk is altyd konseptueel. Dit is gevolglik belangrik om volgens 'n vakgeörienteerde terminografiebenadering terminologie te versamel. Hierdie benadering word gewoonlik saam met ander benaderings, soos taalbeplannings-, vertaalgeörienteerde, taalgemeenskaps-, prosesgeoriënteerde, funksionele

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

en interaktiewe benaderings gebruik om die amptelike Suid-Afrikaanse tale te ontwikkel. Die hegemonie van Engels bemoeilik egter die ontwikkeling van terminologie in die ander amptelike tale. Daar is 'n gebrek aan koördinering en samewerking onder terminologiese instansies wat lei tot die duplisering van projekte. Die nasionale terminologiekantoor, dit is die Terminologiekoördineringsafdeling (TKA — in die praktyk word slegs die Engelse benaming en afkorting gebruik), Nasionale Taaldiens, Departement van Sport, Kuns en Kultuur is die hoofbegunstigde van 'n nasionale terminologiebeleid en behoort riglyne aan ander terminologiese instansies te verskaf. Die uiteindelike roete om die konsep van die nasionale terminologiebeleid te aanvaar, word verskaf. 'n Nasionale terminologiebeleid kan help om die amptelike, provinsiale en plaaslike tale in funksionele tale te ontwikkel wat in alle gebruiksfere van menslike aktiwiteit benut kan word.

Sleutelwoorde: BELEID, TAALGEMEENSKAPSGEÖRIENTEERDE TERMINOGRAFIEBENADERING, MEERTALIG, REGISTER, TAALBEPLANNINGSGEÖRIENTEERDE TERMINOGRAFIEBENADERING, TERMBANK, TERMINOGRAFIE, TERMINOLOGIE, TWEETALIG, OUTEURSREG, VAKGEÖRIENTEERDE TERMINOGRAFIEBENADERING, VAKTAAL, VEELTEMATIES, VERTAALGEÖRIENTEERDE TERMINOGRAFIEBENADERING

1. Introduction

The South African terminology practice has an interesting history mainly influenced by different language policies during the various political dispensations. All the indigenous languages (Afrikaans and the Sintu languages) that later became official languages, had to be reduced to writing, had to be standardised, and their spelling, orthographies and word-formation principles had to be developed before any terminology development could start. The previous bilingual policy favoured the development of Afrikaans terminology. Government policy on language promotion prior to 1994 furthermore stipulated that the promotion of the Sintu languages was the task of the former national states and the self-governing regions and not of the South African government. Since 1995 structural and policy considerations have had an important effect on the practice of terminology in South Africa. It has been accepted as a responsibility of the state to also develop and promote the official Sintu languages and, since it became an official language in 2023, also South African Sign Language (SASL).

The terminologist working in a multilingual society such as South Africa is faced with conflicting situations: the multilingual polythematic terminologies should be developed to enhance the multilingual heritage of the country; however, the number of languages makes this impractical and not economically viable and, to add to the problem, few trained terminologists and even fewer terminology posts are available. The development of SASL terminology also poses a big challenge — the conceptual content of various subject fields needs to be depicted by signs and captured as video clips for the dissemination thereof to the deaf and hearing-impaired community.

Terminology plays a crucial role in language development and the promotion of multilingualism. This article discusses the need for a national termi-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

nology policy as seen from a South African perspective. Special emphasis is given to the South African situation regarding language policies and consequential terminology approaches in terms of subject-oriented terminography, translation-oriented terminography, language planning-oriented terminography, and linguistic community-oriented terminography.

Terminology is never practiced in isolation and various collaborating bodies assist the national terminology office with terminology development. The aspect of training in the principles and practice of terminology and terminography receives attention, and the influence of copyright and human language technologies on terminology development is discussed.

2. Language planning and language policy

Terminology as a discipline is governed by the language policy of the country (cf. Alberts 2003: 134; Wright 2007: 6; Alberts 2008; Alberts 2010: 599-620; Alberts 2017: 149, 154-157). It is at the service of the language policy of the reigning government, e.g. monolingual, bilingual, or multilingual. The South African situation could be taken as example (cf. Alberts 2017: 156-157):

- previous dispensation (i.e. prior to 1994): bilingual technical dictionaries were compiled (English/Afrikaans);
- present dispensation (i.e. after 1994): multilingual term lists are compiled in the eleven official languages of the country (English, Afrikaans, Setswana, Sesotho, Sepedi/Sesotho sa Leboa, isiNdebele, Siswati, isiXhosa, isiZulu, Xitsonga, Tshivenda).
- since 2023 the South African Sign Language was added as official language and concepts need to be depicted as signs through medium of video clips.

South Africa was, prior to 1994, constitutionally a bilingual country. This meant that special emphasis was placed on the development of English and Afrikaans terminology (cf. Official Languages of the Union Act 8 of 1925; Republic of South Africa Constitution Act 32 of 1961; Republic of South Africa Constitution of Act 110 of 1983). After unification, the Constitution (1996) (cf. Constitution of the Republic of South Africa Act 200 of 1993 (Interim Constitution); Constitution of the Republic of South Africa Act 108 of 1996 (now Constitution, 1996) provides for the equitable use of all the official languages of the country, and for the promotion and development of the historically marginalised indigenous languages (cf. Government Gazette 2012, 2013). According to section 6(1) of the Constitution, the official languages of South Africa are Sepedi/Sesotho sa Leboa, Sesotho, Setswana, Siswati, Tshivenda, Xitsonga, Afrikaans, English, isiNdebele, isiXhosa, isiZulu and, since 2023, South African Sign Language (SASL). Apart from the 12 official languages, the Constitution also recognises other languages such as Khoi and San which should be promoted and developed (cf. Figure 1). The Constitution further requires all official languages to enjoy parity of esteem and be treated equitably.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

The 2022 census data (cf. Statistics South Africa 2022) regarding language usage in South Africa could be summarised as follows (cf. Figure 1):

Language	**Percentage of speakers**
Afrikaans	10.6
English	8.7
isiNdebele	1.7
isiXhosa	16.3
isiZulu	24.4
Sepedi/Sesotho sa Leboa	10
Sesotho	7.8
Setswana	8.3
SASL	0.02
Siswati	2.89
Tshivenda	2.6
Xitsonga	4.7
Khoi, Nama, San	0.01

Figure 1: South African Census data 2022 on language usage

A direct relationship exists between language planning, language policy, language practice and terminology development and management (Advisory Panel 2000; Alberts 2003: 134). Terminology occupies a special position within the framework of language and communication, and terminology development is dependent on language planning and language policies (Alberts 2017: 154).

3. Approaches to terminology work

There are various approaches to terminology work that influence terminology development. Most of these approaches are used simultaneously in South Africa.

3.1 Subject-oriented terminography approach

The subject-oriented terminography approach is the traditional way of harvesting and developing terminology. It is usually the methodology used by large terminology offices worldwide, and it is also the traditional method of terminology documentation followed in South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

The point of departure in terminology work is the concept. The objective of the subject-oriented terminography approach is therefore to focus on concepts and the relationship that exists between concepts and conceptual systems and the terms denoting these concepts. There is a one-to-one relationship between concept and term and each concept should preferably be denoted by a specific term in a specific subject area and language. This practice leads to the standardisation of concepts and terms. Standardisation is a process — by frequent usage of source language (SL) and target language (TL) terms these terms penetrate the subject field and language and become standardised (cf. Alberts 2017: 273-300).

The emphasis is placed within the subject-oriented framework on the standardisation of terms and concepts to facilitate effective communication and knowledge transfer within the professional community. In most cases, the subject specialists themselves are responsible for the documentation of specialised terminologies (cf. Alberts 2017: 177-179; Alberts 2019: 60).

This approach can also be regarded as an item-based approach since it limits the terminological research to the structure of languages for special purposes (LSP). The study of word-formation principles is key to this approach, especially the influence of Greek and Latin on terminology creation (cf. Cluver 1989: 150-158; Alberts 2019: 60).

Terminology development goes hand in hand with language development. The first bilingual English/Afrikaans Chemistry dictionary compiled by Dr D.F. Malherbe du Toit was published in 1918. Afrikaans became an official language in 1925. Several technical dictionaries and term lists were since compiled by subject specialists to enable subject-related communication in various subject fields. These subject specialists all worked according to the subject-oriented terminography approach. There were unfortunately several duplications of projects which led to a proliferation of terms for the same subject and/or language. This practice led to miscommunication and a lack of standardisation. Unfortunately, this process is currently repeated for the Sintu languages — various language units and tertiary institutions are creating terms in the official languages, but since these efforts are not properly coordinated, the terminology created by these institutions are also showing a proliferation of terms for the same concept in a given language and subject area (Goosen 2024: 56, 149).

3.2 Language planning-oriented terminography approach

With the language planning-oriented terminography approach the emphasis falls on official intervention in increasing the status of minority languages or languages which have for some or other reason been historically disadvantaged. The underlying belief of this approach is that the use of an unstable language can change with systematic strategic intervention carried out by official bodies. Legislation and other appropriate measures must be put in place to implement the change (cf. Alberts 2017: 181-182). Afrikaans was for instance regarded as a

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

mere kitchen language, but when it became an official language in 1925 its status changed and it became a language of learning and used in various professions.

The language planning-oriented terminography approach is to a certain extent a process-oriented approach and the relationship between language planning and the eventual standardisation of languages is studied (cf. Cluver 1989: 150, 180-193; Alberts 2019: 60).

The language planning-oriented terminography approach can also be regarded as a functional approach since it determines the unique communicative needs of language for special purposes communities. It considers the similarities between terminologies and general vocabulary and between terminography and lexicography (cf. Cluver 1989: 150, 155, 159-173; Alberts 2019: 60).

The language-planning terminography approach is an interactive approach since it places the emphasis on the dynamics of languages for special purposes in its entirety, rather than on its constituent parts. Languages for special purposes are developed through various processes and several factors interact during these activities. One such factor is language politics where it plays a significant role in the development of nationalism, while simultaneously it is the result of nationalism. As such, terminology is influenced by national or official languages, while in turn influencing the development of national or official languages (Cluver 1989: 150, 193-196, 238; Alberts 2019: 60).

Various South African government bodies and other organisations established language bureaux during the bilingual dispensation (i.e. the Department of Culture [later: the Department of National Education]; Transport; SABS; Department of Defence; Iscor; Suid-Afrikaanse Akademie vir Wetenskap en Kuns; Post Office; Wool Board, HSRC; CSIR; etc.). The bilingual policy of the country and language planning prescripts required Afrikaans to be developed as an official language to be on par with the global English language (Botha, Alberts and Kapp 2010)

Terminologists were appointed at the language bureaux to document the terminology of the relevant institution. These bodies all worked according to the language planning-oriented approach and on conceptual/subject-related principles.

To be able to attain the necessary change of status, the Afrikaans language had to have an up-to-date, coherent terminology to ensure professional communication in all subject fields. The objective was to replace terminology imported from languages spoken in technologically advanced and dominant countries with local or native equivalents. The native language is therefore fostered and developed.

As mentioned earlier Afrikaans was once regarded as a so-called "kitchen language" but through inter alia terminology development it became an academic and scientific language. The current multilingual policy of the country requires terminology to be developed in all official languages, especially to develop the official Sintu languages — previously regarded as marginalised languages — to become functional languages in all spheres of human activity. Since 2023 spe-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

cial attention is also given to the development of standardised signs for South African Sign language, captured as video clips.

3.3 Translation-oriented terminography approach

The translation-oriented terminography approach (TOT) is highly developed in institutionally bilingual or multilingual countries and communities. Terminology processing is seen as an activity which supports translation work (cf. ISO 2002; Alberts 2017: 179-180, 210-215; Alberts 2019: 20). This approach establishes terminological equivalents in the various languages which are used as points of reference by translators, and which contributes to the quality of the translated text. One of the major spin-offs of TOT is the creation of multilingual polythematic term banks. Translators consult a term bank for the same reason they consult a dictionary, e.g. to check the spelling of a term, its meaning, its grammatical category, the subject field in which it is used, its synonyms, equivalents in other languages, etc.

In the early 1950's translators at different language bureaux in South Africa started documenting source language (SL) terms and target language (TL) term equivalents to enable them to reuse the terminology in future translation work. They worked according to the translation-oriented terminography approach. This process assisted in the standardisation of SL and TL terms. In South Africa translators are still regarded a major group contributing to terminology development and the standardisation of terms.

3.4 Linguistic community-oriented terminography approach

The linguistic community-oriented terminography approach is a sociolinguistic approach to terminography where the relationship between language for special purposes and the community is determined. Terminology is therefore harvested from rural and urban speech communities and documented (Cluver 1989: 150, 197; Alberts 2017: 180-181).

As noted by Alberts (2010: 615), this process is costly and therefore seldom undertaken. The fact that terms existing in the communities are not documented contributes to the general and erroneous stereotype that the Sintu languages are incapable of naming abstract concepts (ibid.). This is a fallacy. Various terms in a variety of domains already exist in the Sintu languages. The problem is that these terms are not documented in a widely accessible format and therefore not standardised (Alberts 2017: 206). Various text books, study guides and other records containing terminology do exist in the Sintu languages, but are often not shared with the broader community or not available in a centralised term bank, with the result that terminology development happens in silos and therefore not standardised.

The various dialects also contain a wealth of terms which could be harvested and utilized in the standard languages, i.e. the terminology related to animal

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

names, bird names, customs and beliefs, traditional medicines, etc. These terms could be obtained from the older members of the linguistic community. Should these people die their knowledge of terms and related information (i.e. indigenous knowledge systems) dies with them (cf. Alberts 2017: 180-181).

The terminology available in the urban areas should also be documented. Terms are available at language units in translated documents, and at tertiary institutions in textbooks, curricula, master dissertations, and doctoral theses. These terms are new, and if they are not documented (and translated into the relevant official languages where needed), they could not be standardised and disseminated to target users. It is only by documenting terms and related information, standardising the terms and disseminating term lists in various domains and subject areas that the South African indigenous languages will become functional languages in all spheres of professional and academic life.

The current projects at the national terminology office, the Terminology Coordination Section, focus on community needs (i.e. weather terms, basic health, HIV/Aids, Covid-19, election terms) and therefore a linguistic community-oriented terminography approach is followed.

4. Coordination and cooperation

Although the correct terminographical approaches were used to develop terminology in South Africa, there was unfortunately little coordination of projects and every institution developed (and is still developing) the terminology for its own purposes. Several technical dictionaries (i.e. dictionaries for special purposes) were published over the years. There are a duplication of projects and subsequently also a proliferation of terms for the same concept in the same subject and/or language.

Over the years various coordinating bodies were established to coordinate the efforts of the different language bureaux to ensure cooperation and to prevent the duplication of terminological activities:

- In 1958 the *Vaktaalkommissie* (Terminology Commission) was established to assist with coordination efforts.
- It was replaced in 1966 by the *Vaktaalskakelkomitee* (Terminology Liaison Committee).
- In 1974 the *Koördinerende Vaktaalraad* (KOVAK) (Coordinating Terminology Board — COTERM) was established to coordinate the various terminology endeavours of the different language bureaux and to avoid duplication of projects and the proliferation of terms for the same concepts in the same subject and/or language. COTERM replaced all the previous coordinating bodies. The managers heading the big language bureaux at the time were members of COTERM, e.g. Language Bureau of the South African Railways and Harbours Administration, Terminology Division, Department of National Education, Terminology Bureau of the Suid-Afrikaanse Akade-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

mie vir Wetenskap en Kuns, Directorate Language: South African Defence Force, the Woordeboek van die Afrikaanse Taal (WAT), the Language Bureaux of Iscor, SABS, SABC, CSIR, as well as, representatives of the Government Printer, The Municipal Translators' Association, translators from the province and the Johannesburg Technical College. The Director: Arts and Culture of the Department of National Education chaired the meetings and the Terminology Division supplied the secretariat (Alberts 2003: 136; Kapp 2009: 129). Each participating institution had to identify core terms related to their core business and supply these to the WAT for inclusion in the comprehensive dictionary for Afrikaans. Copies of all technical dictionaries and term lists had to be supplied to the WAT and South African libraries.

COTERM developed norms for terminology creation, the standardisation of terminographical principles and practice (i.e. terminology work and the compilation of technical dictionaries) and started with the computerisation of the South African terminology practice (Alberts 1983; Alberts 2003: 137).

COTERM did not fulfil its coordinating role and the various language bureaux continued with their own activities.

— In 1984 a research report by the Human Sciences Research Council (HSRC) (cf. Cluver and Scheffer 1984) determined that the national terminology office, at the time the *National Terminology Service* (NTS) of the Department of Arts, Culture, Science and Technology (DACST), should become the official coordinating terminology office for South Africa. In 1995 the national terminology office was commissioned by the national government to compile multilingual technical dictionaries and terminology lists on national subject-related projects. The Terminology Coordination Section (TCS) commenced operations on 1 April 1998 as a sub-directorate of the National Language Service of the Department of Arts and Culture (DAC). It is currently a directorate of the National Language Service of the Department of Sport, Arts and Culture (DSAC). The TCS is the national terminology office and therefore responsible for the coordination of terminology projects in South Africa. TCS is the largest institution in South Africa that devotes its time entirely to terminology development work and is consequently the primary implementing agency for the national terminology policy.

The terminology practice in South Africa saw many changes over the years that had an influence on terminology development and consequently on the coordination of projects and cooperation with stakeholders:

— Initially bilingual English/Afrikaans translating technical dictionaries were compiled — currently translating and explanatory technical dictionaries are compiled in the 12 official languages of South Africa;
— The terminography process was previously done manually which involved data to be captured on index cards, thereafter the index cards had to be alphabetised in the two languages concerned, then manuscripts were typed, cor-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

rected, and then followed by galley proofs and page proofs — all which had to be proofread several times to ensure the correctness of data. Currently terminological data is entered into a database (i.e. a terminology management system). If data is entered correctly there is no need for the drudgery of the repeated proofreading thereof.

— There were during the bilingual dispensation at its peak 26 terminologists working at the national terminology office on English and Afrikaans terminology in a variety of subject areas — mainly in the higher echelons of science and technology. Currently there is only one terminologist per official language employed at the national terminology office (i.e. TCS) although the staff complement provides for two positions per official language with two terminologists employed for isiXhosa only. Since 2022 no terminologists were employed for English and Afrikaans first-language speakers. These posts, as well as that of a SASL terminologist, were advertised but TCS were informed not to proceed with the selection process until a later date — probably after the approval of the draft national terminology policy (Mnisi 2024b).
— Primary term creation (practiced by the terminography section) is done in the source language (which is currently English), and secondary term creation is done by the terminology section in the other official languages (cf. Figure 2).
— Previously terminologists were highly skilled workers. Novice terminologists received in-service training and worked under the mentorship of trained senior terminologists. Currently there are few trained terminologists in the principles and practice of terminology and terminography, and in-service training is limited to training in the terminology management system (TMS). The TCS currently uses the Autshumato TMS developed by CTexT. Unfortunately, terminology development initiatives in South Africa do not necessarily use TMSs. Sadly, they use flat files such as Excel sheets and MS Word documents which means that data exchange is extremely difficult.
— Prior to 1996 (i.e. before the multilingual dispensation) the national terminology office had a section called Systems Development and Research (cf. Figure 2; Alberts 2005; Alberts 2017: 185). This section was responsible for needs assessment studies, it had to determine terminology-related priorities, and it kept a national terminology register of all terminology projects to ensure the coordination of projects and the cooperation of various terminology creating entities. The national terminology register (NTR) was non-operational for many years, but was reconfigured during 2023 to meet the current needs concerning cooperation and the prevention of duplication of projects. The new NTR was launched at the National Language Forum in 2023 and is fully functional (Machaba 2024).
— At the time, i.e. prior to 1996, the section Data Management and Publication (cf. Figure 2; Alberts 2005; Alberts 2017: 185) dealt with the database, and with the publications of the technical dictionaries. It also published a

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

Directory of Language Resources, a *Catalogue of Technical Dictionaries*, and a monthly newsletter, called *Termbroker*. These resources kept other terminology entities informed of various aspects relating to terminology development, and assisted with coordination and cooperation endeavours.

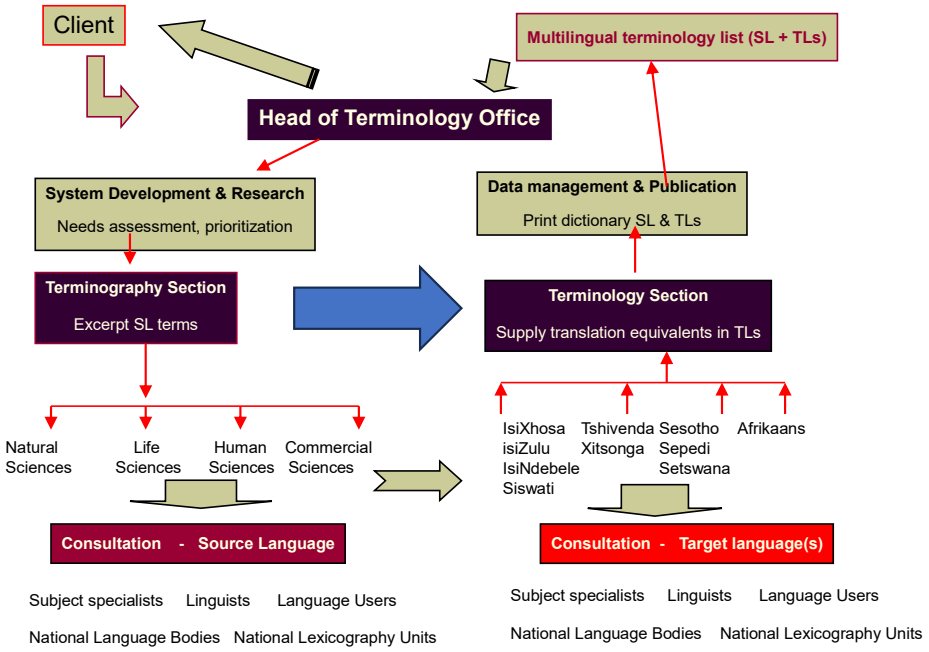


Figure 2: The workflow of the terminology management process in a multilingual national terminology office (cf. Alberts 2005; Alberts 2017: 185)

Despite all these coordination efforts by the national terminology office, i.e. the TCS and its predecessors, there is still a duplication of projects that prevents the standardisation of concepts and their designated terms in the various official languages. One of the reasons for this is the variety of stakeholders who are participating in the strategy towards achieving the goal of developing multilingual terminologies, e.g. the subject specialists, linguists, language units, National Lexicography Units, tertiary institutions, translators, journalists, terminology centrums, i.e. Centre for Political Terminology in Southern Africa (CEPTSA), Centre for Legal Terminology in African Languages (CLTAL), language users, and publishing houses, to name but a few (cf. Figure 3). Some of these bodies create terminology, others evaluate the term creations, e.g. the National Language Bodies of the Pan-South African Language Board (PanSALB). There are, however, no structure, nor norms or a policy in place to regulate these different efforts.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

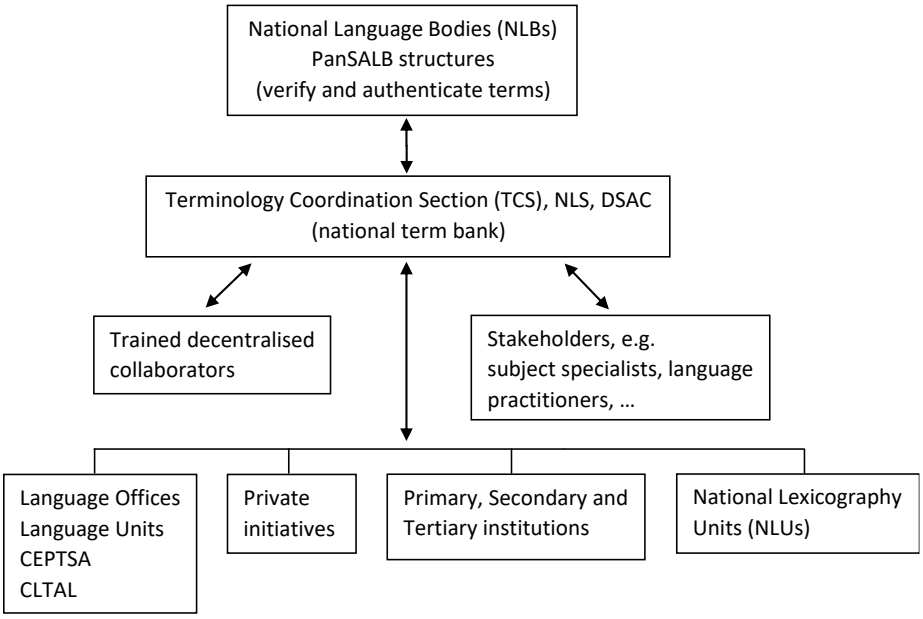


Figure 3: Collaboration between TCS and terminology stakeholders (Alberts 2006; Alberts 2017: 208)

Although there were/are efforts by TCS to ensure cooperation through the coordination of projects, these efforts usually fail because the stakeholders work in silos, and every institution carries on with its own projects which again leads to a proliferation of terms for the same concept in the same subject area and/or language. The reconstructed national terminology register (NTR) managed by TCS to register all terminology projects, enable cooperation, and prevent duplication is currently fully functional and should be able to coordinate terminology-related projects.

5. Terminology policy

5.1 Terminology policies in general

A terminology policy affects everybody who and every institution that needs to understand or acquire specialised knowledge of some kind or other. For this reason, those individuals and institutions that are most dedicated in their commitment to language planning policies are required to be involved. It is, however, also advisable to engage the whole language and subject community in the process of generating policy on terminology (UNESCO 2005: 15; Alberts 2017: 174).

Terminology policy can be defined as a strategy which aims at developing or regulating emerging and existing terminologies (cf. http://www.infoterm/).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

Terminology policy-making at a national level is a complex process, one constantly influenced by various factors, such as demographics, cultural, ethno- and geolinguistic factors as well as socio-psychological, economic, political, and legal factors. A terminology policy may therefore serve quite different purposes (cf. Alberts 2017: 174).

In the strategic management context, a terminology policy may be seen as outlining a vision and a set of instruments for:

— fostering the optimal and rational use of a language's referential resources to support specialised communication, and
— managing the intellectual and other property assets of an organisation/ community to advance given goals (cf. Antia 2008: 10).

This characterisation of a terminology policy implies that vision in terminology policy-making may be directed not only at efficiency and effectiveness of special-purpose communication (professional, academic, scientific, and technical communication), but also at a range of added value goals dealing with knowledge. A terminology policy in the strategic management sense also implies that terminology attempts to communicate an overarching vision which then motivates the activities of phases 1–5 of Figure 4.

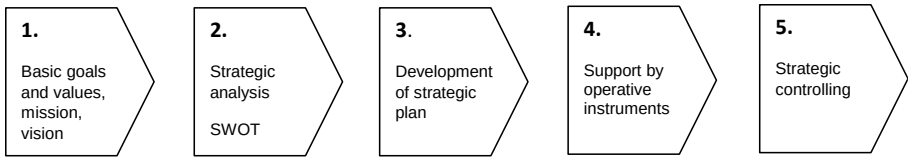


Figure 4: Strategic management process (cf. Antia 2008: 10)

Terminology policy-making seeks to develop goals, values, a mission, and a vision (phase 1) against the backdrop of a strategic environmental analysis of strengths, weaknesses, opportunities, and threats (SWOT analysis) (phase 2), then converts the vision into actions or programmes intended to produce measurable outcomes, i.e. the development of a strategic plan, operative instruments to support the strategic plan, and controlling mechanisms (phases 3–5). The process is cyclical and allows for formative evaluation. A terminology policy may mirror the initial phase or all the phases indicated in Figure 4 (cf. Antia 2008: 10; Alberts 2017: 174-177).

Terminology policy development is also the concern of international organisations. For instance, Subcommittee 1 (SC 1) of ISO/TC 37 drafted a standard: *Terminology Policies — Development and Implementation* (ISO/CD 29383-1). This draft standard is based in part on the *Guidelines for Terminology Policies: formulating and implementing terminology policy in language communities* published by the United Nations Educational, Scientific and Cultural Organization (UNESCO)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

in 2005, a project in which several members of the ISO/TC 37 SC1, including South Africa, collaborated (Alberts 2017: 175).

The ISO standard is designed for terminology policy-makers working in diverse environments, from language planning to for-profit companies. It provides guidance on general principles for the design of an individual policy to be tailored to a specific set of circumstances and it recommends a variety of actions that have proved to be helpful in different situations (e.g. national context (a legal framework)), and in non-governmental organisations (NGO's), private corporations, and small organisations (Alberts 2017: 175).

Language planning and terminology planning are interdependent and complementary activities. There are different phases in the formulating and implementing of a terminology policy, i.e.:

— *Step 1*: Preparation of terminology policy (survey, awareness-raising, consultation)
— *Step 2*: Formulation of terminology policy (draft proposal, coordinate with strategic plan)
— *Step 3*: Implementation of terminology policy (strategies, development plan, presentation, management, publicity, and promotion)
— *Step 4*: Sustaining the terminology infrastructure (operational plan, business model, frequent monitoring, research, network, flexible infrastructure) (cf. UNESCO 2005: 14-21; Alberts 2010).

Any formally conceived terminology policy is likely to address one or more of the following issues:

— establishment of support mechanisms and promotion activities,
— capacity building through the training of terminology experts,
— the development of terminological products, such as term lists, technical dictionaries, terminology databases and services, and
— standardisation activities (cf. http://www.infoterm; Infoterm 2005).

Terminology development clearly occupies a special position within the framework of language and communication, and terminology development is dependent on language planning and language policies (Layton, Alberts and Sanker 2024).

5.2 The need for a national terminology policy for South Africa

As indicated previously, a terminology policy is determined first and foremost by the language policy of a country. During South Africa's previous bilingual dispensation, terms were documented in English and Afrikaans. The new multilingual dispensation, however, demands that terminology be developed in all 11 official languages (Alberts 2017: 176), and, since 2023, also in South African Sign Language.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

As the result of the findings of the Language Plan Task Group (LANGTAG) Report (1996) the terminology section of the Science and Technology branch of the Department of Arts, Culture, Science and Technology (DACST) amalgamated with the translation section from the Arts and Culture branch to form a new National Language Service (NLS). The Terminology Coordination Section (TCS) of the National Language Service, Department of Sport, Arts and Culture (DSAC) is currently the national office commissioned by the national Government to compile multilingual technical dictionaries and terminology lists on national projects (cf. DACST 1996a; 1996b; 1996c; Alberts 2017: 176).

It is important to develop the official languages into functional languages in all spheres of professional life, because transfer, assimilation and retrieval of information should be through the first language (L1). Information is best acquired (the decoding process) and conveyed (the encoding process) through the L1. There is an urgent need for multilingual polythematic terms in all registers — from grassroots level to the higher echelons of science and technology. The national government therefore requires terminology to be developed in the various official languages of the country to communicate with the citizens in the languages they understand best (Alberts 2024: 13).

Terminology work, i.e. terminography, is concerned with the collection, systematisation, description, processing and presentation of concepts and their designations in various subject areas in different language(s). Terminology management is that part of terminology work concerned with the recording and presentation of terminological data in the form of term banks, glossaries, technical dictionaries, term lists, thesauri, or other publications, and online (cf. ISO 1087-1:2000).

Effective academic, economic, scientific, and technical communication skills are developed by using correct and standardised terminology. The availability of multilingual polythematic terminology can, in a sense, be an indicator of a nation's academic and professional achievements and its social, cultural, and educational development. The need for unambiguous communication in the theoretical and applied fields of human activity is constantly growing (Alberts 2017: 173). Terminology, therefore, is a strategic resource and has an important role to play in the functional development of languages and their users — especially in a multilingual country. This is largely because, effective economic, scientific, and technological transfer and assimilation of knowledge and skills among subject specialists and language practitioners together with the communication skills of the citizenry are developed through correct terminology usage (Alberts 2017: 173-174).

The provision of standardised terms and definitions in various target languages assists these languages to enhance their utilization in various subject fields, assisting the languages to become functional languages in all domains of human activity. Even a global language such as English is not functional in all domains and needs continued development of terminologies and terms for new inventions, discoveries, technologies, etc. Terminology work is therefore also an exercise in continual and sustained language development, addressing the availability of concepts in the languages of the users, rather than only in English. All official languages, including South African Sign Language (SASL), have the poten-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

tial to develop their respective terminologies, and should not be oppressed by the hegemony of English (Layton, Alberts and Sanker 2024).

Once the terminology of the previously marginalised languages is developed into functional terminologies, South Africans will be better equipped with effective communication tools. Languages can benefit immensely in their continual striving towards functionality in all aspects of human endeavours through terminology development projects that can be carried out by terminology offices, private initiatives, publishers, and developers of human language technologies (Layton, Alberts and Sanker 2024).

Unfortunately, there are very few trained terminologists in South Africa. In-service training is often limited to training in the use of terminology management systems rather than in terminology principles (theory), and in methodologies and practice (terminography). Terminology training at tertiary institutions is often limited to modules in lexicography, linguistics, and translation studies.

The TCS currently employs only one terminologist per official language, since 2024 a position for a SASL terminologist is approved. Posts are currently filled when a vacancy arise, but only for one terminologist per official language. No English first-language terminologist is employed although English is currently the documentation language and definitions are supplied in English only (Mnisi 2024a).

The national term bank is currently not operational and the national terminology register was until recently under construction which prevented cooperation with and among other bodies concerned. The various term-creating bodies work in silos. Since there is no coordination of projects every institution develops the terminology for its own purposes — this is reflected in the duplication of projects and subsequently also the proliferation of terms for the same concept in the same subject and/or language as experienced at tertiary institutions, government (local, provincial, national), and private initiatives.

These challenges have arisen in South African terminology development, as documented, and as discussed over several years, and it is these problems which have highlighted the need for a national terminology policy. The challenges therefore include the coordination of terminology activities in the country to ensure consistency and to avoid duplications of effort; systems to support the development and management of terminologies; and the provision of skills development for those working on terminology work at all levels. Also of concern is how terminologies will be made available and used by the subject specialists, and the linguistic and language for special purpose communities whose work can benefit from these terminologies (Layton, Alberts and Sanker 2024).

Another issue concerns the significant backlog in the development of the terminologies of the official languages, including in the digitisation of the published bilingual (English/Afrikaans) technical dictionaries. These dictionaries should be digitalised, revised, updated, and extended for the addition of Sintu language term equivalents, as well as SASL signs. These should be disseminated to target users, such as subject specialists, translators and interpreters, and the linguistic community (Layton, Alberts and Sanker 2024).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

The suggested solutions offered in the draft national terminology policy (cf. 5.3) are based on best practice concerning the current South African terminology scenario. The general underlying problems to be addressed by a national terminology policy can be summarised as follows:

- Lack of a national terminology policy
  - Lack of uniformity of terminology development methodologies
  - Lack of clearly defined roles of stakeholders
  - Stakeholders work in silos — no liaison with a lack of guidelines, cooperation
- Lack of a national terminology coordination strategy
  - Duplication of projects
  - Proliferation of terms for the same concept in the same subject and/or language
  - Problems regarding the standardisation of terms — leads to miscommunication
- Management issues
  - Inadequate national term bank
  - Inoperable national terminology register
  - Personnel structure
  - Training issues — in-service as well as tertiary training
  - Incompatible terminology management systems, and data therefore not interchangeable
  - Developers of software/service providers do not always supply technical assistance
  - Previously compiled/out of print terminology-related publications should be digitised, revised, updated and other official language terms and SASL signs should be added
  - Terminological data is not open and not readily available to all users
  - Copyright should not apply to terminology — multilingual polythematic terminology should be open to all users
  - Insufficient funding for terminology-related work
- Dissemination
  - Various platforms are available to disseminate multilingual polythematic terminological data and they should be utilized, i.e. national term bank, SADiLaR, VivA
  - Dissemination should be online and publishing houses, i.e. the Government Printer could be requested to publish hard copies when needed
- Training
  - Tertiary institutions should be requested to present terminology training modules. Currently there is only a proper terminology module at the University of Pretoria and at other institutions it forms part of modules of lexicography, linguistics, translation, and interpreting courses.
  - Practicing terminologists should be encouraged to enroll for post-graduate terminology training (Layton, Alberts and Sanker 2024).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

These underlying problems are the starting point for the process which has led to the drafting of a national terminology policy.

5.3 A national terminology policy for South Africa

There is no existing national terminology policy currently operational in South Africa, even though the UNESCO Guidelines on terminology policies are readily available, and the international ISO TC/37 standard on national terminology policy was adopted as a national SABS TC/37 standard. These guidelines and standards are available and can be applied to South African institutions engaged in terminology-related work (cf. PanSALB 2001; ISO 2005, 2020, 2022; Layton, Alberts and Sanker 2024).

Language management in South Africa is coordinated by the National Language Service (NLS) of the Department of Sport, Arts and Culture. Terminology development is coordinated through the Terminology Coordination Section (TCS), a directorate within the NLS, which has the responsibility to develop and publish terminologies. However, terminology development is not the exclusive domain or responsibility of the TCS and is conducted by several organisations, both within and outside of the three spheres of government (i.e. local, provincial, and national government), tertiary institutions, publishers, and private individuals. For the purposes of the national terminology policy, the TCS, as the national coordination body for terminology in South Africa, is the implementing and coordinating agency (cf. Layton, Alberts and Sanker 2024; Mnisi 2024a).

The TCS therefore needs to establish relationships with all the other bodies concerned. To be functional, the TCS must have fully-trained terminologists in all twelve official languages, receive sufficient funding, and must be supported by a sufficient infrastructure to fulfil its task (cf. Layton, Alberts and Sanker 2024).

The national terminology policy (NTP) for South Africa (cf. Layton, Alberts and Sanker 2024) was drafted in compliance with the *Constitution of the Republic of South Africa* Act 108 of 1996 (now *Constitution*, 1996) relevant pieces of legislation, known structures, and prescripts of UNESCO and standardising bodies such as ISO TC/37 and SABS TC/37. The process of drafting a national terminology policy started in August 2023 and a final draft was submitted to the Department of Sport, Arts and Culture in June 2024.

The main objective of the national terminology policy for South Africa is to provide policy makers in government and other institutions with guidelines to support the development of terminology and the implementation of practices and methods in terminology management and terminology work (i.e. terminography) by the national terminology office, i.e. the Terminology Coordination Section (TCS) (cf. Layton, Alberts and Sanker 2024).

This policy is regarded as a strategy in coordination with, and supporting, the general development of such policy for South Africa, including all stake-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

holders from government (national, provincial, regional, local), organisational, educational (primary, secondary, and tertiary levels), publishing houses or private terminological enterprises. The NTP is evidence based and has a visionary approach. Various steps were recommended to realize the objectives by the implementation thereof (cf. Layton, Alberts and Sanker 2024).

The NTP was drafted after an extensive literature study, consultation with stakeholders (also with the SASL community), and the gathering of information from various sources. Its aim is to provide government, administration, tertiary institutions, publishers, non-profit and profit organisations, and private enterprises with guidelines for the development and implementation of a comprehensive approach to the planning and management of terminology. It further aims to stress the importance of a functional national terminology service that comply with the prescripts of the multilingual policy as prescribed by the Constitution (1996) (cf. Layton, Alberts and Sanker 2024). The aim of the NTP is furthermore to provide the TCS with guidelines and implementation plans to improve its services, its coordination role, and its support to language units, various stakeholders, education (primary, secondary, tertiary), subject specialists, language practitioners, publishers, etc. (Alberts 2024).

The NTP addresses the specific role of terminology, as a concept-based structure. It addresses the coordination and management of terminologies for special contexts to improve the encapsulation of knowledge, communication in subject-related or professional environments, and to provide for concepts and terms to be available in the diverse languages of South Africa. This policy therefore addresses languages for special purposes (LSP), which are located within a specific subject area, domain, discipline, or profession. Arising from this range of issues, this policy has a universal reach throughout society, and the activities which this policy demands are an essential and deep-rooted component of an effective society (cf. Layton, Alberts and Sanker 2024).

Issues raised in terminology work in the past and highlighted during various engagements with the stakeholder community during the national terminology policy development project, included the following gaps which required attention (cf. Layton, Alberts and Sanker 2024):

- Training in the basic principles and practices of terminology and terminography
- Needs assessment studies
- Planning and prioritisation of terminology projects
- Terminography process: excerption/harvesting, documentation, systematisation, standardisation
- Personnel structure
  - all official languages should be represented
  - If SL is English there should be an English first-language speaker employed to compile English definitions
  - vacancies should be filled immediately

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

- Job performance
- Terminology training
  - Trained terminologists should mentor newly appointed terminologists.
  - Terminologists should receive training in terminology (theory and principles); terminography (methodology and practices); terminology management system (TMS) — the training should include a range of different types of software for terminology management should the type of software change in future.
  - Post-graduate training at a tertiary institution.
- Copyright issues — terminology should not be copyrighted, rather open and available to everyone always.
- Language planning, harmonisation, standardisation.
- Terminology management through a terminology management system (TMS).
- TMS should be compatible with other TMS' to allow data exchange. The format in which the terminological data can be downloaded, exchanged and uploaded onto another terminology management system should also be compatible.
- National term bank (NTB) — currently not operational, but should be a national repository for multilingual polythematic terminology.
- Feedback — a possibility to comment on available terminological data to validate the data — this would help to identify the needs of target users.
- National terminology register (NTR) — coordination of projects to ensure collaboration and avoid duplication.
- Digitisation of published/out of date technical dictionaries/language for special purpose (LSP) dictionaries.
- Revision of published/out of date LSP dictionaries and the adding of term equivalents in the other official languages.
- Dissemination of terms and related information in printed and online format.
- Communication (i.e. printed/online newsletter, terminology forum) with target users such as language units, subject specialists, translators, interpreters, academics, National Language Boards (NLB)s, National Lexicography Units (NLU)s, SADiLaR, etc.
- Collaboration with national target users, role-players, stakeholders, SABS TC/37, tertiary institutions, terminology centrums, private enterprises, publishing houses.
- International liaison (UNESCO, ISO TC/37, tertiary institutions).

The project team addressed these issues in detail and implementation plans were put forward in the national terminology policy as presented to the NLS, DSAC.

The TCS is the largest institution in South Africa that devotes its time entirely to terminology work and is consequently the primary implementing agency for this policy. The policy therefore emphasises the role and responsibilities of the

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

TCS in terms of its coordination role, and that no other institution should overlap with these responsibilities, which may dilute the effectiveness of the TCS, and which may increase the challenges arising from working in silos on problems in the national interest for which integration and coordination are essential characteristics. Although the role and responsibilities of the TCS received attention during the drafting of the NTP, it was noted that the TCS has the infrastructure to execute this task but lacks sufficient capacity and funding to meet its mandate. These were key issues which the NTP sought to mitigate (cf. Layton, Alberts and Sanker 2024).

6. Process for acceptance of the National Terminology Policy

The project team responsible for the drafting of a National Terminology Policy for South Africa, i.e. Dr Roger Layton, Dr Mariëtta Alberts, and Ms. Sholeen Sanker, presented various drafts of the NTP to the NTP steering committee, stakeholders, and collaborators. Comments received from collaborators were included in the final version of the NTP that was presented to DSAC in June 2024.

The inclusion of South African Sign Language at a late stage of the process was a challenge but a very successful workshop was held with the SASL community and they also commented on the draft NLP. The specifications for systems that were incorporated in the draft NTP include the need for terms (written format) and signs (video clips).

The contents of the national terminology policy could only be introduced officially to the South African terminology community once the following process for accepting the NTP is concluded (cf. DSAC 2024):

- The NTP steering committee appointed by the Director General of DSAC had to endorse the NTP
- The NTP steering committee meeting for endorsement of the NTP was scheduled for the middle of July 2024
- After endorsement the NTP would be tabled at the meeting of the Executive Management Team (EMT) of DSAC scheduled for a later date in July 2024
- Thereafter, around September 2024, DSAC would present the NTP to the DG's Forum (Technical Working Group) Social Protection, Community and Human Development (SPCHD)
- By the end of December 2024 DSAC should present the NTP to the Forum of South African Directors-General (FOSAD)/DG SPCHD cluster Technical Committee
- By the end of February 2025 DSAC should present the NTP to the SPCHD Cabinet Committee
- The NTP would then officially be gazetted for public comments — probably by the end of March 2025

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

— Public comments received would by the end of June 2025 be incorporated into the NTP by DSAC
— An updated version of the NTP would be submitted by DSAC to FOSAD SPCHD by the end of September 2025
— DSAC should be able to submit an updated version of the NTP to the SPCHD Cabinet Committee for its approval by the end of Dec 2025 (Alberts 2024).

Although certain dates for the endorsement process were indicated in the above-mentioned roadmap, the process involves high-ranking officers and is subject to their availability. At the time the calendar of the meetings of all Clusters was according to the DG's office not available, which meant a delay in the progress. However, the TCS team indicated their commitment and dedication to ensure the endorsement of the draft NTP (Mnisi 2024b).

Although the formal roadmap for the approval of the NTP seems to be taking a long time, this process needs to be finalised by the relevant authorities to enable the terminology practice of South Africa to finally work according to a national terminology policy, specifically designed to deal with all the aspects mentioned in the policy dedicated to the entire South African terminology practice.

7. Conclusion

The availability of an operational national terminology policy (NTP), aligned to the Constitution of the Republic of South Africa (1996), all relevant legislation, frameworks and policies predicts a prosperous future for terminology development in South Africa — especially if the recommendations and implementation plans described in the national terminology policy is followed (cf. Layton, Alberts and Sanker 2024). This could lead to effective coordination of terminology development activities nationally. Challenges regarding coordination issues could be resolved by a functional national terminology register (NTR), preventing the duplication of projects, and assuring collaboration among stakeholders. The availability of a national term bank (NTB) as national repository for terminological information and the printing of hard copy dictionaries and/or the online dissemination of terminological data on open platforms could allow access to all citizens as end-users thereof. The various TMS' used by different stakeholders should be interoperable and compatible to allow for terminological data exchange. The NTP should also be able to create synergy among terminology developers by providing a comparative benchmark regarding the involvement of stakeholders and collaborators. It should also through a terminology forum, and regular newsletters ensure proper liaison with terminologists, language practitioners, subject specialists, collaborators, National Language Bodies and National Lexicography Units.

The NTP emphasises the importance of needs assessment studies to deter-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

mine the terminology needs of the country, to prioritise projects, to coordinate projects, to collaborate, and to establish working relationships with stakeholders.

The NTP stresses the importance of ongoing terminology training (in-service and post-graduate training) in terminology principles, terminography practices and methods, and terminology management systems.

The NTP gives guidelines on the development of the terminology of a variety of subject fields, domains, and disciplines in all registers — from grass-roots level to the higher echelons of science and technology to accommodate the needs of target users, i.e. laypeople, semi-specialists, and experts. The equal development of all official languages into functional languages in all spheres of human activity would adhere to the multilingual dispensation of the country in accordance with Section 6 of the Constitution of the Republic of South Africa (1996) — no language should be left behind.

8. References

Advisory Panel, Minister of Arts, Culture, Science and Technology. 2000. *Language Policy and Plan for South Africa.* Final draft. 29 February 2000. Unpublished DACST document.

Alberts, M. 1983. *Aspekte van die rekenarisering van vakwoordeboeke.* M.A. Thesis. Pretoria: University of Pretoria.

Alberts, M. 2003. Lexicography and Terminology Training at University Level. *PanSALB News*, April–June 2003: 3-4.

Alberts, M. 2005**.** National Language Bodies (NLBs) and Technical Committees (TCs). Paper read at the PanSALB Colloquium on National Language Bodies, 11 November 2005, PanSALB, Pretoria.

Alberts, M. 2008. National Language and Terminology Policies — A South African Perspective. *eDITion Terminologiemagazin* 1(2008) (Deutscher Terminologie-Tag e.V. (DTT), Deutsches Institute für Terminologie Ev (DIT)): 18-21.

Alberts, M. 2010. National Language and Terminology Policies — A South African Perspective. *Lexikos* 20: 599-620.

Alberts, M. 2017. *Terminology and Terminography Principles and Practice. A South African Perspective.* Milnerton: McGillivray and Linnegar.

Alberts, M. 2019. *Terminologie en terminografie: 'n Handleiding.* Pretoria: SAAWK.

Alberts, M. 2024. Current State and Practices of Terminology Development in South Africa. Unpublished preconference workshop presentation delivered at the African Association for Lexicography 28th International Conference, Pretoria, 1 July 2024. Pretoria: University of Pretoria.

Antia, B.E. 2008. Vision and Terminology Policy. *eDITion Terminologiemagazin* 1(2008). (Deutscher Terminologie-Tag e.V. (DTT), Deutsches Institute für Terminologie Ev (DIT)): 10-13.

Botha, W., M. Alberts and P.H. Kapp. 2010. Historical Experiences in Developing Afrikaans as a Language. Unpublished paper read at the 'Steering Committee: African Languages Roundtable', Department of Higher Education and Training, 22 October 2010, Unisa, Pretoria.

Cluver, A.D. de V. 1989. *Dictionary of Terminography/ Terminografiewoordeboek*. Pretoria: Human Sciences Research Council.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

Cluver, A.D. de V. and C.J. Scheffer (Eds.). 1984. *RGN-ondersoek na vaktaalaangeleenthede in Suidelike Afrika: Verslag van die ondersoek na organisatoriese aangeleenthede op vaktaalgebied.* (Term-1). Pretoria: HSRC.

DACST. 1996a. *Proceedings of a LANGTAG Workshop on Issues in Language Development. Focusing on The Future of the Apartheid Language Boards and The Harmonisation and Development of African Languages.* 28 March 1996. Pretoria: State Language Services, DACST.

DACST. 1996b. *LANGTAG Report.* Department of Arts, Culture, Science and Technology, Pretoria.

DACST. 1996c. *Information on the Language Plan Task Group* (*LANGTAG*). Pretoria: State Language Services, Department of Arts, Culture, Science and Technology.

Department of Sport, Arts and Culture. 2024. *National Terminology Policy Roadmap — 13 June 2024.* Unpublished document. Pretoria: Department of Sport, Arts and Culture.

Goosen, M. 2024. *'n Kritiese analise van terminologiebestuur binne die Suid-Afrikaanse konteks*. Unpublished PhD dissertation. Pretoria: University of Pretoria.

Government Gazette. 2012. Use of Official Languages Act 12 of 2012. *Government Gazette*, Vol. 568, No. 35742, 2 October 2012. Cape Town: Government Printer.

Government Gazette. 2013. Use of Official Languages Act — Proposed Regulations. *Government Gazette* No. 36392, Government Notice 411 of 2013, 26 April 2013: 56-67. Cape Town: Government Printer.

Infoterm. 2005. Available at: http://www.infoterm/

International Standards Organisation (ISO). 2000. ISO 1087-1. *Terminology Work — Vocabulary — Part 1: Theory and Application.* International Organisation for Standardisation.

International Standards Organisation (ISO). 2002. ISO/TC37/SC2. *Translation-oriented Terminography*. International Organisation for Standardisation.

International Standards Organisation (ISO). 2005. ISO/CD 29383-1. *Guidelines for Terminology Policies. Formulating and Implementing Terminology Policy in Language Communities*.

International Standards Organisation (ISO). 2020. ISO/CD 29383-1. *Terminology Policies — Development and Implementation*.

International Standards Organisation (ISO). 2022. ISO 704:2022 (en). *Terminology Work — Principles and Methods.*

Kapp, P.H. 2009. *Draer van 'n droom. Die geskiedenis van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns, 1909–2009.* Hermanus: Hemel-en-See Boeke.

Layton, R., M. Alberts and S. Sanker. 2024. *National Terminology Policy*. Pretoria: Terminology Coordination Section, National Language Service, Department of Sport, Arts and Culture.

Machaba, H. 2024. SADiLaR Workshop — Preconference Workshop at Afrilex Conference. Private email 2024-06-21.

Mnisi, S. 2024a. Official Name of the Terminology Office, Please. Private emails 2024-08-28, 2024-08-29.

Mnisi, S. 2024b. NTP — Update. Private email 2024-11-14.

Pan South African Language Board (PanSALB). 2001. *Guidelines for Language Planning and Policy Development*. Pretoria: PanSALB.

Random House. 1987. *Random House Dictionary of the English Language.* Second Edition, Unabridged. New York: Random House.

Republic of South Africa. 1961. *Constitution of the Republic of South Africa.* Act 32 of 1961.

Republic of South Africa. 1983. *Constitution of the Republic of South Africa*. Act 110 of 1983.

Republic of South Africa. 1993. *Constitution of the Republic of South Africa*. Act 200 of 1993 (interim Constitution).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1976 (Article)

Republic of South Africa. 1996. *Constitution of the Republic of South Africa*. Act 108 of 1996.

Statistics South Africa. 2022. *South African National Census of 2022*. Available at: http://www.statssa.gov.za

Union of South Africa. 1925. *Official Languages of the Union*. Act 8 of 1925.

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2005. *Guidelines for Terminology Policies*. Vienna: Infoterm.

Wright, L. 2007. Language Policy and Planning: General Constraints and Pressure. English National Language Body, PanSALB. 2007. *English and Multilingualism in South African Society: Papers from the English National Language Body:* 6-11. Unpublished document. Pretoria: PanSALB.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

An Analysis of Word-Formation Processes in Neologisms: A Case Study of the *Oxford Advanced Learner's Dictionary*

Aling Shi, *School of Foreign Studies, Tongji University, Shanghai, China (shialing2011@163.com) (https://orcid.org/0009-0001-6053-0930)*

Abstract: The swift progress of technology and globalization has given rise to an abundance of fresh ideas, sparking the frequent generation of neologisms. This study examines 932 newly added entries in the *Oxford Advanced Learner's Dictionary* from March 2020 to March 2024, investigating their word formation and features. Besides common processes like compounding, derivation, abbreviation, conversion, and blending, the OALD has included new words formed through less conventional methods like eponymy, onomatopoeia, and variation. Compound nouns mainly consist of two nouns in open compound words, while derivatives are often created through suffixation, and abbreviations are primarily uppercase initialisms. Most converted words are verbs, shifted from nouns. Among blends, the noun usually dominates, with the first word often clipped. And the study reveals diverse origins for new words, with loanwords playing a significant role, and a tendency towards simplicity emerges over time. Researching the word-formation of English neologisms is vital for lexicography because it helps ensure dictionaries accurately reflect language evolution. By understanding how neologisms are formed, lexicographers can effectively document and categorize emerging words, while also aiding vocabulary acquisition by highlighting patterns that enhance learning and usage.

Keywords: NEOLOGISM, *OXFORD ADVANCED LEARNER'S DICTIONARY*, QUANTITATIVE METHOD, QUALITATIVE METHOD, WORD-FORMATION, FEATURE, INCLUSION

Opsomming: 'n Analise van woordvormingsprosesse in neologismes: 'n Gevallestudie van die *Oxford Advanced Learner's Dictionary*. Die snelle ontwikkeling van tegnologie en globalisering het gelei tot baie nuwe idees wat die gereelde vorming van neologismes tot gevolg gehad het. In hierdie studie word 932 inskrywings ondersoek wat vanaf Maart 2020 tot Maart 2024 nuut in die *Oxford Advanced Learner's Dictionary* ingevoeg is deur 'n ontleding van die woordvorming en eienskappe daarvan te doen. Buiten algemene prosesse soos samestelling, afleiding, afkorting, funksiewisseling, en reduksiesamestelling, sluit die OALD ook nuwe woorde in wat deur minder konvensionele metodes soos eponimie, klanknabootsing, en variasie gevorm is. Engelse naamwoordsamestellings bestaan hoofsaaklik uit twee selfstandige naamwoorde wat los van mekaar geskryf word, terwyl afleidings dikwels deur agtervoeging geskep word, en afkortings is hoofsaaklik letterklankwoorde wat uit hoofletters bestaan. Die meeste woorde wat van funksie ver-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

wissel, is werkwoorde, wat afgelei is van selfstandige naamwoorde. By reduksiesamestellings domineer die selfstandige naamwoord dikwels, met die eerste woord wat meermaal verkort is. Die studie toon ook diverse herkomste vir nuwe woorde, met leenwoorde wat 'n beduidende rol speel, en 'n geneigdheid tot eenvoud kom mettertyd na vore. Die ondersoek na die woordvorming van Engelse neologismes is noodsaaklik vir die leksikografie aangesien dit help verseker dat woordeboeke taalontwikkeling akkuraat weerspieël. Wanneer leksikograwe verstaan hoe neologismes gevorm word, kan hulle woorde wat toenemend gebruik word, effektief dokumenteer en kategoriseer, terwyl hulle terselfdertyd woordeskatverwerwing help deur patrone uit te lig wat leer en gebruik verbeter.

Sleutelwoorde: NEOLOGISME, *OXFORD ADVANCED LEARNER'S DICTIONARY*, KWANTITATIEWE METODE, KWALITATIEWE METODE, WOORDVORMING, EIENSKAP, INSLUITING

1. Introduction

Language is inherently dynamic, undergoing continuous transformation across time and space. New words are either invented to meet evolving linguistic needs or borrowed from other languages, and the meanings of old words shift with each passing generation. This perpetual evolution underscores the necessity for lexicography to continually adapt and update dictionaries in order to accurately reflect the changing linguistic needs and preferences of language users. In its latest update, more than 500 new words, phrases, and senses have been added to the Oxford English Dictionary[1]. All these new words originate from various spheres of life and serve to describe and capture social changes and advancements. T.S. Eliot (1943: 35), a renowned modernist poet and literary critic, once reflected, "For last year's words belong to last year's language, and next year's words await another voice". The rate of linguistic change is intertwined with the pace of social progress. Hence, the purpose of this paper is to analyze neologisms recorded in the OALD in recent years and identify the prevailing types of word formation and their features. In this paper, "new words" and "neologisms" are used interchangeably to denote identical concepts.

Researchers from diverse academic backgrounds present different definitions of neologism. According to Greenough and Kittredge (1901: 78) it is "a new word or phrase that has not yet been fully recognized in the standard variant but has gained undisputed status". Algeo (1991: 2) argues that "a new word is a language form or the use of a language form not included in general dictionaries, this language form can be a single word, a compound or even phrase". Crystal (1995: 455) believes "the term neologism refers to the creation of a new word out of existing elements (post-person, linguist field); also called a coinage". Gao (1998: 11) suggests, "[f]rom a reference standpoint, new words are those vocabularies that have not been documented in any dictionary, some dictionaries, or all existing dictionaries". Hartmann and James (2000: 99) define it as "a word or phrase has entered the language (as a borrowing or a coinage or through semantic change) relatively recently, often commented on and collected in spe-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

cialized dictionaries". Newmark (2001: 140) puts forward that "neologisms can be defined as newly coined lexical units or existing lexical units that acquire a new sense". A neologism can be initially defined as 'an item newly introduced into the lexicon of a language' (Malmkjær 2006: 601). According to Boulanger (2010: 63), "neologisms are nothing but the initial stage of a lexical unit, just a useful label to track words throughout the chronological axis of their existence".

For the purposes of this study, drawing on the foundation laid by predecessors, new words are defined as entries in a dictionary either as standalone lexical items or as additional meanings under existing entries. The inclusion of new words in dictionaries underscores the critical role lexicography plays in documenting and standardizing language. Dictionaries not only reflect current usage but also serve as authoritative references for understanding evolving meanings and word formations. The implication for lexicography is that it must remain adaptive, capturing the dynamic nature of language while maintaining clear and coherent entries. Word formation processes — such as compounding, derivation, and blending — are highly relevant to lexicography, as they represent the mechanisms by which new lexical items emerge. Lexicographers must assess the frequency, significance, and stability of these forms before incorporating them, ensuring that dictionaries remain comprehensive and reliable records of language development.

2. Methodological framework

In this study, the methodology involved constructing a detailed corpus of 932 newly introduced lexical items from the "Recent Additions to OALD" section on the official website of the *Oxford Advanced Learner's Dictionary* (OALD)[2]. These entries, ranging from 2020 to 2024, were carefully selected as the primary dataset, providing a diverse range of neologisms for scholarly investigation.

The research followed a two-pronged approach, combining quantitative and qualitative methods to ensure a comprehensive analysis of the data. First, in the quantitative phase, each neologism was systematically classified according to its word formation mechanism. This involved identifying whether the word was a result of processes such as compounding, affixation, blending, clipping, conversion, or borrowing. The frequency of each formation type was then calculated to reveal dominant patterns and trends in contemporary English word creation. In the qualitative phase, an in-depth examination of each neologism's syntactic and semantic properties was conducted. This included analyzing how the words function within sentences, their grammatical roles, and any shifts in meaning or nuance compared to similar existing terms. Special attention was paid to how the meanings of these new words reflect current social, technological, and cultural changes.

By integrating both quantitative data — highlighting the structural trends in neologism formation — and qualitative insights into the specific uses and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

meanings of these words, this methodological approach allowed for a nuanced understanding of the linguistic and cultural dynamics influencing modern English. This framework provided a thorough exploration of the emerging lexical items, revealing key characteristics that contribute to the development of English vocabulary.

3. Studies on word-formation processes of English neologisms

New words in English are not only increasing in number, but also coming from a variety of sources. O'Dell (2016: 95) remarks, "interestingly, it is actually very rare for an English word that is completely new to be formed. Often, repurposing takes place; in other words, a new sense is added to an already existing word".

Bradley (1904) discusses three types of word-formations in English namely composition, derivation and root-creation. Cannon (1978) identifies 21 word-formation categories such as shifts, borrowings, shortenings, initialisms, additions, and affixations in his 13,683-word corpus drawn from the three dictionaries: the *1973 Barnhart Dictionary of New English Since 1963*, the *1980 Second Barnhart Dictionary of New English*, and the *Merriam's 1981 Addenda Section to Webster's Third New International Dictionary of the English Language*. Algeo (1991) provides six categories of English word-formation based on neologisms from *Among the New Words* and they are creating, borrowing, combining, shortening, blending, and shifting. Bauer (1994) categorizes word-formation into abbreviations, condensing, compounding, and affixation in his book *Watching English Change*. In China, a multitude of scholars equally dedicate their attention to the formation of neologisms (Yang 1997; Lu 1999; Wang 2000; Li 2001; Gao 2001; Yuan 2019), exploring the processes by which new words are coined. Lu (1999) presents a list of word-formation processes that includes blending, compounding, eponym, back-formation and borrowing. Zhou (2000) analyzes neologisms within a 10-year span and reaches a conclusion that new words have three main characteristics which are large quantity, extensive usage, and flexible word-formation method. Wang (2000) takes all neologisms from *Oxford Neologisms Dictionary* and his findings indicate that the majority of these new words are formed through conventional word-formation methods such as compounding, abbreviation, conversion, back-formation, and onomatopoeia, with only two words being truly newly coined. Chen (2000) works on the word-formation trend of new words in the 20th century. Wang (2005) classifies English neologisms into 12 types which refer to coinage, derivatives, blending, shortening, initialism, functional shift, back-formation, eponym, borrowing, analogy, compounding and secretion. Yang (2008) analyses new words from a cognitive perspective and points out compounding, blending, deriving, conversion and abbreviation as the main word-formation methods of neologisms. Yu et al. (2011) divides the word-formation of neologisms into five categories. The first is neologisms created by putting

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

existing words to new uses, which refers to lexical changes in meanings and word class. The second is neologisms formed by combining existing words or word parts, which refers to compounding, parallel construction, affixation, blending, and words combined as a result of causal pronunciation, and reduplication. The third is a neologism formed by shortening existing words or expressions, which refers to clippings, backformation, abbreviation, shortened compounds, noun phrases, and verb phrases. The fourth is neologisms borrowed from other languages. And the last is a neologism created out of nothing. Gao (2012) suggests six important word-formation methods of new words and they are compounding, derivation, shortening, blending, borrowing and conversion. Liu (2015) discusses the word-formation of cyber neologisms from three angles: flexible word-formation like coinage and borrowing, euphemistic word-formation which refers to the traditional way of word-formation like compounding, and abbreviation like blending.

The discussions above demonstrate that most research on word formation encompasses processes such as coinage, borrowing, compounding, blending, shortening, and back-formation, among others. While scholars generally agree on broad classifications of word formation, there are differences in opinion regarding specific aspects. In the context of lexicography, this paper draws on the OALD collection and the established discourse on word formation to examine processes relevant to dictionary compilation. The focus includes common mechanisms like compounding, derivation, abbreviation, conversion, and blending, as well as less frequent but lexicographically significant processes such as eponymy, onomatopoeia, and variation. These processes highlight the complexity of documenting new lexical items and reflect the need for a comprehensive approach in lexicography to accurately represent language evolution.

4. Word-formation of English Neologisms in the OALD

This section examines the various word formation processes that contribute to the creation of neologisms, including compounding, derivation, abbreviation, conversion, blending, eponymy, onomatopoeia, and variation, all of which are critical from a lexicographical perspective. Compounding and derivation are identified as the primary methods, frequently shaping new entries in dictionaries. Abbreviation and conversion, though less common, are equally significant in expanding the lexicon and necessitate careful consideration by lexicographers. These processes, despite their lower occurrence, play a vital role in enriching the vocabulary and must be adequately documented to reflect language development accurately in lexicographical work.

4.1 Compounding

Jackson (1988: 30) states that "one of the most productive ways in which new

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

words have been coined, especially in modern times, is by the process called compounding". Words that result from compounding are commonly referred to as compounds or compound words. Among recent additions to the OALD, a significant portion, totaling 465 entries, constitutes compound words, representing an impressive 49.89% of the total. From the perspective of orthography, compound words manifest in three primary forms: open, closed, and hyphenated. Within this corpus, open compounds dominate with 347 instances, followed by 61 hyphenated compounds and 57 closed compounds. It is noteworthy that, while the formation of English neologisms generally favors open compounds over closed compounds, the current dataset reveals a significant departure from this expected pattern. From a lexicographic standpoint, this divergence underscores the importance of accurately categorizing and treating different types of compounds in dictionary entries. Lexicographers must carefully document the structure and usage of both open and closed compounds, ensuring that these distinctions are clearly represented to reflect contemporary language trends. This approach aids in providing users with precise definitions and contextual examples, enhancing the overall utility of the dictionary. There are four patterns among open compound words. The first pattern, constituting 94.3% of the total at 327 occurrences, involves open compounds formed by separating ordinary words, such as "capstone project". The second pattern comprises open compound words with hyphens. Typically, in these instances, the left part of the compound employs hyphenation to form adjectival modifiers for the subsequent ordinary words. There are a total of 6 occurrences, exemplified by "ultra-processed food". The third pattern includes three compound words composed of common words and combining forms, such as "back four", and the fourth pattern includes 10 compound words, such as "R number", consisting of ordinary words and letters. In closed compound words and hyphenated compound words, the overwhelming majority of terms are comprised of two common words combined, exemplified by "wingman" and "deaf-blind".

In the OALD, the predominant compound words include compound nouns, compound adjectives, and compound verbs, delineated by their respective quantities and proportions as depicted in Table 1 below. Additionally, "young adult" stands as a dual-functional term, serving both as a noun and an adjective.

Table 1: Part of speech of compounds

Part of speech	Total	Percentage	Example
Compound nouns	409	87.95%	artistic gymnastics
Compound adjectives	53	11.39%	deaf-blind
Compound verbs	3	0.64%	live-tweet

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Among the compound nouns, the predominant process involves combining two nouns, followed by compounds formed by pairing adjectives with nouns. This phenomenon likely occurs due to the inherent flexibility and versatility of noun combinations in English. Nouns often serve as the building blocks for compound nouns, allowing for the creation of new concepts and ideas by combining two or more nouns. Additionally, compound nouns formed from adjectives and nouns are also common, reflecting the descriptive nature of compound formations in English. In compound adjectives, the predominant formation involves combining a noun and an adjective, comprising 13 instances, followed by compounds formed by pairing two nouns, totaling 12 instances. And compound verbs predominantly result from the combination of nouns and verbs.

Among all compound words, a total of 31 contain combining forms, such as "agri-food" (agri- comb. form + food n.). Many combining forms used in word formation originate from Latin, Greek, French, and other languages. For instance, "multi-" is derived from Latin, and "hepta-" comes from Greek. These elements are commonly known as "neoclassical elements." Additionally, some combining forms are shortened versions of existing words, like "-ology" from "theology". In lexicographic practice, it is important to carefully document these combining forms, including their linguistic origins and variations, to ensure a thorough and accurate representation of their role in modern compound words.

4.2 Derivation

In the OALD, a total of 123 derivatives are documented. While this figure may appear modest in contrast to the abundance of compound words, it's crucial to acknowledge that derivation serves as a significant cornerstone for the expansion of English vocabulary. Among these derivatives, 72 are nouns, 37 are adjectives, and 14 are verbs. While the distribution trend of word classes is similar to that of compound words, with no instances of words having multiple parts of speech, the abundance of verbs surpasses that of compound words.

Derivation involves the formation of new words through the addition of prefixes or suffixes. Of these derivatives, 80 are created through the addition of suffixes, with common suffixes listed in Table 2. The remaining 43 words result from prefixation, with common prefixes also detailed in Table 2. The predominance of suffix-derived words over prefix-derived ones may stem from several factors. Firstly, English tends to utilize suffixes more frequently for word formation compared to prefixes. Suffixes often modify the grammatical function or meaning of a word, making them versatile tools for creating new vocabulary. Additionally, suffixes can be added to a wide range of words, allowing for a broader scope of derivational possibilities.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Table 2: Commonly-used affixes

Affix		Meaning	Example
suffix	-er	person or thing belonging to or associated with something	blueliner
	-ity	quality, state, or degree	inclusivity
	-ed	having; possessing; affected by	staged
	-less	not having something specified; without something	maskless
	-ist	a person who has a particular quality	ableist
prefix	de-	remove (a specified thing) from something	delist
	e-	electronic	e-scooter
	pre-	earlier than; before	pre-existing
	un-	opposite of; contrary to	unboxing
	anti-	opposite to something	anti-vaccine

4.3 Abbreviation

According to Bauer et al. (2013: 22), an abbreviation is "a shortened form of a word or phrase that retains the meaning of the original". They classify abbreviations into three types: (1) acronyms, which are pronounced as words and formed from the first letters of a phrase (e.g., NASA); (2) initialisms, which are pronounced as individual letters and formed from the first letters of a phrase (e.g., FBI); and (3) short forms, which are pronounced as a whole word and formed by shortening a longer word (e.g., ad for advertisement). And Gao (2021: 114) points out that abbreviations are mainly presented in three forms: shortened form, initialism and acronym, and clipping. This paper centers on the analysis of initialisms, acronyms, and clipped words, based on the actual collection of new words found in the OALD.

In the OALD, there are a total of 55 listed abbreviations. Among these, initialisms comprise 31 entries, making up a significant proportion of 56.36% of the total abbreviations. Of these initialisms, 22 are explicitly labeled as nouns, such as ASMR (autonomous sensory meridian response), while 9 initialisms lack explicit word class labels in the OALD, being only marked as "abbreviation" following the entry, for example, BBD (body dysmorphic disorder). Regarding their written representation, all 54 initialisms are displayed in full uppercase within the OALD. The sole exception is IoT (Internet of Things), where the middle "o" appears in

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

lowercase. These initialisms exhibit diverse applications across various fields, including technology (e.g., IoT, IP), medicine (e.g., MND, ALS), environmental science (e.g., AQI), transportation (e.g., BRT), and gender and sexual orientation (e.g., LGBTQIA+), among others. They serve to condense lengthy terms into concise representations, facilitating ease of communication and memorization, as seen in examples like ASMR and PPE. Many are widely recognized and utilized in both everyday language and professional discourse, while others, like DRS and SOP, have become industry standards within specific domains. Additionally, certain initialisms, such as NFT and IoT, signify emerging concepts and technologies, reflecting ongoing social and technological advancements.

In the OALD, the presence of only 9 acronyms compared to the larger number of initialisms reflects several notable distinctions between the two categories. Acronyms, unlike initialisms, are formed by combining the initial letters of a phrase to create a pronounceable word. This process typically requires a specific combination of words that lend themselves to such formation, which may be less common in comparison to the broad range of terms suitable for the creation of initialism. Additionally, acronyms often arise in specialized fields or contexts where the abbreviated term becomes widely recognized and adopted as a single word with a distinct meaning, as seen with COVID-19 (Coronavirus Disease 2019). While initialisms are prevalent across diverse domains due to their simplicity and efficiency in abbreviation, acronyms tend to emerge in specific niches or during pivotal events, such as in the case of HIIT (High-Intensity Interval Training) within the fitness industry or BRICS (Brazil, Russia, India, China, South Africa) in global economics. Therefore, the relatively limited number of acronyms in the OALD suggests that while they hold significant importance within certain specialized fields and cultural contexts, they may not be as prevalent or as easily formed as initialisms within the broader scope of language usage.

In the OALD, 12 clipped words are identified, primarily consisting of nouns, with the sole exception being the adjective "hetero," derived from "heterosexual." From a lexicographic perspective, these clipped forms highlight two main truncation patterns that must be carefully documented in dictionary entries. The first involves tail-end clipping, where the final portion of the word is removed, as seen in "crypto" (from "cryptocurrency") and "vax" (from "vaccine" or "vaccination"). The second pattern involves clipping from the middle of the word, as illustrated by "alt-rock" (from "alternative rock"). Occasionally, exceptions such as "inspo" (from "inspiration") appear, lacking a clear formation pattern. Lexicographers must pay attention to these diverse clipping processes, ensuring that the truncated forms are accurately recorded and linked to their full forms to provide users with a clear understanding of the word's origin and usage.

4.4 Conversion

Throughout the history of the English language, its vocabulary has undergone

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

significant changes. One of the most notable changes is the disappearance of inflectional endings in most words. Inflectional endings are suffixes or affixes added to a base word to indicate its grammatical function, such as tense, number, and case. Instead of relying on inflectional endings, English users have embraced the conversion method as the most straightforward and efficient approach to generating new words within English. According to Plag (2003: 134), conversion can be defined as "the derivation of a new word without any overt marking".

In the OALD, there are a total of 32 converted words. Within this category, verbs represent the largest group, comprising 13 entries, followed by nouns with 10, and 5 adjectives. Furthermore, there are 4 converted words that serve dual roles as both adjectives and adverbs, exemplified by terms like "trackside". Within converted words, the predominant transformation involves the transition from nouns to verbs. This conversion process reflects the dynamic nature of language, where nouns are adapted to convey actions or processes. For instance, "deadname" originates from the noun "dead name" and is converted into a verb to describe the act of referring to a transgender person by their birth name rather than their chosen name. Similarly, "mentor" transforms from a noun describing a guide or advisor into a verb representing the act of providing guidance or support to someone in their personal or professional development. This phenomenon also sheds light on why verbs dominate among converted words. Verbs, by their nature, encapsulate actions and processes, making them particularly adaptable for linguistic transformation. When nouns are converted into verbs, they often acquire dynamic qualities, allowing language to express a broader range of actions and behaviors.

Furthermore, within converted words, various other forms of conversion occur, including verbs transitioning into nouns, nouns transforming into adjectives, adjectives evolving into nouns, and even combining forms adapting into adjectives. This diverse array of conversion processes highlights the dynamic nature of language evolution, where words undergo transformations to fulfill different syntactic roles and semantic functions.

4.5 Blending

In the OALD, there are 21 blends listed, comprising primarily nouns, with only one adjective, "Paralympic". This adjective is a blend of "para-" from "paraplegic" and "-lympic" from "Olympic". For the sake of convenience in describing the pattern of blends, the first and second source words in the two-word blends are respectively labeled as A and B, with A1 and A2 representing the beginning and end of the first source word, and B1 and B2 representing the beginning and end of the second source word. Table 3 illustrates the different patterns of the blends.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Table 3: The structure of blends

Pattern	Total	Example
A1+B	9	alderwoman: alder- (in alderman n.) + woman n.
A1+B1	3	hazmat: haz- (in hazardous adj.) + mat- (in material adj.)
A1+B2	4	Paralympic: para- (in paraplegic n.) + -lympic (in Olympic n.)
A+B2	2	mockumentary: mock adj. + -umentary (in documentary n.)
Others	3	metaverse: meta- + -verse (in universe n.)

Most blends are created by truncating the end of the first word and merging it with the second word, or by removing parts of both words to form a new term. Additionally, some blends involve combining forms and affixes in the word-formation process. To determine whether a neologism should be added to the dictionary, lexicographers assess factors such as the blend's frequency in written and spoken language, its widespread acceptance, and its relevance to current cultural, social, or technological trends. The blend must demonstrate consistent usage and be recognized across various contexts before being incorporated as an official dictionary entry.

4.6 Other word-formation methods

Beyond the commonly observed processes of compounding, derivation, abbreviation, conversion, and blending, the lexicon of the OALD is enriched by the incorporation of new words formed through less conventional means. These include eponymy, where words are derived from the names of people or places, imbuing them with historical or cultural significance. Onomatopoeia contributes words that phonetically resemble the sounds they represent, offering vivid auditory imagery. Variation introduces linguistic diversity through regional or dialectal differences, showcasing the dynamic nature of language across different contexts and communities. These less-common processes not only broaden the scope of vocabulary but also reflect the diverse array of human expression and creativity woven into the ever-changing fabric of language evolution. Table 4 illustrates the number of new words generated through eponymy, onomatopoeia, and variation, accompanied by examples.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Table 4: Neologisms from other word-formation methods

Word-formation Methods	Total	Example
eponymy	6	Tabata™: from the name of Izumi Tabata, the Japanese physiologist who developed the activity.
onomatopoeia	3	nana: child's pronunciation of nanny or gran.
variation	6	masa: variant of master

4.6 Comparative analysis of the formation of English neologisms

In 1978, Cannon embarked on a comprehensive study delving into the intricacies of English word-formation. He focused on analyzing *6000 Words: A Supplement to Webster's Third New International Dictionary*. Subsequently, in 1982, Cannon extended this analysis to encompass 4520 newly coined words found in *The Second Barnhart Dictionary of New English*. Building upon this foundation, further examination ensued post-1987, encompassing a thorough scrutiny of 13,683 additional neologisms from *A Supplement to Webster's Third New International Dictionary*. In 1980, Algeo conducted a meticulous analysis by randomly selecting 1000 words from a pool of over 5000 newly coined words found within *The Barnhart Dictionary of New English*. Moreover, Bauer conducted a parallel investigation, randomly extracting 2082 words from *The Supplement to the Oxford English Dictionary*. These words were then segmented into three distinct temporal periods, demarcated by the years 1913 and 1938, serving as historical benchmarks. This enabled a comprehensive longitudinal analysis within the framework of the discussed morphological principles. The resulting statistical findings are detailed in Appendix 1.

In the OALD, the distribution of word-formation methods, including compounding, derivation, blending, and others, is shown in Appendix 2. Compounds account for nearly half of the total instances, indicating their substantial role in recent lexical expansion. Following compounding, derivation appears with a lower but still significant frequency. Abbreviation, encompassing initialisms, acronyms, and clippings, contributes to word formation as well, albeit to a lesser degree compared to compounding and derivation. Conversion and blending also play roles in word creation, albeit less prominently. Other methods such as eponymy, variation, and onomatopoeia are relatively infrequent. This data underscores the continued importance of compounding and derivation as primary drivers of lexical growth in contemporary English, with other mechanisms serving as supplementary avenues for word formation.

Upon thorough examination, it becomes evident that compounding is the primary method of word formation across various contexts. Particularly within the OALD, the number of words formed through compounding exceeds those formed through derivation. This observation aligns with Cannon and Algeo's

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

research, indicating a consistent trend. However, it contrasts with Bauer's findings. Bauer's methodology was influenced by his belief that newly created words resulting from shifts in meaning often display unique patterns, making systematic classification and statistical analysis difficult and potentially biased. Additionally, Bauer's study utilized random sampling, while Cannon's analysis concentrated on the complete collection of new words in the dictionary. These methodological differences help explain the discrepancies in findings between the studies.

5. The feature of English neologisms in the OALD

5.1 Wide-ranging sources for neologisms

The English lexicon is in a constant state of flux, characterized by ongoing evolution rather than stagnation. Minkova and Stockwell (2009: 5) note that "new words enter the language every day, words acquire or lose meanings, and words cease to be used". Neologisms in English play a pivotal role in mirroring the fluid and dynamic nature of language, serving as tools to encapsulate emerging concepts, technologies, and cultural shifts. As social changes unfold, the emergence of new ideas and phenomena necessitates the creation of precise vocabulary to articulate and convey these developments accurately. Reflecting this dynamism, the OALD regularly updates its entries every quarter, incorporating at least 160 newly coined or revised terms with each update.

These neologisms of the OALD stem from various fields including politics, economics, science, and culture. Economically, terms like "direct-to-consumer" (D2C) and "asset base" represent strategic shifts in business dynamics. "Direct-to-consumer" reflects a change in sales strategies, engaging directly with consumers, while "asset base" indicates financial stability and investment potential. In politics, terms such as "brand identity" and "brand recognition" highlight the marketing-driven nature of contemporary politics, shaping public sentiment. In science, concepts like "deep learning" and "data obfuscation" signify technological advancements in AI and data management, addressing privacy concerns. These linguistic changes reflect social evolution, showcasing progress across various human endeavors. For lexicographers, these entries underscore the necessity of tracking how new terms mirror shifts in societal trends and advancements across various domains. Documenting these neologisms helps to keep dictionaries aligned with current language use and emerging concepts.

5.2 The influx of foreign words into the vocabulary

English, as a universal language, surpasses national borders and linguistic boundaries. In addition to its primary use in countries such as the United States, the United Kingdom, and Canada, English holds official status in 55 other nations and regions, including India, Ghana, Jamaica, and New Zealand. Moreover, more

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

than 100 countries, including China and Russia, have embraced English as their preferred second language, solidifying its worldwide influence. According to Murugova and Evtushenko (2023: 2), "this complex geopolitical process creates a new worldview of modern man, makes great changes in languages, and strengthens the tendency to borrow vocabulary".

Within the OALD, there are precisely 52 loanwords, the number strikingly akin to the influx of novel terms facilitated by abbreviations. These loanwords traverse not only the linguistic realms of Greek, Latin, and French but also extend their roots into diverse language families. Among them are borrowings from Arabic, Korean, Japanese, Vietnamese, and beyond. For instance, "anosmia" hails from Greece, highlighting a loss of sense of smell, while "djembe" originates from French, denoting a West African drum. Additionally, loanwords include "bibimbap" from Korean, a savory mixed rice dish, "Wagyu" from Japanese, referring to prized beef, and "inshallah" from Arabic, conveying the hope for a desired outcome. These loanwords, originating from a variety of languages and carrying significant cultural weight, act as linguistic bridges that enhance the English lexicon. Lexicographers are essential in capturing and detailing these words, ensuring that their meanings and usages are accurately represented and integrated into the dictionary.

5.3 The simplicity of English neologisms

The trend of conciseness in English neologisms epitomizes the evolution of language in the digital era, where brevity and efficiency reign supreme. The proliferation of abbreviations and acronyms, such as "D2C" for direct-to-consumer, reflects the demand for streamlined communication in a fast-paced world driven by technology. Influenced by the Internet and social media in particular, people increasingly gravitate towards shorter forms of expression that can be quickly written, read, and understood. This linguistic efficiency not only saves time and effort but also aligns with the fast-paced exchange of information that defines modern digital communication.

Moreover, the trend towards conciseness in English neologisms mirrors broader social values and preferences. In an age where attention spans are shrinking and information overload is rampant, concise language offers clarity amidst the noise. By condensing complex ideas into succinct terms, English neologisms cater to the need for immediate comprehension and efficient communication across various domains, from business and technology to social media and everyday conversation. Thus, the rise of abbreviated and acronymic neologisms reflects not only linguistic adaptation to the digital landscape but also a broader cultural shift towards brevity and convenience in communication.

6. Conclusion

This research aims to analyze the word formation processes and characteristics of new entries introduced into the OALD lexicon between March 2020 and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

March 2024. In addition to commonly recognized processes like compounding, derivation, abbreviation, conversion, and blending, the OALD lexicon is enriched by incorporating new words formed through less conventional means, including eponymy, onomatopoeia, and variation. Among compound nouns, the majority are formed by combining two nouns, typically in the form of open compound words. Derivatives mostly arise from the addition of suffixes, while abbreviations primarily consist of initialisms written in uppercase letters. Converted words tend to be predominantly verbs and blends typically involve the removal of part of the first source word and its combination with the complete second source word. Overall, the sources of new words are diverse, with a notable presence of loanwords, and they generally tend to evolve towards simplicity.

The findings highlight the evolving nature of word formation and underscore the need for lexicographers to adapt their methodologies to accurately document new entries. By understanding the diverse processes through which new words are created, lexicographers can ensure that dictionaries remain current and reflective of contemporary language usage. This research contributes to the field of lexicography by providing a detailed analysis of recent word-formation trends and characteristics. It offers insights into the mechanisms behind new entries and their integration into the lexicon, thereby aiding in the development of more comprehensive and accurate dictionary entries. However, the study is limited by its focus on entries from a single dictionary (OALD) and a specific timeframe (March 2020 to March 2024). It does not account for potential variations in word formation trends across different dictionaries or periods. Future research could expand the scope to include additional dictionaries and longer timeframes to provide a broader understanding of word formation trends. Comparative studies could also be conducted to analyze how different dictionaries treat new words and to explore the influence of cultural and technological changes on word formation.

6. Endnotes

1. https://www.oed.com/information/updates/december-2024/
2. https://www.oxfordlearnersdictionaries.com/wordlist/new_words

References

Algeo, J. 1991. *Fifty Years among the New Words: A Dictionary of Neologisms, 1941–1991.* Cambridge/New York: Cambridge University Press.

Algeo, J. 1980. Where Do All the New Words Come from? *American Speech* 55(4): 264-277.

Bauer, L. 1994. *Watching English Change: An Introduction to the Study of Linguistic Change in Standard Englishes in the Twentieth Century.* London: Longman.

Bauer, L., R. Lieber and I. Plag. 2013. *The Oxford Reference Guide to English Morphology.* Oxford: Oxford University Press.

Bradley, H. 1904. *The Making of English.* New York/London: Macmillan.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Cannon, G. 1978. Statistical Etymologies of New Words in American English. *Journal of English Linguistics* 12(1): 12-18.

Cannon, G. 1982. A Linguistic Analysis of 4,520 New Meanings and New Words in English. *Dictionaries: Journal of the Dictionary Society of North America* 4: 97-109.

Cannon, G. 1987. *Historical Change and English Word-Formation.* New York: Peter Lang.

Chen, Jiansheng. 2000. A Study on the Word-Formation Trends of Newly Created English Words in the 20th Century. *Journal of Sichuan International Studies University* 3: 58-61.

Crystal, D. 1995. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Eliot, T.S. 1943. *Four Quartets.* Harcourt, Brace and Co.

Gao, Fengjiang. 2001. The Formation and Characteristics of English Neologisms. *Chinese Science & Technology Translators Journal* 4: 54-45.

Gao, Yongwei. 1998. Research on Neologisms in English-Speaking Countries in the Past 20 Years. *Foreign Languages and Foreign Language Teaching* 11: 2.

Gao, Yongwei. 2012. *Research on English Neologisms and Dictionaries.* Shanghai: Fudan University Press.

Gao, Yongwei. 2021. *A 21st Century English–Chinese Dictionary of Neologisms.* Shanghai: Fudan University Press.

Greenough, J.B. and G.L. Kittredge. 1901. *Words and Their Ways in English Speech.* New York: Macmillan.

Hartmann, R.R.K. and G. James. 2000. *Dictionary of Lexicography.* Beijing: Foreign Language Teaching and Research Press.

Jackson, H. 1988. *Words and Their Meaning*. London: Routledge.

Li, Ying. 2001. *An Analysis of Word-Formation of English Neologisms.* Liaoning: Dalian Maritime University.

Liu, Kelan. 2015. *Network English Neologisms from a Cultural Perspective.* Wuhan: Wuhan University Press.

Lu, Guoqiang. 1999. *Modern English Lexicology*. Shanghai: Shanghai Foreign Language Education Press.

Malmkjær, K. (Ed.). 2006. *The Linguistics Encyclopedia.* London/New York: Routledge.

Minkova, D. and R. Stockwell. 2009. *English Words: History and Structure.* Cambridge: Cambridge University Press.

Murugova, E. and S. Evtushenko. 2023. Neologisms and its Role in the Development of Modern Sustainable Ecosystemic Society *E3S Web of Conferences* 420.

Newmark, P. 2001. *A Textbook of Translation*. Shanghai: Shanghai Foreign Language Education Press.

O'Dell, F. 2016. Creating New Words: Affixation in Neologisms. *ELT Journal* 70(1): 94-99.

Plag, I. 2003. *Word-formation in English.* Cambridge: Cambridge University Press.

Wang, Rongpei. 2000. Tracking English Neologisms. *Foreign Languages and Foreign Language Teaching* 8: 10-14.

Wang, Yin. 2005. *Exploration of Cognitive Linguistics.* Chongqing: Chongqing Publishing House.

Yang, Bin. 2008. *The Doorbell of the Mind: A Cognitive Approach to English Neologisms.* Jiangsu: Soochow University.

Yang, Yonglin. 1997. The Expression of Culture in the Formation of English Neologisms: An Analysis of "-gate" and Its Compounds. *Journal of Northwest Normal University (Social Sciences)* 1: 51-56+104.

Yu, Nini, Zhang, Li, Ping, He, Jian. 2011. *The Generation, Formation, and Usage of English Neologisms.* Guangzhou: Guangdong People's Publishing House.

Yuan, Ningning. 2019. *Trend Analysis of English Word-Formation Based on OED Neologisms.* Guandong: Guangdong University of Foreign Studies.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Appendix 1: Previous scholarly studies on word-formation

Word-formation	Source	Cannon 1978	Cannon 1982	Cannon 1987	Algeo 1991	Algeo 1991		Bauer 1994		
	Number	4881	4520	13683	1000	3200	1200	824	613	641
Compounding		27.97%	33.76%	29.6%	29.8%	40.3%	36.3%	16.0%	15.8%	18.4%
Shortening	Back-formation	0.53%	2.01%	1.1%	1.4%	7.6%	10.0%	1.8%	2.6%	4.7%
	Abbreviation	4.14%	4.96%	3.4%	8.3%					
	Initialism		2.17%	1.1%						
	Clipping		1.73%	4.6%						
	Others	3.67%	14.04%							
Blending		0.72%		7.9%	4.8%	4.6%	9.8%	0.9%	2.3%	2.5%
Shifting	New Meaning	18.35%	11.60%	14.4%	7.7%	11.4%	19.4%			
	Conversion	1.99%	3.34%	4.1%	6.4%	6.0%				
	Other			1.2%	0.1%					
Borrowing		6.08%	7.06%	7.5%	6.9%	1.6%	4.3%	31.4%	22.3%	19.2%
Others	Variation	0.57%	1.51%	0.4%	0.5%	0.5%	2.2%	4.4%	5.4%	5.9%
	Uncertain	0.76%	0.24%							
	Onomatopoeia	0.23%								
	Eponymy	0.39%						2.1%	3.6%	3.6%

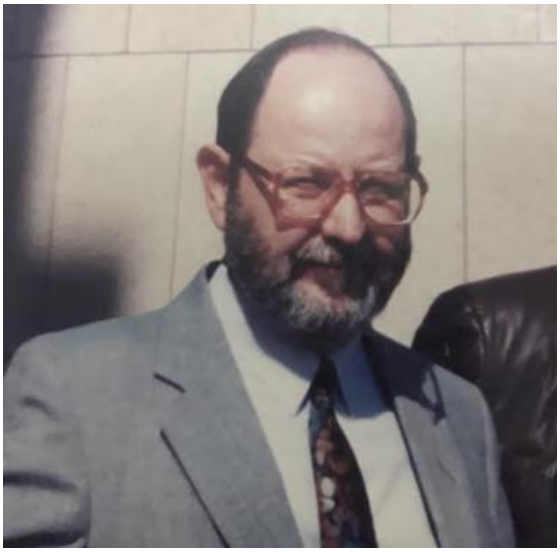
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1977 (Lexiconote)

Appendix 2: The proportion of word-formation methods in the OALD

Word-formation		Total	Percentage
Compounding		465	49.89%
Derivation		123	13.19%
Abbreviation	Initialism	31	3.32%
	Acronym	9	0.96%
	Clipping	12	1.28%
Conversion		32	3.43%
Blending		21	2.25%
Eponymy		6	0.64%
Variation		6	0.64%
Onomatopoeia		3	0.32%
Others		230	24.08%

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1978 (Lexikotribute)

Reinhard R.K. Hartmann
1938 – 2024



Ons het met diepe leedwese verneem van die heengaan van Reinhard R.K. Hartmann op 12 September 2024. Dr. Reinhard Hartmann het 'n sleutelrol in die ontwikkeling van die leksikografie oor die wêreld heen gespeel. Hy geniet erkenning en word gehuldig vir sy rol in die vestiging van die European Association for Lexicography (EURALEX), die African Association for Lexicography (AFRILEX) en die Asian Association of Lexicography (ASIALEX). Hy het die LEXeter-konferensie in 1983, wat 'n kernrol in die totstandkoming van EURALEX en sy vlagskippublikasie, *International Journal of Lexicography*, gespeel het, georganiseer. Reinhard het as EURALEX-sekretaris en vierde president van hierdie vereniging gedien. Erelidmaatskap van EURALEX is later aan hom toegeken.

Hy is welbekend vir sy omvattende baanbrekerswerk op praktiese en teoretiese aspekte van die leksikografie in publikasies soos *Contrastive Textology: Comparative Discourse Analysis in Applied Linguistics, Lexicography: Principles and Practice, The History of Lexicography, Lexicography in Africa* en *Dictionary of Lexicography. Lexicographica Series Maior.* Hy het baie jare lank as mederedakteur van die boekereeks *Lexicographica Series Maior* opgetree.

Op 14 Julie 1995 het verskeie leksikograwe, akademici en belanghebbendes by 'n konferensie van die African Languages Association for Southern Africa (ALASA) byeengekom om 'n professionele vereniging vir die Leksikografie te vestig, en so het AFRILEX tot stand gekom. Dr. Reinhard Hartmann, voorsitter van die European Association for Lexicography (EURALEX), was op hierdie

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1978 (Lexikotribute)

geleentheid teenwoordig en hy het die hele proses gefasiliteer. In 2005 het hy na Suid-Afrika teruggekeer om die 10de AFRILEX-konferensie in Bloemfontein by te woon. Dit was 'n gedenkwaardige geleentheid waar al die AFRILEX-presidente teenwoordig was.

Reinhard Hartmann het die wêreldbekende *Dictionary Research Centre* in Exeter, wat later na Birmingham verskuif is, tot stand gebring. Alhoewel hy altyd heel beskeie oor die waarde, omvang en impak van die Navorsingsentrum was, het hierdie sentrum oor die laaste twee dekades van die vorige eeu vir diegene met 'n belangstelling in die leksikografie dié plek geword om te besoek. Die Sentrum het baie navorsers van oor die wêreld wat die omvattende versameling gepubliseerde en ongepubliseerde leksikografiese hulpbronne as 'n insiggewende geleentheid vir die beplanning en uitvoering van pioniers- leksikografiese navorsing beskou het, gelok. Hierdie ervaring is verder versterk deur die stimulerende besprekings met Reinhard.

Reinhard was 'n uitstekende leermeester en studieleier en menige M.A.- en Ph.D.-verhandelings van studente van oor die wêreld is onder sy toesig voltooi. Hierdie studente sit nou die werk wat deur Reinhard aangevoor is, voort.

Reinhard Hartmann was instrumenteel in die bevordering van internasionale samewerking en het gehelp om 'n netwerk van leksikograwe wat toegewyd is daaraan om die beste praktyke met mekaar te deel, te vestig en om die veld te bevorder. Sy sogenaamde lys van kundiges in die leksikografie dien as voorbeeld hiervan.

Reinhard sal onthou word as een van die prominentste figure in die leksikografieveld.

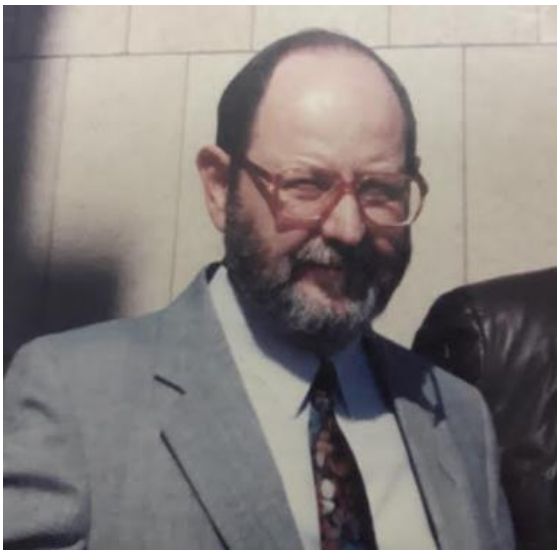
Danie J. Prinsloo
Universiteit van Pretoria
Suid-Afrika
(danie.prinsloo@up.ac.za)
(https://orcid.org/0000-0003-0054-4676)

en

Rufus H. Gouws
Universiteit Stellenbosch
Suid-Afrika
(rhg@sun.ac.za)
(https://orcid.org/0000-0002-3423-058X)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1978 (Lexikotribute)

Reinhard R.K. Hartmann
1938 – 2024



We were deeply saddened to learn that Reinhard R.K. Hartmann passed away on 12 September 2024. Dr. Reinhard Hartmann played a key role in the development of lexicography around the globe. He is recognised and honoured for his role in the establishment of the European Association for Lexicography (EURALEX), the African Association for Lexicography (AFRILEX) and the Asian Association of Lexicography (ASIALEX). He organised the LEXeter conference in 1983 which played a major role in the formation of EURALEX and its flagship publication, the *International Journal of Lexicography*. Reinhard served as EURALEX secretary and fourth president. Honorary membership of EURALEX was later bestowed on him.

He is well-known for his extensive ground-breaking work on practical and theoretical aspects of lexicography in publications such as *Contrastive Textology: Comparative Discourse Analysis in Applied Linguistics, Lexicography: Principles and Practice, The History of Lexicography, Lexicography in Africa* and *the Dictionary of Lexicography*. For many years he served as co-editor of the book series *Lexicographica Series Maior*.

On 14 July 1995 several lexicographers, academics and stakeholders got together at a conference of the African Languages Association for Southern Africa (ALASA) to establish a professional association for Lexicography, and so AFRILEX was born. Dr. Reinhard Hartmann, chairperson of the European Association for Lexicography (EURALEX) was present at this occasion and he facil-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1978 (Lexikotribute)

itated the whole process. In 2005 he returned to South Africa to attend the 10th AFRILEX conference in Bloemfontein, a memorable occasion when all of the AFRILEX Presidents were present.

Reinhard Hartmann established the world-famous *Dictionary Research Centre* in Exeter which was later moved to Birmingham. He was always quite modest about the value, size and impact of the Research Centre, but during the last two decades of the previous century this centre became the place to visit for those with an interest in lexicography. The Centre attracted many researchers from around the world who found the wide-ranging collection of published and unpublished lexicographic resources an informative opportunity for planning and executing pioneering lexicographic research. This experience was enhanced by the stimulating discussions with Reinhard.

Reinhard was an excellent teacher and supervisor and many MA and PhD dissertations by students from around the globe were completed under his supervision. These scholars now continue with the work started by Reinhard.

Reinhard Hartmann was instrumental in advancing international collaboration, helping to establish a network of lexicographers dedicated to sharing best practices and advancing the field, his so-called list of who's who in lexicography being an example.

Reinhard will be remembered as one of the greats in the field of lexicography.

Danie J. Prinsloo
University of Pretoria
South Africa
(danie.prinsloo@up.ac.za)
(https://orcid.org/0000-0003-0054-4676)

and

Rufus H. Gouws
Stellenbosch University
South Africa
(rhg@sun.ac.za)
(https://orcid.org/0000-0002-3423-058X)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

Du Ruiqing and Jiang Yajun. *New Century Chinese–English Dictionary*. Third Edition. 2025, XXXVIII + 2490 pp. ISBN 978-7-5213-5418-8 (Hardback). Beijing: Foreign Language Teaching and Research Press. Price $54.43.

The *New Century Chinese–English Dictionary* (NCCED) was first published by Foreign Language Teaching and Research Press in 2003. Adopting the guiding principle of "Readers first, practicality foremost" advocated by the late Professor Hui Yu, its editor-in-chief, the dictionary aimed to include words and expressions that could reflect the present times while paying special attention to the recording of neologisms that had emerged in the previous decade. With a coverage of 140,000 entries and a heavy focus on the use of phrasal illustrative examples, the dictionary gradually established itself as one of the most popular C–E dictionaries in China. Du Ruiqing took over as editor-in-chief of the second edition which was published in 2016, and the dictionary had 8,100 additional entries and 74,000 revisions, boasting a total of 150,000 entries. The inclusion of neologisms in the dictionary was rather praiseworthy as its compilers had made extra effort to record new words and expressions created because of economic and social development. The word 网络 (online or network) had 98 subentries, which is evidence enough of its comprehensiveness in the coverage of neologisms. Thanks to its wide coverage, excellent translation quality and successful app launch, NCCED2 has so far stood out as one of the best C–E dictionaries now available. In consequence, it has been widely reviewed or written about. A keyword search in CNKI, a database of Chinese academic papers, came up with over 50 reviews and papers.

The publication of NCCED3 has been long expected and desired as the past few years have witnessed the creation of hundreds or even thousands of neologisms in the Chinese language. The second edition, though still popular among its wide readership, has failed in many ways to meet their needs. The third edition, co-edited by the late Professor Du and Jiang Yajun, came out after a seven-year revision. As is mentioned in the introduction written by Jiang, the revision mainly revolved around the collection of new words and new meanings, and the improvement upon deficiencies related to the coverage, translation, and illustrative examples in the previous edition. The newness of the dictionary can be illustrated by the following two examples. The first example is related to the addition of the new medical term 新冠 (COVID-19) and many examples containing the term. In the entry for 新冠, three sub-entries have been furnished, namely 新冠肺炎 (old term for 新冠感染), 新冠感染 (COVID-19; coronavirus disease 2019), and 新冠疫苗 (COVID-19 vaccine). Besides, dozens of related examples have been provided throughout the dictionary text, such as 注射新冠疫苗 (vaccinate sb. against COVID-19), 新冠病毒感染疫区 (COVID-19-affected area), 新冠病毒奥密克戎变异株 (the Omicron variant of SARS-CoV-2), 保持社交距离 (enforce/practice social distancing), and 他新冠核酸检测呈阳性 (His COVID-19 nucleic acid test came back positive). The second example is related to one of the most popular buzz-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

words — 人工智能 (artificial intelligence; AI). The editors not only revised the entry through the addition of two examples (namely "人工智能技术 artificial intelligence technology" and "人工智能算法 artificial intelligence algorithm") and the provision of related entries such as 生成式人工智能 (generative artificial intelligence; generative AI) and 大语言模型 (large language model; LLM), but also provided three more related examples in other entries, namely 人工智能的创始人 (founding father of artificial intelligence), 人工智能产业方兴未艾 (The AI industry is on the upswing), and 人工智能是一种赋能技术 (Artificial intelligence is an empowering technology).

NCCED3 has included about 4,000 new entries, many of which were selected with the help of existing Chinese dictionaries, corpus data, and online search. These new words are used in various aspects of human life, as can be exemplified by the entries shown in Table 1:

Table 1: Types of neologisms recorded in NCCED3

Different types	Entries
Political or social buzz-words	人类命运共同体 (global community of shared future), 新质生产力 (new quality productive force)
Technological terms	人形机器人 (humanoid robot), 月面着陆器 (lunar/moon lander)
Environmental terms	碳达峰 (carbon dioxide peaking; carbon peak), 碳中和 (carbon neutrality)
Online words and expressions	秒回 (reply instantly; respond immediately; shoot back instantly), 破防 (breach one's psychological defenses; be emotionally overwhelmed)
Fashion terms	汉服 (*hanfu* [traditional styles of clothing worn by the Han ethnic group]), 马面裙 (*mamianqun* [a type of traditional Chinese skirt])
Proper names	长三角 (Yangtze River Delta), 粤港澳大湾区 (Guangdong-Hong Kong-Macau Greater Bay Area)

Besides, NCCED3 has also recorded the new uses of many existing words. 数字, for instance, had three senses in the second edition, namely "characters that denote amounts", "numeral; figure; digit", and "quantity; amount", and the new edition has recorded its adjective use — "digital". Other new meanings that have found their way into NCCED3 are shown in Table 2:

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

Table 2: New meanings recorded by NCCED3

Entries	Original sense(s)	New sense(s)
弹幕	barrage	bullet screen; bullet comment
拍	clap; racket; beat; shoot; flatter; auction	purchase on an online store
奇葩	exotic flower; extraordinarily beautiful flower	extraordinary literary or artistic work; oddball; weirdo; freak
搜索	look for; search (for)	[computing] search
团	round; roll; regiment; body; group; ball-shaped object …	group-buy
雾化	atomize	nebulize; nebulizer therapy/treatment
应用	apply; use; put to use; applied	app; application
云	cloud	[computing] cloud

Thirdly, the editors of NCCED3 have provided many existing entries with illustrative examples which not only encompass phrases and short sentences, but also include compound words or terminology formed with the entry word, as is exemplified by the entries in Table 3:

Table 3: New illustrative examples in NCCED

Entries	New illustrative example(s)
电子烟（e-cigarette）	吸电子烟 (vape; vape an e-cigarette)/吸电子烟会损伤你的DNA (vaping could damage your DNA)
峰会 (summit)	全球气候峰会 (global climate summit)
付款 (pay a sum of money)	扫码付款 (pay by scanning a QR code)
射击 (shooting)	残奥会射击 (shooting Para sport)
首演 (debut)	参加首演 (attend the premiere)
搜索引擎 (search engine)	搜索引擎优化 (search engine optimization)
网购 (online shopping)	网购可以节约大量时间 (shopping online is a great time-saver)
新常态 (new normal)	进入新常态 (enter a new normal)
延迟 (delay; postpone)	延迟退休 (delay retirement)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

Moreover, considerable improvement has been made to the new edition, which is manifested in the following two aspects: the improvement of the translations for headwords and the replacement of existing illustrative examples. The editors made special efforts to amend the translations of many entries. 香菇, for instance, was translated as "fragrant mushroom; wood mushroom" which has been changed into "Chinese (black) mushroom", apparently a much better equivalent. Other entries whose translations have been amended are shown in Table 4:

Table 4: The amendment of translations

Entries	Original translation(s)	Amended translation(s)
热搜	hot search list; trending list	hot/trending search; trending hashtag/topic
包子	baozi; steamed stuffed bun	bao; baozi; bao bun; steamed stuffed bun
飞碟射击	skeet shooting	clay pigeon/target shooting
电子手账	electronic organizer; personal digital assistant	digital planner; electronic planner
群主	group master	group administrator/owner
木鱼	wooden fish	wooden fish; fish-shaped woodblock; temple block
电子烟	electronic cigarette	electronic cigarette; e-cigarette
相声	xiangsheng; comic dialogue; crosstalk	xiangsheng; cross talk

Translation-wise, the adoption of the pinyin system for culture-specific words in NCCED3 is rather commendable. As a matter of fact, such practice has long been adopted by C–E dictionary-makers. In NCCED1, the editors translated 粽子 through a combination of transliteration and explanation, namely "*zongzi* — pyramid-shaped dumpling made of glutinous rice wrapped in bamboo or reed leaves [usu. eaten during the Dragon Boat Festival]". In the third edition, the editors have adopted transliteration in the translation of more culture-specific terms, such as *fensi* for 粉丝 (cellophane noodles), *Huagu* for 花鼓 (flower drum), *meihuazhuang* for 梅花桩 (plum-blossom stake), *tangbao* for 汤包 (steamed dumplings filled with juicy meat and gravy), *tangyuan* for 汤圆 (stuffed dumplings made of glutinous rice flour served in soup), and *tanghulu* for 糖葫芦 (skewered fruit coated in hardened sugar syrup). Besides, they have also made some changes to the existing body of illustrative examples, the purpose of which is to make them keep pace with the changes of the times or to improve the wording. In the entry for 邮件 (email), for instance, one of the original examples — 备份邮件到 CD (backup [sic] emails to CD) — has been replaced by 备份电子邮件到云存储账户

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

(backup emails to one's cloud storage account). Similar changes have been made to the entries shown in Table 5:

Table 5: The replacement of some illustrative examples

Entries	Previous example	Current example
倡导 (advocate)	倡导计划生育 (advocate family planning)	倡导全民阅读 (promote nationwide reading)
手机 (mobile)	3G 手机 (3G mobile)	5G 手机 (5G mobile)
倘若 (if; supposing)	倘若那嫌疑犯开枪，就当场击毙他 (Should the suspect open fire, shoot him dead right away.)	倘若那嫌疑人开枪，就当场击毙他 (Should the suspect open fire, shoot him dead right away.)
物有所值 (be good value for money)	买这辆车要花5万美元，但还是物有所值 (The car will cost $50,000 but it is cheap at the price.)	买这辆车要花30万元，但还是物有所值 (The car will cost 300,000 yuan but it is cheap at the price.)
一线 (front line)	抗击非典的一线医护人员 (doctors and nurses in the front line of the fight against SARS)	抗击新冠的一线医护人员 (doctors and nurses in the front line of the fight against COVID-19)

NCCED3 is also praiseworthy for its provision of some new features that include 翻译点津 (Translation suggestions), 翻译速查 (Translations for related terms), and 话题词汇 (Related words and expressions). There are altogether 132 boxes containing very useful information that is conducive to improving one's translation skills and enlarging one's vocabulary as well.

However, NCCED3 still leaves room for improvement when it comes to the selection of headwords, the coverage of neologisms, the translations of the entries and illustrative examples, and the provision of examples. Some of the words in the dictionary, neologisms in particular, are chosen purely for the fact that their English equivalents are neologisms in the first place even though these Chinese counterparts are not being frequently used in Chinese. This deficiency can be illustrated by its inclusion of 网购踩点 (literally "examination before online shopping"). The term can be considered a nonword as there is virtually no result when searched in search engines like Google, and it is a far-fetched equivalent for the English neologism *showrooming* which is defined by *Oxford Dictionary of English* as "the practice of visiting a store or stores in order to examine a product before buying it online at a lower price". Other terms in NCCED3 that have been selected not in accordance with its usual criteria

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

include 不首先使用 (no first use), 池浮游生物 (heleoplankton, known more as 沼泽浮游生物), 臭氧消耗物质 (ozone-depleting substances, known more as 消耗臭氧层物质), 数字基础设施 (digital infrastructure), 网站人流量 (web traffic), etc.

The failure to record some of the latest Chinese neologisms has plagued C–E dictionaries for years as it was quite difficult for Chinese dictionary-makers to fully record the new lexicon due to various reasons. NCCED3 is not immune to this failure. As the second volume of *The Chinese English Dictionary (Unabridged)* (CEDU) was published in June 2024, we can make a comparison of the new entries recorded in the two dictionaries (as is indicated in Table 6).

Table 6: A comparison of NCCED and CEDU in their coverage

Main entry	NCCED	CEDU
免疫 (immunity)	30 entries (e.g. 免疫反应 [immune response] and 免疫抗体 [immune antibody])	15 entries (e.g. 免疫缺陷 [immunodeficiency] and 免疫原 [immunogen]) and 26 sub-entries (e.g. 免疫测定 [immunoassay] and 免疫信息学 [immunoinformatics])
纳米 (nanometer)	3 entries (e.g. 纳米材料 [nanomaterial] and 纳米技术 [nanotechnology])	4 entries (e.g. 纳米机器人 [nanorobot] and 纳米纤维 [nanofibre]) and 17 sub-entries (e.g. 纳米计算机 [nanocomputer] and 纳米医学 [nanomedicine])
气候 (climate)	5 entries (e.g. 气候疗法 [climatic treatment] and 气候学 [climatology])	one entry (气候变化 [climate change]) and 30 sub-entries (e.g. 气候罢工 [climate strike] and 气候危机 [climate crisis])
碳 (carbon)	52 entries (e.g. 碳汇 [carbon sink] and 碳信用 [carbon credit])	93 entries (e.g. 碳补偿 [carbon offset] and 碳足迹 [carbon footprint])

Some newly emerged Chinese words have been conspicuously missing from NCCED3. The character 脑 (brain) has formed many compounds in the past couple of years, such as 脑洞 (avenue to the brain; brain crease), 脑补 (figure out the picture or actual meaning), 脑雾 (brain fog), and 脑腐 (brain rot). However, these words have failed to make their way into NCCED3. Also missing from NCCED3 are some widely known Chinese food items or culinary terms like 韭菜盒子 (fried Chinese leek dumpling), 麻辣烫 (hot, spicy soup of meat and vegetables), 螺蛳粉 (*luosifen*), 香锅 (spicy numbing stir-fry pot), etc. As far as the recording of new scientific terminology is concerned, NCCED3 has also done a poor job. China National Committee for Terms in Sciences and Technologies (CNCTST), a governmental agency in charge of regulating the use of terminol-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

ogy, has recently released the top ten scientific and technological terms of the year 2024, namely 人工智能+ (artificial intelligence plus), 低空经济 (low-altitude economy), 月背采样 (take samples from the far side of the moon), 量子科技 (quantum technology), 具身智能 (embodied intelligence), 混合现实 (mixed reality), 生物制造 (biomanufacturing), 近零碳 (nearly zero-carbon), 实景三维 (real-scene 3D), 微核糖核酸 (microRNA, miRNA). Unfortunately, NCCED3 has recorded none of them. In the same vein, some new meanings of existing words have been not recorded by NCCED3, such as 钓鱼 (entrap somebody; originally "angle; fish"), 队友 (mate, one's husband or wife; originally "teammate; fellow player"), 落地 (implement, put into effect; originally "fall to the ground"), 赛道 (track; originally "racetrack"), 虚拟 (virtual; originally "suppositional; hypothetical; invent"), 野蛮 (unrestrainedly; originally "uncivilized; savage"), and 移动 (its adjective use "mobile"; originally "move; shift").

Moreover, NCCED3, though generally viewed as the go-to dictionary for translators, has a mountain to climb as regards the quality of translations for some of its headwords and illustrative examples. The translations for many words beginning with the character 网 are rather problematic. 网红店 (Instagrammable store), for instance, has been translated into "trending shop" and "trending store", but these two translations are not idiomatic in the English language: there is no record of *trending shop* and one occurrence of *trending store* (appeared in an Indian media outlet) in the NOW corpus. Other entries whose translations need improvement are listed in Table 7:

Table 7: Suggested amendment for the translation of headwords containing 网

Entry word	Current translation(s)	Suggested version(s)
网红打卡地	trending photo spot	Instagrammable spot
网络红人	Internet sensation; Internet/web celebrity; blue-eyed boy/girl on the Internet	Internet sensation; online celebrity; influencer
网络水军	online hirelings	army of paid posters online
网络文学	net literature	online literature
网络语言	net talk; net language; cybertalk; Internetese; netspeak	cyberspeak; cybertalk; netspeak
网络主播	network anchor	online anchor; livestreamer

Similar problems can be identified in the translations of other categories of words. A case in point is 肉包子 which was translated into "steamed bun with stuffing of minced meat" in NCCED1, and in the second edition, two more equivalents — *bakpau* and *bah-pau* — were put before the previous translation. These two additions, however, are not English words. According to *Wiktionary*, *bakpau,* a variant

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

spelling of *bakpao,* originated from Xiamen Hokkien and is an Indonesian word while *bah-pau* seems to be its variant form. What caused the confusion may be attributed to the fact there is indeed a kind of Hokkien snack popular in Singapore and Malaysia that is known as *kong bak pau* (扣肉包). Therefore, it is suggested that deletion should be made of these two equivalents. Improvement of translations can also be suggested for the entries in Table 8:

Table 8: Suggested translations for various kinds of words

Entries or subentries	Existing translation	Suggested translation or improvement
低调	low-key	low-key; low-profile
木须肉	muxu pork	moo shu (pork)
人肉搜索	*renrou* search; human flesh search; cyber manhunt ...	doxxing
手机	mobile/cellular phone; mobile; handset	mobile/cellular phone; cell phone; mobile; celly
小白菜	pak choi; bok choy, small Chinese cabbage [a variety of Chinese cabbage]	bok choy; pak choi; Chinese cabbage
远程办公	telework; telecommute	telecommute; work remotely

Finally, the excessive use of transliteration may sometimes backfire. 龙, for instance, has been widely accepted as the Chinese equivalent for *dragon*. But as 龙 conveys a positive meaning in the Chinese culture and *dragon* a negative one in the Western culture, many people believe the use of *dragon* as the equivalent of 龙 to be inappropriate. As a result, there has been some clamor for the use of *loong* to replace *dragon* in academia, and the editors of NCCED3 have obviously sided with this view by adopting *loong* (as opposed to *long* which unfortunately is a homonym of the English word *long*). As *loong* is by no means in accordance with the pinyin system, we believe *Chinese dragon* which has been recorded by the OED will do in this case. Other entries in which the use of pinyin is unnecessary include 国潮 (*guochao*; Chinese trend; China chic), 摔碗酒 (*shuaiwanjiu; shuaiwan* wining [drink-and-smash-wining]), 唐装 (*tangzhuang; tang* suit [a type of traditional Chinese garment]), 网红 (*wanghong*; Internet/online celebrity ... influencer), etc.

Meanwhile, the translations for some illustrative examples in NCCED3 are unidiomatic or may contain grammatical errors. Let's look at one example for 构建 (construct; establish; constitute) — 构建中国话语和中国叙事体系 which was translated into "establish our own discourse and narrative". This translation apparently sounds unidiomatic and even smacks of Chinglish, therefore it should be replaced by "develop China's discourse and narrative system". Similarly, the

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1981

translations for the following examples should be improved upon, such as "营造清朗的网络空间 ensure a clean and healthy cyberspace" (the article *a* should be deleted), "国产片又甩出了王炸 Domestic cinema has thrown out another ace." (it might be replaced by "There is a new blockbuster in the domestic box office."), "春节相互拜年是中国的传统习俗 It is a traditional Chinese custom to exchange visits at the Spring Festival" (the latter part might be changed into "pay each other visits during the Spring Festival"), and "党内斗争不久就公开了 The struggles within the party soon came/broke into the open" ("infighting" is a much better choice than "struggles").

NCCED3 would have been better if more examples were provided for some frequently used terms. 社交媒体 (social media), for example, has been a popular term for years, but no examples have been furnished for the entry. In contrast, there are three illustrative examples for this entry in CEDU. There are many other entries that would prove more useful if illustrative examples were furnished, such as 炒热 (boost the popularity through repeated media coverage/exposure), 房奴 (mortgage slave), 各行各业 (all trades and professions; all walks of life), 路怒 (road rage), 始作俑者 (the originator of an evil), and 网申 (apply online).

It has been more than 200 years since the first Chinese–English dictionary (namely Robert Morrison's *A Dictionary of the Chinese Language*) was compiled. Traditional lexicography has been facilitated by technological advances. In this era of generative AI, dictionary-makers should make good use of what AI can offer when making and revising dictionaries. The editors of NCCED3 did indeed resort to the use of AI chatbots in proofreading part of the dictionary text, but AI should have played a bigger role throughout the whole revision process. Should they have made full use of AI, the quality of the dictionary, especially that of the translations, would have been greatly improved. Despite its inadequacies, NCCED3 will continue to prove to be one of the most sought-after C–E dictionaries in China.

References

Du, R. 2016. *New Century Chinese–English Dictionary*. Second Edition. Beijing: Foreign Language Teaching and Research Press.

Hui, Y. 2003. *New Century Chinese–English Dictionary*. Beijing: Foreign Language Teaching and Research Press.

Lu, G. 2024. *The Chinese–English Dictionary (Unabridged).* Volume II. Shanghai: Fudan University Press.

Oxford Dictionary of English. Oxford: OUP. Available at: http://www.oed.com

Oxford English Dictionary. Oxford: OUP. Available at: http://premium.oxforddictionaries.com

Yongwei Gao
College of Foreign Languages and Literature
Fudan University
China
(ywgao@fudan.edu.cn)
(https://orcid.org/0009-0008-1532-8742)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1982

Edward Finegan and Michael Adams. *The Cambridge Handbook of the Dictionary*. 2024, 801 pp. ISBN 978-1-108-83624-1 (Hardback), ISBN 978-1-108-81887-2 (Paperback), ISBN 978-1-108-86443-5 (eBook). Cambridge/ New York, NY: Cambridge University Press. Price: US$ 170.00 (Hardback), US$ 170.00 (eBook)

Before the publication of *The Cambridge Handbook of the Dictionary*, a number of handbooks on lexicography (Durkin 2016; Fuertes-Olivera 2018; Hanks and De Schryver 2020; Hartmann and James 1998; Jackson 2013, 2022) had already established a clear and well-defined process of dictionary making. These earlier works focused primarily on practical aspects such as orthography, sense division, phonetic transcription, collocation, corpus construction, and the use of labeling and metalanguage, treating dictionaries largely as utilitarian reference works. However, they frequently disregarded the broader social implications of lexicography, thus omitting crucial non-lexicographical subjects necessary for a deeper understanding of dictionaries. Dictionaries, in fact, serve purposes far beyond mere tools — they act as lenses that illuminate cultural nuances, as well as records that capture the essence of specific times, places, and human activities. It is within this context that *The Cambridge Handbook of the Dictionary* emerges, offering a holistic perspective of dictionaries in their social contexts. By examining the dynamic interplay among dictionary makers, users, cultural contexts, historical aesthetics, and the material embodiment of ideas, this handbook bridges the gap left by earlier works. Bringing together cutting-edge research from an international team of experts, it aims to foster dialogue among lexicographers while inviting a broader audience to approach dictionaries with more thoughtfulness.

This handbook consists of 31 chapters in six parts, covering a wide array of topics, from dictionary types and visual design to their role in book history and significance as material objects. It also delves into the ideologies embedded in dictionaries, their role in specialized fields like law and history, and the business contexts of their production. Additionally, it looks ahead to the future of dictionaries. Below is a brief introduction to each part within the handbook.

Part One introduces dictionary typology as a classification system to delineate the characteristics of dictionaries. However, due to the variations both within and across the types, their classification remains inherently complex. This part provides a wide scope and depth of dictionary types, including historical and specialized dictionaries, thesauri, general-purpose or commercial dictionaries, bilingual dictionaries, learner's dictionaries, children's dictionaries, and dictionaries of quotations and of proverbs. It stresses their crucial roles in cross-cultural communication, historical documentation, systems of cultural concepts, pedagogy, reference, as well as entertainment. Each type is meticulously curated by lexicographers to meet the diverse needs and interests of their respective user bases.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1982

Part Two concentrates on dictionaries as books, exploring their distinctive features such as typography, illustrations, page layout, and book design. These elements enhance dictionaries' appeal to readers and facilitate ease of reference. Additionally, the chapters in this part emphasize the significance of dictionaries as tangible artifacts in book history, highlighting their materiality, intrinsic value, influence, and status.

Part Three delves into ideological underpinnings of dictionaries, both directly and indirectly embedded in the usage, language contact, language attitudes, religious history, etc. Usage information in dictionaries reflects ideological preferences, mirroring potential social hierarchies. The complexity of language contact in dictionary production is inextricably bound up with socio-political relationships. Attitudes towards dictionaries reveal the extent of authority granted to them and how ideologies — political, social, or personal biases — are manifested, thereby influencing the inclusion and exclusion of dictionary entries, definitions, paratexts and beyond. Also, dictionaries involved in interpreting texts in domains such as religion are inherently shaped by ideological considerations.

Part Four discusses the domains of use of dictionaries, accentuating their mutual impact and value to the fields of linguistics, philology, writing and editing, and law. It also reflects their roles in intellectual, social, and cultural history, as well as their significance as literary artifacts where literature unfolds.

Part Five details the commercial aspects of dictionary making, such as preparation, planning, organization, editing and publication. Operating under organizational and commercial constraints, dictionary projects require extensive collaboration among diverse teams, encompassing editors, assistants, copyeditors, proofreaders, marketers, and sales teams. Legal and ethical considerations further affect the internal dynamics of dictionary business, ultimately shaping the nature and quality of the final product.

Part Six reviews the past and present, and primarily explores the future trajectory of dictionaries. As interest in print dictionaries wanes, attention has increasingly shifted towards digital dictionary compilation, a process driven by machines, enabling faster, automated analysis and continuous real-time updates. This transition is particularly evident in the adaptation of dictionaries to web structure and online use, revealing the evolving nature of dictionary use in the digital era.

Overall, one of the most notable features of *The Cambridge Handbook of the Dictionary* is its all-encompassing account of dictionaries in context. Throughout the handbook, each part, to some degree, concerns dictionaries in their social contexts, and addresses the relationships such as those among dictionary makers, users, and cultural contexts. No other work on this topic today is as detailed or compelling. For instance, previous handbooks like Jackson's *The Bloomsbury Companion to Lexicography* (2013, 2022), Hanks and De Schryver's *International Handbook of Modern Lexis and Lexicography* (2020), and Durkin's *The Oxford Handbook*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1982

of Lexicography (2016) have given scant attention to the issues of ideology and society. Similarly, in Fuertes-Olivera's *The Routledge Handbook of Lexicography* (2018), ideology is only briefly mentioned in the context of cultural considerations, whereas *The Cambridge Handbook of the Dictionary* dedicates considerable space to portraying how values, ideologies and attitudes shape dictionaries, revealing societal norms and personal biases. Linguists, in fact, have identified ideology in dictionaries in some detail (e.g. Kachru and Kahane 1995). Arguably, the most significant contribution of this handbook is its extensive treatment of lexicographic subjects such as ideology and relational dynamics, enriching readers' understanding of dictionaries in the present day. This emphasis makes the handbook particularly valuable for scholars and practitioners concerned with the convergence of lexicography, ideology, and social contexts.

Another significant aspect of *The Cambridge Handbook of the Dictionary* is its effort towards achieving sufficiency in the coverage of important and diverse dictionary types. These types range from historical dictionaries to thesauri and to children's dictionaries and dictionaries of quotations and proverbs, with this handbook exploring their development, structures, features, and influence in great depth. By comparison, *The Bloomsbury Companion to Lexicography* and *International Handbook of Modern Lexis and Lexicography* neglect to address specific dictionary types. *The Oxford Handbook of Lexicography* categorizes specialized dictionaries based on their microstructural features such as place names, surnames, spelling, and pronouncing. Approaching dictionary classification through functions and purposes — text production, reception, and translation, *The Routledge Handbook of Lexicography*, though it touches on specialized dictionaries, does not provide as detailed a breakdown of types as *The Cambridge Handbook of the Dictionary*. Instead, it emphasizes broader criteria such as the degree, range, and type of specialization, along with the number of languages used, and macro- and microstructural aspects. This distinction underscores *The Cambridge Handbook of the Dictionary* as an invaluable resource for those seeking a comprehensive understanding of the diverse types of dictionaries and their uses.

Additionally, the handbook presents a unique perspective by viewing dictionaries not only as reference tools but also as books, material artifacts, and commodities. It unveils the complexities of the relationship among dictionary makers, dictionary content, and their users. These aspects receive substantial attention, particularly in the parts covering domains of use and business of dictionaries, which were often treated briefly in earlier handbooks. By exploring the commercial and contextual dimensions of dictionaries, this handbook highlights their role influenced by business decisions, consumer demands, technological advancements, and social contexts.

Like any scholarly endeavor, *The Cambridge Handbook of the Dictionary* is not without its limitations. While the handbook offers a comprehensive overview of the state of the field within this significant sub-discipline of linguistic research, it notably falls short in its discussion of the use of corpora in the lexicographic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1982

process. In contrast, earlier handbooks such as *The Routledge Handbook of Lexicography* and *The Oxford Handbook of Lexicography* devote substantial sections to the construction and practical application of corpora in dictionaries, acknowledging corpus-based lexicography as the standard practice (Kilgarriff 2000; Kilgarriff and Kosem 2012). *International Handbook of Modern Lexis and Lexicography* explores electronic and corpus-driven approaches in modern lexicography across all major and rare languages. Similarly, *The Bloomsbury Companion to Lexicography* discusses the role of corpora as essential data sources for dictionaries and highlights their significant contributions to various aspects of dictionary compilation. The lack of a more in-depth discussion on this topic in *The Cambridge Handbook of the Dictionary* — likely due to thematic constraints rather than an oversight of lexicographic principles — indicates a missed opportunity to thoroughly engage with advancements in corpus-based research relevant to lexicography.

To conclude, *The Cambridge Handbook of the Dictionary* addresses pivotal issues in lexicographic scholarship, offering an unparalleled and comprehensive overview of the field. By integrating diverse perspectives, contemporary theories, and findings, it presents a unified perspective on dictionary studies, demonstrating both breadth and depth. These elements collectively establish it as an indispensable reference for scholars and practitioners. Moreover, the handbook catalyzes future scholarly inquiry into the dynamic relationship between dictionaries and society, pointing towards potential avenues for lexicographic studies.

Acknowledgements

This study was supported by the Jiangxi Province Social Science Fund Project（江西省社会科学基金项目）(Grant No. 23YY27D) and by the Jiangxi Province Higher Education Institutions Humanities and Social Sciences Research Project（江西省高校人文社会科学研究项目）(Grant No. YY22214).

References

Durkin, P. (Ed.). 2016. *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press.

Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*. Abingdon, UK/New York: Routledge.

Hanks, P. and G.-M. de Schryver (Eds.). 2020. *International Handbook of Modern Lexis and Lexicography.* Berlin: Springer.

Hartmann, R.R.K. and G. James (Eds.). 1998. *Dictionary of Lexicography.* London/New York: Routledge.

Jackson, H. (Ed.). 2013, 2022. *The Bloomsbury Companion to Lexicography*. London/New York: Bloomsbury.

Kachru, B.B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies and the Dictionary: Studies in Honor of Ladislav Zgusta.* Tübingen: Max Niemeyer.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1982

Kilgarriff, A. 2000. Business Models for Dictionaries and NLP. *International Journal of Lexicography* 13(2): 107-118.

Kilgarriff, A. and I. Kosem. 2012. Corpus Tools for Lexicographers. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 31-55. Oxford: Oxford University Press.

Huahua Jiang
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou, China;
School of Foreign Studies
Jiangxi University of Science and Technology
Ganzhou, China
(jasonjiang41513@163.com)
(https://orcid.org/0009-0008-1471-7746)

and

Hai Xu
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou, China
(Corresponding Author, xuhai1101@gdufs.edu.cn)
(https://orcid.org/0000-0003-4644-9033)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1983

Stephen Turton. *Before the Word was Queer: Sexuality and the English Dictionary, 1600–1930.* 2024, xv + 333 pages. ISBN 978-1-316-51873-1 (Hardback). Cambridge: Cambridge University Press. Price: US$ 94.67 (Hardback)

Positioning English dictionaries as crucial historical witnesses, Stephen Turton's monograph *Before the Word was Queer* provides an in-depth analysis of sexual language over a period of 330 years (1600–1930). It not only enhances our understanding of sexual linguistics but also highlights the role of historic dictionaries in unlocking the queerness of sexual history.

This volume comprises five chapters, culminating in a conclusion. Each chapter delves into specific themes related to sexuality, especially focusing on terms concerning acts, desires, and identities grounded in relations between women and between men. Through its analytical approach, the work contributes to a deeper understanding of queer linguistics.

The first half of Chapter 1 examines the interplay between sexual practices and the laws of divine, natural, and human standards. Prior to the establishment of the term *homosexual*, lexicons pertaining to same-sex relationships were dominated by the terms *buggery* and *sodomy*. These terms exhibit three main characteristics. First, they were heavily laden with the implications of legal, moral, and natural transgression. Second, their interpretations demonstrated semantic ambiguity throughout the seventeenth and early eighteenth centuries, lacking explicit definitions and clear references to same-sex acts. Third, although both terms were predominantly depicted as acts occurring between sexes, their participatory scope usually displayed variability. Early legal lexicographers defined *buggery* as encompassing bestiality, male-male relations, and sometimes female-female relations, though some dictionaries excluded the latter. In contrast, *sodomy* primarily referred to male-male acts. However, by the seventeenth century, these two terms had converged to share identical participatory scopes. The second half of Chapter 1 examines *sodomy* and *buggery* in relation to semantically broader sexual terms such as *copulation*, *coupling*, *occupying*, and *fucking*, revealing two prominent features of sexual language: underlying androcentrism and cross-sex normativity. Historically, lexicographers refrained from offering explicit details about *copulation*, limiting its definition to heterosexual intercourse. It was not until the seventeenth century that the term *copulation* embraced erotic senses. Coupled with this cross-sex normativity, androcentrism also shaped the dominant paradigm of sexual vocabularies in historic dictionaries. Verbs like *swive*, *fuck* and *occupy*, despite their unmarked grammatical subject, always presupposed male agencies. This lexical framework reinforced androcentrism perspectives while marginalizing female identities. Dictionaries serve as essential repositories of sexual terminologies, documenting both semantic interpretations and syntactic structures. The definitions of *sodomy* and *buggery* in historical dictionaries were influenced by legal and religious perspectives, reflecting the socio-cultural contexts of their respective eras. Linguistic analysis through dictionaries acts as a key

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1983

to unlocking the queerness of sexual linguistics, offering valuable insights into the historical discourse on sexuality and the nuanced shifts in the semantic interpretations of sexual terms over a period of 330 years (1600–1930).

Chapter 2 begins by examining the etymologies of *sodomy* and *buggery*. The term *sodomy* was rooted in the biblical story of Sodom and Gomorrah, while *bugger* was associated with medieval Bulgaria. Historically, these terms were framed as vices associated with ancient pagans, depicted as unnatural and foreign sins opposed to Protestant ethics, British customs, and national identity. To preserve the national identity and prevent the perceived negative impact of such loanwords, lexicographers often marginalized and relegated them to the periphery of linguistic discourse. Dictionaries, as repositories of language, not only offered definitions but also included etymologies, revealing the intricate relationship between sexual language and national identity. By excluding unnatural and inauthentic elements, dictionaries could better preserve and nurture a nation's unique character and cultural heritage. The origins of these sexual terms enable us to better understand their perceived exoticism and moral condemnation, as well as the rationale behind their marginalization or exclusion in historical dictionaries. Beyond *buggery* and *sodomy*, practices such as *pederasty* — denoting men's affection for boys and women's affection for girls — were also commonly labeled as unnatural and immoral, a view rooted in ancient Greece and Rome. Then, Chapter 2 transitions to an analysis of *pederasty*-related terminologies, particularly male-boy relations, to explore the concepts of male sexual passivity and dyadic recursivity. Terms like *catamite*, *ingle*, and *bardash*, which denote receptive male partners, illustrate both syntactic and sexual passivity. *Pederasty* challenges cross-sex normative conventions by establishing a unique paradigm characterized by an active older male and a passive younger male, which is contrary sharply to the androcentrism and the cross-sex normativity discussed in Chapter 1. This analysis provides a clearer framework for understanding how marginalized sexual concepts are contextualized against dominant norms. Historic dictionaries not only documented etymologies, male sexual passivity, and dyadic recursivity associated with *pederasty*, but also preserved related classical myths. From the sixteenth to the nineteenth centuries, lexicographers reinterpreted the myths of male stories, highlighting the boundary between fiction and reality to express reverence for ancient Greece and Rome. This strategy allowed them to retain the erotic aspects of classical mythology while keeping a cautious distance from the cultural mainstream of English lexicography. While myths of male love could find a place on the cultural periphery of hard-word and general dictionaries, myths surrounding female affection remained entirely excluded, reflecting the boundaries of acceptability in the historical linguistic tradition. The term *lacunae* most aptly describes this absence of female-female love myths. Finally, this chapter examines same-sex love in classical and bibliographical dictionaries, with a special focus on Sappho, who was renowned for her desire for women. In the early modern period, Sappho was depicted primarily in three ways: as the foremost female poet, the jilted lover of

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1983

Phaon the ferryman, and the admirer of her own sex. While largely absent from hard-word and general dictionaries, Sappho found a place in classical references during the eighteenth century. Diverse representations of Sappho across different genres highlight the inherent tensions in cultural representation. Dictionaries, functioning as linguistic repositories, not only document language but also unveil the marginalized and complex narratives of same-sex love.

Chapter 3 begins by exploring the translation of the term *Tribade*. Historically, lexicographers addressed sexual terminology in two primary ways: by borrowing definitions from bilingual dictionaries or by creating new terms. Headwords related to female sexuality, such as *fricatrix* and *tribade*, derived from Latin and Greek, respectively, and were often considered unnatural, impolite and improper. While these terms appeared in some Latin–English word lists, they were conspicuously absent from certain hard-word dictionaries, likely due to the inclusion of ladies and gentlewomen among their target audience. Both lexicographers and users play a crucial role in shaping dictionary content. By examining the historical lexicographic treatment of terms and definitions associated with erotic relations between women, we can better understand how sexual linguistics was shaped by the intended audience and propriety constraints. Turton subsequently shifts focus to the representation of lesbian relationships in general dictionaries. Beyond *fricatrix* and *tribade,* the word *Sapphic* also signifies same-sex intimacy between women. Historically rooted in ancient Greek narratives, the concept of *Sapphic* love acquired negative connotations within the Catholic theological framework. This perspective subsequently influenced its reception in Victorian Britain, where moral objections significantly impeded both its social acceptance and linguistic integration into mainstream discourse. Consequently, the term *Sapphic* turned into a "ghost-word", occurring in classical writing but lacking entries in dictionaries. Similarly, many obscene or non-procreative sexual terms were excluded due to prudent and pious concerns. Turton's survey of same-sex terminologies in dictionaries revealed a sharp decline in their representation between 1755 and 1830, followed by a revival from 1831 to 1883, reflecting the complex interplay of reader's comprehension, moral consideration, and linguistic documentation. Despite the reviving coverage of *pederasty*, *buggery* and *sodomy* in general dictionaries in the nineteenth century, explicit terms for sexuality between women remained absent. The latter half of Chapter 3 contrasts dictionaries of proper English with lexicons of cant and fashionable slang to trace the limits of the polite and impolite expressions. Notably, terms about same-sex love between women remained overlooked by cant and slang lexicographers. By tracing historical dictionaries, we can analyze how sexual vocabularies entered or exited these lexicons from the sixteenth to the nineteenth centuries. The omission of certain words from English dictionaries serves as a means of reinforcing the exclusion of certain groups from English society. Through this analysis, historic dictionaries offer deeper insights into the evolution of sexual language shaped by societal scruples and cultural discourses of deflection.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1983

Chapter 4 focuses on sexual language in medical dictionaries, beginning with an analysis of the pathologization of unnatural sexual practices. Medical dictionaries not only documented diseases associated with such behaviors, but also served as frameworks for public health management, behavioural regulation, and bodily control. The discussion then shifts to entries on the clitoris and the tribade's phallus, noting that an enlarged clitoris was often linked to immoral practices and female same-sex desire. Medical dictionaries frequently recommended surgical intervention for this condition, a practice more prevalent in Arab and Egyptian contexts than in Europe. In medical dictionaries, terms like *Confricatrices* were employed to describe learned or imitative female same-sex interactions. Driven by the revival of classical scholarship and advancements in forensic medicine, the topic of female same-sex relationships gained particular attention. This chapter concludes by emphasizing the material constraints of medical lexicons, arguing that the scope of sexual knowledge produced by lexicographers was deeply influenced by material limitations of the dictionaries themselves. Terms related to sex bears distinct interpretations in different types of dictionaries. By examining medical dictionaries, we can gain insights into the representations of sexual intercourse, particularly female homosexuality.

Chapter 5 examines the intersection of science and sexuality within the *Oxford English Dictionary* (OED). The OED employed a scientific methodology in its compilation, characterized by empirical principles and extensive volunteer involvement. The OED's treatment of sexual vocabularies aligned with the moral, legal, and natural standards that were prevalent from the sixteenth to the eighteenth centuries. Old discourse could be found in OED as evidence. While the OED surpassed other general dictionaries in documenting same-sex terms, its scientific rigour was compromised by incomplete linguistic coverage, inadequate definitions, and insufficient or unsatisfactory quotations as supporting evidences. These limitations partly stemmed from the diverse backgrounds and biases of its contributors, as well as the potential expurgation by lexicographers, despite volunteers providing direct and faithful reports about sexual lexis. Then, Turton explores the evolution of sexology and its representation in the OED1 (1888–1928) and its supplement (1933). Sexology redefined sexual desire as a psychological phenomenon rather than a matter of sinfulness, criminality, or anatomy, marking a significant shift from legal to medical, and ultimately to psychological interpretations. This evolving perspective influenced Western articulations of sexuality. Notably, OED1 omitted sexology. But the continuing advance of sexology resulted in vast nomenclatures, which posed challenges for lexicographers in incorporating the expanding terminology of sexual terms. Space and time constraints during OED1's compilation precluded adequate coverage of sexology, making it a crucial subject for the OED Supplement in 1933. The OED Supplement typically included citations from terms related to homosexuality and sexual inversion to same-sex relationships, with most citations drawn from medical sources. At the end of this chapter, the author turns to the Sapphic and Uranian writers in the supplement. This chapter examines the OED's editorial methods

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1983

and the representations of sexology in both OED1 and its Supplement. It underscores that scientific methodologies are invariably influenced by the prevailing social mores of their respective contexts. While the OED serves as an authoritative instrument for documenting language, it exhibits inherent limitations and biases, particularly in its treatment of language related to homosexuality and female same-sex relationships. Despite these shortcomings, the essential role of dictionaries in elucidating the trajectory of sexual history should not be ignored. After all, through the analysis of the treatment of sexology in both OED1 and its Supplement, this chapter deepens our understanding of the development of sexual linguistics in lexicography.

The conclusion of this volume extends beyond queer perspectives and the present, examining the OED's lasting impact on lexicography. Its descriptivist principles have become foundational, shaping modern dictionary compilation. Besides, compilation process has evolved from individual editors to broader public participation, reducing personal biases while introducing collective ones. Furthermore, technological advancements, particularly the use of databases, have revolutionized language analysis, contributing to more objective semantic and grammatical documentation in both print and electronic formats. Additionally, social movements like the 1969 Stonewall Riots, alongside with legal reforms and shifting attitude, have promoted a reevaluation of language surrounding sexual identities. In modern times, online dictionaries tend to adopt more inclusive language regarding gender and sexuality. However, true linguistic democracy still remains elusive. Because descriptivist principles cannot fully eliminate social inequalities, volunteer compilers bring their own biases, and databases, while valuable, often amplify certain perspectives at the expense of others. Although OED3 (2000–) has made progress in refining sexual terminology, such as redefining *homosexuality* and *uranian*, some definitions of sex-related vocabulary still remain consistent with those from the 18th century, perpetuating the same dilemmas.

In summary, gender-related topics such as sodomy, buggery, sapphic relationships, and pederasty still remain sensitive and underexplored even today. Turton's *Before the Word was Queer* bridges dictionary compilation, queer theory, literary criticism, and linguistics, addressing a critical gap in the study of sexual language. This work demonstrates that dictionaries should not be viewed solely as scholarly tools but as texts requiring critical analysis. Linguistics analysis through dictionaries not only unlocked the evolution of queer history and the queerness within sexual narratives but also revealed how lexicographers and their audiences interpreted, neglected, or accepted human sexuality from the 16th to the 19th centuries. Today, lexicographers continue to grapple with challenges in addressing gender and sex-related entries. Turton's work underscores that historical dictionaries serve not only as keys to unlocking sexual history but also as frameworks for articulating historical contexts and refining definitions of sexual terminology in the future.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1983

References

Murray, J.A.H., H. Bradley, W.A. Craigie and C.T. Onions (Eds.). 1888–1928. *A New English Dictionary on Historical Principles; Founded Mainly on the Materials Collected by the Philological Society (Vols. I–X).* Oxford: Clarendon Press. (OED1)

Murray, J.A.H., H. Bradley, W.A. Craigie and C.T. Onions (Eds.). 1933. *The Oxford English Dictionary: Being a Corrected Re-issue with an Introduction, Supplement, and Bibliography of A New English Dictionary on Historical Principles (Vols. I–XII, Supplement).* Oxford: Clarendon Press. (OED Supplement)

Simpson, J.A. and M. Proffitt (Eds.). 2000–. *OED Online.* Third edition. (OED3) Retrieved from www.oed.com

Xiaotong Zhao
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou
China
(20242510009@mail.gdufs.edu.cn; 1657555795@qq.com)
(https://orcid.org/0009-0008-2301-1067)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

Analyse des Termes Émergents en Épidémiologie Post-COVID-19

Ma'moun Alshtaiwi, *Département des langues modernes, Université de Yarmouk, Jordanie (m.alshtaiwi@yu.edu.jo) (https://orcid.org/0000-0003-4458-7694)*

Résumé: Après la crise de la COVID-19, de nombreuses ressources terminologiques ont actualisé les définitions des termes liés au domaine de l'épidémiologie. Des unités terminologiques simples déjà existantes telles que «pandémie», «épidémie» et «endémie» ont été particulièrement examinées; leurs définitions ont été révisées pour refléter les leçons tirées de la pandémie de COVID-19 et son impact sur la compréhension de ces concepts épidémiologiques. Cette adaptation est le résultat de la modification du système conceptuel du domaine. De plus, il a été nécessaire d'attribuer de nouveaux sens aux lexèmes déjà existants afin d'engager la société dans l'avancement des situations de crise sanitaire. Ces unités se sont imposées en réponse aux situations d'urgence, surtout dans une période où le monde entier suit massivement les informations liées à l'épidémiologie. Dans ce travail, nous analysons la terminologie émergente ainsi que le système conceptuel du domaine de l'épidémiologie. Les résultats ont révélé que cette terminologie s'est développée à travers différents domaines de spécialité tels que la médecine, la psychologie, l'informatique, ainsi que le droit, la sécurité et les maladies anciennes.

Mots-clés: ÉPIDÉMIOLOGIE, TERMINOLOGIE, CONCEPT, SYSTÈME TERMINOLOGIQUE, POSTE COVID-19

Abstract: Post-COVID-19 Review of Epidemiological Terms and Their Multidisciplinary Implications. Several dictionaries and terminological databases have updated numerous definitions of terms related to the field of epidemiology following the COVID-19 crisis, for instance, definitions of terms like "pandemic," "epidemic," and "endemic." This adaptation is the result of a modification in the conceptual framework of the studied domain. Furthermore, it was necessary to coin lexemes expressing new concepts to engage society in addressing public health crises. These terms have become commonplace in response to emergencies, especially when the entire world closely follows epidemiology-related information. In this work, we analyze the existing terminology in epidemiology. The comprehensive results have revealed that epidemiological terminology has evolved across various specialized fields such as medicine, psychology, informatics, as well as law, security, and ancient diseases.

Keywords: EPIDEMIOLOGY, TERMINOLOGY, CONCEPT, TERMINOLOGICAL SYSTEM, POST COVID-19

Introduction

La pandémie de COVID-19 a donné lieu à plusieurs formulations et dénomina-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

tions de concepts liés aux pratiques sanitaires, aux mesures préventives ou aux actions. Ainsi, le *Grand dictionnaire terminologique* (OQLF 2010) a répertorié 68 termes relatifs au concept d'épidémiologie en 2020, dont les termes « syndrome post-COVID-19 », « pistage numérique » ou « défilement morbide ». Parmi ces termes, quelques-uns ont été attribués par l'OMS, comme l'acronyme « COVID-19 » et sa forme étendue « maladie à coronavirus 2019 ». D'autres termes relèvent de différents domaines, récents ou anciens, et ont été actualisés par de nouvelles significations. Bien que ces termes existent depuis longtemps, l'avènement de la pandémie les a popularisés auprès du grand public (quarantaine, couvre-feu, …). Ces termes décrivent des maladies du passé et sont utilisés pour décrire la situation de la pandémie. De plus, des termes ont été empruntés à d'autres langues et souvent des anglicismes ont été ajoutés au français. Certains linguistes voient d'un mauvais œil ces emprunts, par exemple le terme « coronavirus », qu'ils considèrent comme un emprunt à l'anglais qui place l'élément spécifique avant l'élément général. De même, le terme « distanciation sociale » a été critiqué pour être calqué sur l'anglais et certains veulent s'en éloigner. Cependant, pour autant que nous le sachions, ce terme est utilisé dans la plupart des langues du monde. Dans ce cas, le calque sémantique de ce terme est avantageux dans la mesure où l'aspect sémantique universaliste du terme est modélisé dans plusieurs langues. D'autres termes n'ont pas nécessairement été calqué à l'anglais, et cette diversité se reflète clairement dans la réaction des locuteurs à la stratégie d'auto-isolement que les individus du monde entier ont commencé à exercer. Dans cette étude, nous examinons les concepts émergents en épidémiologie, impliquant la mise à jour de nouvelles définitions et de nouveaux termes nouvellement apparus dans l'ère post-COVID-19.

1. Cadre théorique

Les différents principes et orientations partageant l'idée que la terminologie est construite à partir de deux composantes: composante conceptuelle et composante linguistique, elle repose sur deux propriétés: (a) le terme désigne un concept ; (b) le concept est attaché à un domaine spécialisé. Pour bien saisir les structures et les usages, y compris l'équivalence, le développement et l'efficacité des différentes ressources terminologiques utilisées en épidémiologie, il est nécessaire de définir les deux hypothèses sur lesquelles repose la théorie de la terminologie. La première approche repose sur la théorie onomasiologique de la formation des termes donnant des noms à partir du sens ou du concept (Kocourek 1991; Wüster 1976, 1979). Dans ce cas, on recherche les différentes désignations ou termes correspondant à un sens spécifique et on les qualifie de synonymes, par exemple, les deux termes « agent infectieux » et « agent pathogène ». La seconde approche est souvent utilisée en lexicographie, elle est basée sur le rassemblement de différents termes dont il faut définir les significations en sémantique (approche sémasiologique). Ainsi, un même terme peut avoir plusieurs sens:

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

on parle alors d'homonymie ou de polysémie. La désignation peut être faite par une unité terminologique, appellation d'un concept individuel ou par un symbole. Dans notre corpus, nous recherchons donc des énoncés descriptifs ressemblant à des définitions, car l'ensemble de ces énoncés est la représentation d'un concept servant à le différencier des concepts connexes. Ces énoncés peuvent aussi être utilisés afin de formuler des définitions par compréhension ou par extension, de décrire les caractères essentiels ou non essentiels (tel que la forme, la couleur et la dimension …) à l'existence du concept et d'enrichir des ressources lexicales et ontologiques existantes (Dahlberg 1981). Ainsi, nous devrions combiner ces deux approches précédentes, car dans tout travail terminologique, il est indispensable d'analyser le système conceptuel hiérarchique afin de pouvoir distribuer les désignations terminologiques existantes liées au domaine (Alshtaiwi 2016). De plus, étant donné que nous travaillons avec un corpus pour dépouiller la terminologie existante, ce qui signifie que notre recherche est textuelle, nous utilisons donc une approche sémasiologique, notamment lors de l'utilisation d'outils électroniques pour la recherche d'un corpus textuel. Dans l'analyse de l'unité terminologique, nous pouvons observer que la différenciation se situe principalement au niveau des approches adoptées. L'analyse des termes se base souvent sur un corpus où les termes sont pris en compte.

2. Évolution de la terminologie en épidémiologie

a. Intégration de concepts de différents domaines

L'épidémiologie est un domaine intégré à la médecine préventive et à la santé publique. Les nouveaux termes qui ont fait leur apparition en épidémiologie proviennent de divers domaines, notamment la médecine, la sociologie, l'informatique par exemple, « apéritif virtuel », la biologie, la psychologie, et d'autres domaines d'activité. Il est important de noter que cette liste n'est pas exhaustive et puise dans de multiples systèmes de connaissance. En réaction à la crise sanitaire, des mises à jour ont été apportées à la classification des domaines concernés. En 2020, le *Grand dictionnaire terminologique* (GDT) a actualisé la définition du terme « épidémiologie » comme suit: « L'étude de la fréquence et de la répartition des problèmes de santé et des maladies dans le temps et dans l'espace, ainsi que des facteurs qui conditionnent leur apparition, leur diffusion et leur évolution au sein des groupes ou des populations » (OQLF 2010). Par conséquent, le terme « étude des épidémies » ne couvre désormais qu'un sous-domaine de l'épidémiologie moderne (OQLF 2010). Les termes empruntés à divers domaines ne sont en réalité qu'une appropriation du vocabulaire utilisé dans différentes spécialités. Dans cette perspective, nous pouvons observer comment plusieurs domaines connexes contribuent à l'enrichissement conceptuel du domaine de l'épidémiologie. À travers l'identification des termes spécifiques et leur fréquence dans le corpus, nous avons pu dégager sept domaines qui jouent un rôle crucial dans cette évolution conceptuelle à divers niveaux.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

#	Domaine	Représentation de la terminologie dans le corpus	Exemple
1	Biologie et médecine	45 %	Variante delta, variante gamma, variante omicron … délestage
2	Pharmacologie	18 %	Remdésivir
3	Informatique	17 %	Pistage numérique
4	Psychologie et sociologie	10 %	traumatisme de réclusion
5	Sécurité	5 %	Couvre-feu
6	Commerce	3 %	surenchère immobilière
7	Loisir	2 %	Apéritif virtuel

Tableau 1: Représentation conceptuelle des domaines connexes de l'épidémiologie

Ces concepts, provenant de champs variés, ont émergé de manière particulièrement notable au cours de la pandémie, et nous les classons en fonction de leur prévalence, du plus fréquent au moins fréquent. Cette révision approfondit davantage l'analyse en expliquant de manière plus cohérente l'origine des nouveaux termes en épidémiologie et en détaillant comment ces termes ont été adoptés dans divers domaines. En outre, elle précise de manière explicite le processus de mise à jour de la terminologie et la classification des concepts issus de domaines connexes. Notons que ces concepts englobent d'autres domaines, par exemple le terme « apéritif virtuel » peut être classé dans trois domaines: en informatique, en sociologie et en loisirs. En termes de hiérarchie, cette classification peut porter sur des domaines principaux ou sur des sous-domaines. Par exemple, le terme « délestage » renvoie à la gestion des soins de santé, un sous-domaine de la médecine.

b. Analyse des formes de regroupement lexical

Les concepts peuvent être classifiés de manière générique, notamment dans le cas de la « prévention» qui peut être « internationale », « nationale », ou « locale ». Ils peuvent aussi être structurés de manière partielle, comme la « technologie médicale » appartenant à une « ressource », ou « par des liens épidémiologiques » et «en laboratoire» étant des types des « cas confirmés » sur la base de critères spécifiques. De plus, au sein d'une même unité lexicale, les concepts peuvent être divisés en plusieurs éléments lexicaux distincts, chacun représentant un concept spécifique. Par exemple, le terme « confinement à durée déterminée » fait référence à une période de confinement relativement courte, contrairement au concept de « confinement », qui est plus commun. En conséquence, il est

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

possible de composer des concepts simples pour former des représentations conceptuelles plus complexes. Un autre exemple de cette composition conceptuelle est la « transmission par contact direct », où le concept procédural général de « propagation d'une maladie » est associé à une topographie caractérisant le concept via un lien sémantique de type « Has_Site », afin de créer un concept composite significatif tel que « par contact ». De plus, un autre lien, tel que « has_Direction », peut être utilisé pour préciser le mode de transmission, comme « direct ». En outre, il est envisageable de subdiviser les concepts en différentes catégories, comme le concept de « cas » qui peut être classé en sous-catégories telles que « cas confirmé », « cas confirmé en laboratoire », « cas suspect », etc. Cette approche de la classification et de la structuration des concepts permet mieux comprendre comment la langue évolue pour traiter des questions spécifiques, notamment dans le domaine étudié.

Concept général	Concept spécifique	Concept spécifique
Cas	confirmé index primaire rétabli suspect	en laboratoire par lien épidémiologique
Dépistage	à la frontière à l'arrivée à la sortie	
Calendrier	de vaccination	de masse

Tableau 2: Représentation compositionnelle pour « cas », « dépistage » et « calendrier »

Nous avons identifié les concepts généraux tel que «cas», «dépistage» et «calendrier». Ensuite, nous avons identifié tous les concepts spécifiques attaché aux concepts généreux, tel que «cas confirmé», «cas primaire» et «calendrier de vaccination» qui leur sont associés. Nous avons créé une représentation conceptuelle et en utilisant des symboles exprimant les relations entre les termes.

L'intégration de nouveaux termes dans le langage courant depuis le début de la pandémie de COVID-19 a permis aux individus d'exprimer et de comprendre les situations d'urgence. On constate que certains termes médicaux ont intégré le langage courant, car ils sont devenus indispensables pour décrire des concepts de plus en plus répandus. Nous avons entrepris une analyse typologique des concepts, examiné leurs caractéristiques et exploré les relations qui les unissent. Cette démarche nous a permis d'identifier cinq catégories conceptuelles distinctes liées à l'épidémiologie.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

Composante	Exemple
Composante humaine	« cas confirmé »
Composante micro-organique	« maladie infectieuse », « virus », « acide désoxyribonucléique », « aérosol infectieux », « agent immunisant »
Phénomène physique naturel	Phénomènes de mouvement relatif à l'air et les virus dans l'espace.
Mathématiques	Mesure logique et logiciel, distance sociale.
Composantes logicielles	Procédures nécessaires au fonctionnement d'un système informatique. Exemple « application de recherche de contact », « assistant virtuel »
Composantes matérielles	« banque des vaccins », « désinfectant pour les mains »
Processus, action et mesure	« traitement de données », « technologie » et « transmission », « auto-isolement », « couvre-feu », « dépistage »

Tableau 3: Nature des concepts relatifs à l'épidémiologie

Un système terminologique est une structure qui organise les liens entre les unités terminologiques et les concepts d'un domaine spécifique, tout en établissant des règles, des relations, des définitions et des codes associés. Ces systèmes portent différents noms tels que terminologie, nomenclature, thésaurus, vocabulaire, classification, système de codage, taxonomie et ontologie. Les concepts sont interconnectés dans divers domaines. Les termes examinés présentent des difficultés particulières dans la recherche d'équivalences entre les langues.

3. Développement post-COVID de la terminologie de l'épidémiologie

a. Terminologie relative à la médecine et à la psychologie

Les termes « morbide et comorbidité » et « coping » sont souvent utilisés dans le cadre de la priorisation des vaccins. Cette terminologie a été utilisée très fréquemment après l'épidémie de COVID-19. Pour les personnes n'appartenant pas au domaine de la santé, les deux termes « morbide », « morbidité » et « comorbidité » sont probablement associés au concept « malsain », mais dans le domaine médical (sous-domaines: sémiologie et pathologie), le concept exprime les causes « relatif à la maladie » (*Usito,* dictionnaire). On parle de « comorbidité » d'une maladie primaire avec de multiples conditions cliniques spécifiques ayant un caractère « primaire ». Cette notion diffère de la multimorbidité, qui n'a pas de maladie primaire après l'ajout de conditions cliniques. Afin de pouvoir comprendre le sens de ce terme, nous pouvons aussi identifier un certain nombre de caractéristiques à travers les occurrences identifiées dans le contexte. La « comorbidité » peut être mesurée, elle est fréquente dans un certain nombre de maladies et elle peut être classée et mesurée. Des occurrences nominales et verbales ont été associées au terme « comorbidité »: « Mesurer X », « La présence de X »,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

« X fréquents », « Facteur de X », « X augmente », « X identifié », « X chez », « Connaître X », « Classer X ». Ainsi, ce terme est moins familier dans le langage courant. En principe, les non-spécialistes des domaines médicaux auront du mal à comprendre à quoi ces termes font référence. La « comorbidité » renvoie donc à une « présence simultanée chez une personne d'un problème de santé donné et d'un ou plusieurs troubles, maladies, facteurs de risques ou processus pathologiques » (Bureau de la traduction — Services publics et Approvisionnement Canada 1976). Les comorbidités impliquent la présence d'une ou plusieurs maladies associées à la maladie dite principale. À titre d'exemple, on discuterait en psychiatrie de la possibilité d'une « association comorbide entre anxiété, dépression, addiction, etc., » puis le rôle diagnostique du thérapeute serait de définir l'émotion primaire la plus importante du patient afin d'adapter son schéma thérapeutique. En français, l'OQLF sur le Grand dictionnaire terminologique définit le terme « coping » comme suit: « Ensemble des efforts cognitifs, émotionnels ou comportementaux qu'une personne met en œuvre afin de maîtriser ou de tolérer les tensions internes ou externes qui menacent ou dépassent ses ressources ou ses capacités à s'ajuster à une situation » (OQLF 2010). En anglais, Le GDT (OQLF 2010) définit le terme en question « coping » comme suit: « stratégies mises en œuvre par le sujet pour s'adapter à une situation stressante ». Les termes privilégiés sont « adaptation » et « ajustement » en psychologie (ibid.). Dans le contexte de COVID-19, nous pouvons observer un autre terme pour le même concept, le « faire-face ». Ainsi, la base de données terminologique *France Terme* a proposé le terme « faire-face » en 2016 à une situation pour l'équivalent du terme anglais « coping » (France Terme, Ministère de la Culture, 2000). Le terme français « coping » est ainsi un anglicisme courant en français, particulièrement en la psychologie de la santé. Il désigne les stratégies comportementales et les ressources d'adaptation qu'on utilise lorsqu'on est confronté à une situation stressante. Selon le *GDT,* cet emprunt intégral […] manque d'adaptation sur tous les plans: graphique, phonologique, morphosyntaxique et sémantique (OQLF 2010). Ainsi, l'équivalent proposé par le *GDT* est le terme « adaptation ». Un certain nombre d'occurrences nominales et verbales associées au terme « coping » notons les suivantes: « X centré sur … », « X comporte … », « X positif », « X dépend », « stratégies de X », « styles de X », « X et le stress », « X de l'individu ». Ainsi, il est important de distinguer entre « adaptation » et « coping ». Le terme « adaptation » renvoie à un concept général, par ailleurs, le coping renvoie à un concept plus spécifique. Le dernier terme exprime différentes stratégies d'adaptation et de réaction dans un environnement stressant. Il s'agit d'un processus spécifique que l'individu pense et fasse face à une telle situation.

b. Terminologie relative à l'infectiologie

L'utilisation d'un terme peut être trompeuse, comme la « grippe espagnole » qui, en fait, était loin d'être espagnole. C'est ainsi que le coronavirus a été appelé à l'origine lorsqu'il a fait son apparition en Chine. De nombreux noms ont été

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

déjà attribuées à ce virus, par exemple l'utilisation de l'expression « virus chinois » ou la variante « britannique » ou « brésilienne », etc., mais tous ces noms ne figurent jamais ni dans le dictionnaire ni dans les bases de données lexicales et terminologiques, parce qu'ils sont peu utilisés ou pour d'autres raisons ségrégationnistes.

L'OMS joue un rôle indispensable dans la normalisation de la terminologie existante. Le terme COVID-19 a été officiellement introduit par l'OMS, qui est une abréviation de (coronavirus disease), traduite par « maladie à coronavirus 2019 ». De nombreux noms de maladies ont été simplifiés. Par exemple: de nombreux médias ont utilisé des termes géographiques pour décrire la variante du coronavirus.

Nom lié à un lieu géographique	Nom scientifique	Code scientifique
Variante britannique	Variante Alpha	B.1.1.7
Variante sud-africaine	Variante Bêta	B.1.351
Variante brésilienne	Variante Gamma	P.1
Variante indienne	Variante Delta	B.1.617. 2
Variante botswanaise	Variante Omicron	B.1.1.529

Tableau 4: Les variantes épidémiologiques

Comme indiqué ci-dessus, la terminologie de l'épidémiologie se caractérise par des noms liés à une localisation géographique. L'OMS a désigné ces noms par des noms et des codes scientifiques.

c. Terminologie relative aux sciences de l'information et de la communication

Près de 2,6 milliards de personnes ont été confinées dans le monde entier et ont dû recourir principalement à l'Internet pour communiquer, travailler, voire se détendre ou faire du sport. Dans ces conditions, il n'est pas surprenant que ces personnes issues de cultures et de milieux différents interagissent les unes avec les autres et aient besoin d'une terminologie unifiée commune, ou du moins d'un ensemble d'expressions communes pour communiquer. Lorsque des restrictions sanitaires sont déclarées, certains États utilisent le terme anglais « tracking » pour désigner le « traçage » et la « reconstitution du parcours d'une personne ». Le terme « tracking » est utilisé dans des domaines divers (en aéronautique, en armée, en management, en transport, etc.) et exprime principalement le concept de « suivre des objets et non des personnes ». Dans ce contexte, il s'agit de tracer « des individus par leur téléphone, via un dispositif de géolocalisation. » Ainsi, le terme « géolocalisation » pourrait exprimer un moyen d'identifier ces individus, en se référant à une technique utilisée pour décrire l'identification d'une

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

personne par sa position géographique sur son téléphone portable. En termes épidémiologiques, le terme « *tracking* » fait référence à l'identification des personnes qui ont été en contact étroit avec une ou plusieurs autres personnes infectées, et à la manière d'identifier une personne qui a été en contact avec un ou plusieurs porteurs du virus en question. On parle de « *tracking* » lorsqu'il s'agit d'objets, mais dans le contexte de la pandémie, il désigne les personnes infectées. Il fait référence à la localisation géographique des porteurs, que ce soit par le biais d'un téléphone mobile. Pour déterminer l'emplacement géographique d'une personne, on peut recourir à la technique de « géolocalisation ». Si l'on veut connaître la localisation du mobile qu'il possède, on utilise le « *tracking* ». Le terme sera utilisé pour la police et pour le déplacement d'un objet. Mais dans les débats publics, qu'ils n'utilisent pas le terme « *tracking* ». Il est étrange que le terme ne soit anglicisé que lorsque le concept est pertinent pour tout le monde. En ce qui concerne le terme « *back tracking* », nous avons trouvé deux équivalents, les termes « reconstitution de parcours » ou aussi « reconstitution de trajectoire », « reconstitution de parcours » d'une personne. Le terme « *back tracking* » a également d'autres termes possibles en français pour le décrire: « reconstitution de son parcours » de personnes ayant eu le virus.

Ces termes ne sont pas toujours clairs, car ils sont pour la plupart empruntés ou copiés de l'anglais. Par exemple, le terme « *super spreader* » a été traduit par « *super contaminateur* » ou « *super propagateur* » pour décrire une maladie qui infecte un grand nombre d'autres personnes. Le terme anglais « *cloud rave* » exprime un concept numérique de plus en plus répandu. Ce concept renvoie à l'idée d'une diffusion en direct pour mettre en relation des artistes et des publics du monde entier afin que les gens se divertissent tout en restant isolés. Le concept fait référence aux « rituels » accompagnant le confinement et qui sont nés du manque d'interaction humaine. Ces rituels sont notamment les chants sur le balcon en Italie et les salutations qui donnent aux gens un sentiment d'identité et d'appartenance à un groupe. Enfin, si certains termes peu ou pas d'ambiguïté en anglais, ce n'est pas le cas en français. Ce problème d'équivalence des termes est une source de confusion pour les francophones, et il n'est pas spécifique. Il est donc important d'avoir des équivalences claires afin que les locuteurs puissent communiquer sur la gravité de la maladie et soient en mesure de comprendre les médias.

d. Terminologie relative à l'environnement social

Des termes liés à l'espace et à l'environnement social sont apparus depuis les premières manifestations du COVID-19, tels que « décontamination », « quarantaine », « distanciation sociale ou physique ». Ces termes existaient avant le début de la pandémie, mais ils ont aujourd'hui des définitions différentes et sont utilisés par la population pour décrire la nouvelle réalité. On observe ainsi un glissement sémantique d'un terme réservé aux hôpitaux ou aux voyageurs vers un terme commun. Dans notre contexte, selon la *base de données terminologiques*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

canadienne TERMIUM Plus®, le concept d'isolement fait référence à une « mesure préventive contre la propagation d'une maladie infectieuse, consistant à séparer la personne infectée des personnes non infectées pendant la période de transmissibilité de la maladie » (Bureau de la traduction — Services publics et Approvisionnement Canada 1976). En outre, selon le *Dictionnaire COVID: Le Lexique de La Pandémie de A–Z*, 2021 note que cette mesure peut être aussi un « auto-isolement » au cours duquel « une personne atteinte décide de se séparer des personnes non infectées pendant la période de transmissibilité de la maladie ». Examinons les trois termes suivants: « distance sociale », « confinement » et « quarantaine », ainsi que leurs variantes.

i. Distanciation sociale, distanciation physique, éloignement social et éloignement physique

Ces termes sont principalement liés à la distance sanitaire. Selon le *GDT*, il s'agit d'une distance recommandée « par les autorités sanitaires » et selon *TERMIUM Plus*® « en maintenant une distance entre les personnes » pour éviter « la propagation d'une maladie infectieuse caractérisée par la non fréquentation des lieux surpeuplés, des lieux de rassemblement et des contacts étroits » (Bureau de la traduction — Services publics et Approvisionnement Canada 1976). Le concept exprimé par le terme « distanciation physique » peut être exprimé par plusieurs désignations. Les deux premiers termes « distanciation sociale » et « distanciation physique » sont inspirés des termes anglais « *social distancing* » et « *physical distancing* ». Quant aux deux autres termes, « éloignement social », « éloignement physique » ou « éloignement sanitaire », ils sont controversés. Ces cinq termes sont désormais communément utilisés pour exprimer une nouvelle contrainte à la protection contre le virus. Certains observateurs critiquent cet usage, soulignant qu'il serait plus judicieux de dire « physique » que « social », alors que l'élément « social » peut impliquer un isolement social, voire un fossé entre classes sociales. Certaines sources d'information utilisaient le terme de « physique », plutôt que « social », car « il a l'avantage » de réduire « la connotation négative associée à l'isolement social » (OQLF 2010). Par ailleurs, « l'utilisation de l'adjectif «social» est motivée lorsqu'il fait référence à la «*distance sanitaire*» devant être maintenue en société » (Bureau de la traduction — Services publics et Approvisionnement Canada 1976). Grâce aux technologies contemporaines, « les personnes éloignées physiquement ne renoncent pas la possibilité » de maintenir « leurs liens sociaux » (OQLF 2010). Bien que le terme « distanciation » renvoie généralement à « une forme imagée de recul par rapport à une chose ou à un événement, il est également acceptable, par extension, pour désigner un éloignement physique entre deux choses ou deux personnes » (OQLF 2010). Ainsi, depuis l'apparition de la pandémie de COVID-19, le sens s'est élargi et englobe désormais la distance sanitaire.

Le terme de « distanciation sociale », dans cette acception, est « utilisé par les autorités sanitaires […] depuis le milieu des années 2000 » (OQLF 2010). Ce terme

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

nous semble approprié, car il exprime l'établissement d'une distance sécuritaire. Notons par ailleurs que le terme « distanciation sociale » fait sans doute écho à la notion plus ancienne de «facture sociale» qui l'a précédée.

Concernant le terme « éloignement », il renvoie à « la distance, dans l'espace, de deux objets ou d'un objet au point d'observation » (International Federation of Surveyors. Technical Dictionary Commission 1963). Contrairement à la distanciation, l'éloignement social, telle que décrite par l'OMS, n'explique pas l'idée d'être loin d'un objet « éloignement de la mer ». Le terme distanciation implique une mesure préventive en accordant une distance (de 2 mètres par exemple) entre deux personnes.

Il convient de noter que le concept en question est devenu synonyme. Le terme fait référence à une mesure préventive et volontaire établie par les autorités et utilisées uniquement pour des raisons de santés et par les personnes en bonne santé. Enfin, la notion de « distanciation » renvoie aussi à différents exemples, dont les salutations et les différents types de salutation sans contact telles que « agiter la main, « placer la main sur son cœur », « *namasté* », « lever les pouces ou les sourcils », « hocher la tête ou le menton », et les salutations avec contact comme « par le coude, poignée de main », etc.

Cependant, des désaccords surviennent parfois entre les utilisateurs quant à la nécessité d'utiliser terme. Par exemple, certaines personnes sont critiques à l'égard du terme « distanciation sociale » en raison de son origine anglaise. Ce terme est utilisé depuis de nombreuses années, depuis l'épidémie de grippe H1N1 et la crise du SRAS, par les autorités sanitaires telles que l'OMS, mais certains voudraient le voir banni, car le terme français « éloignement » renvoie au même concept.

ii. Confinement, confiner, restez chez vous !

Selon *TERMIUM Plus*®, la notion de « confinement » fait référence à un « protocole d'urgence visant à limiter les mouvements et les rassemblements d'une population pour des raisons de santé ou de sécurité publique » (Bureau de la traduction — Services publics et Approvisionnement Canada 1976). Comme l'explique le linguiste Jacques Poitou, il s'agit d'un nom verbal masculin dérivé du verbe « confiner », qui signifie « enfermer dans un lieu déterminé ». En outre, c'est un « nom d'action … » qui exprime, «par métonymie, le résultat (*nomen acti*) et la période pendant laquelle dure l'isolement». Dans le contexte de la pandémie, son sens actuel en France est «limité par la spécification de lieu» au domicile. Ce terme s'accompagne de la fermeture de tous les espaces susceptibles de recevoir du public, à l'exception de ceux jugés indispensables comme les magasins d'alimentation et les pharmacies. Dans le langage courant, le terme «confinement» remplace celui de «quarantaine» (ibid.), qui désigne «l'isolement d'un petit groupe d'individus par rapport à tous les autres», alors que l'enfermement désigne «l'isolement» de chaque individu «par rapport à tous les autres» (Poitou 2022)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

Parmi les termes apparus au cours de la pandémie figure le terme «grand confinement». Ce terme était utilisé en milieu hospitalier ou carcéral, mais son utilisation n'est devenue courante que depuis le début de la COVID-19. Par analogie avec la Grande Dépression des années trente du XXe siècle (souvent appelée crise de 1929 ou crise de 29), le Fonds monétaire international (FMI) a baptisé mi-avril la crise économique mondiale résultant des mesures de confinement prises dans différents pays». C'est l'origine du terme anglais «*Great Lockdown*» (Poitou 2022).

Toutefois, le participe passé de confiner, «confiné», est utilisé à la fois comme adjectif et comme nom. On peut citer comme autre exemple une équivalence de l'injonction à se confiner: «Restez chez vous !». Cette expression complexe a été traduite dans quasiment toutes les langues et exprime de manière simple une notion de proximité avec l'individu. 'Mais d'une manière générale, le confinement est utilisé «pour désigner une restriction de la liberté de» (Poitou 2022) mouvement, dont «les périmètres peuvent en être variables. […]». Ainsi, à Wuhan, par exemple, «les premières mesures de confinement mises en place à la fin du mois de janvier portaient sur l'interdiction» de franchir les limites de la ville. «Elles ont été renforcées en mi-février par l'interdiction de sortir de son périmètre de résidence sans autorisation» (Poitou 2022). Enfin, les trois termes «confinement», «le grand confinement», «restez chez vous !» font référence à des mesures préventives, établies par les autorités. Ces mesures consistent principalement en un isolement complet de l'environnement social, réservé exclusivement à des raisons sanitaires et pratiqués par des personnes saines.

iii. «Quarantaine» et « quatorzaine »

Dans son sens premier, le terme « quatorzaine » est destiné à remplacer le terme « quarantaine » dans un sens original. Cependant, nombreux sont ceux qui soutiennent qu'il n'est pas légitime, car il ne figure pas dans le dictionnaire, mais qu'il est réputé être un terme plus précis. La détermination des termes corrects est un processus long et difficile. En effet, le seul moyen de le savoir consiste à le rechercher dans le dictionnaire. Mais, comme il s'agit de termes relativement nouveaux, le dictionnaire n'a pas encore fait son choix. Cette décision fait intervenir de nombreuses personnes (journalistes, chercheurs, auteurs, spécialistes de la langue, etc), ce qui prolonge le processus. La probabilité que ces nouveaux termes apparaissent dans le dictionnaire dépend de leur « pertinence » ou de leur « importance » à long terme. Le terme « quatorzaine », désigne une période de quatorze jours de quarantaine. Entre autres, le terme « quarantaine » joue un rôle clé dans la formation du nom « *quarantini* », lui-même inspiré de la marque d'alcool *Martini*. Si le terme « *quarantini* » est apparu en 2005 dans un épisode de la série télévisée Scrubs, il a gagné en popularité et en intérêt pendant l'épidémie et la pandémie de COVID-19 entre 2020 et 2022. Il convient

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

également de noter que le terme anglais « *locktail* », désigne un cocktail préparé durant le confinement.

e. Terminologie relative à la sécurité

Le terme « couvre-feu » (pl. couvre-feux) est une mesure imposée dans de nombreux pays déclarant l'état d'urgence. Depuis les temps modernes, et depuis la guerre, ce terme est utilisé comme néosémie, à l'image de notre environnement et de notre communauté. Il est probable que les termes utilisés tout au long de cette pandémie changeront à nouveau. En effet, la pandémie pourrait également introduire des expressions pérennes, telles que « Dieu vous bénisse », introduites lors des pestes du Moyen Âge, et le terme « patient zéro », adopté tout au long de l'épidémie de SIDA. Pourtant, le terme « quarantaine » représente notion toute nouvelle, qui remonte au Moyen Âge, à l'époque où les médecins portacovidient des masques pour se couvrir, limitant ainsi l'exposition, et où on envisageait de séparer les malades et de les embarquer sur des bateaux. En raison de l'existence d'une forte communauté religieuse, une période d'isolement de 40 jours, à l'instar des prêtres, a été choisie. Pour la pandémie de COVID-19, la durée d'isolement a été de quatorze jours, comme ce fut le cas pour le typhus et la variole. Pour chaque maladie, la durée de l'isolement pendant la période de quarantaine est différente. Cette notion de quarantaine remonte à l'époque médiévale et c'est en 1821 que les Français ont inventé le terme « cordon sanitaire ». Ils ont posté 30 000 soldats à la frontière espagnole pour tenter d'arrêter la propagation de la fièvre jaune. En raison de la gravité de la COVID-19 en Chine et en Italie, de nombreuses épiceries italiennes et restaurants chinois sont complètement désertés, pour la seule raison que leur nom est associé à une maladie. Certains termes peuvent également être utilisés pour blâmer les autres et induire les gens en erreur, notamment dans le cas de la « fièvre espagnole », qui n'était pas vraiment espagnole. Pendant la Première Guerre mondiale, les Français et les Britanniques ont censuré des informations douteuses telles que des soldats malades pour sauver la face. Dans les années où différentes maladies étaient répandues dans le monde, leurs noms étaient parfois utilisés de manière trompeuse. Par exemple, pour calmer les gens, on utilisait le terme «grippe aviaire» au lieu de «peste aviaire» et de «grippe espagnole», qui n'était pas non plus espagnole. C'est ce qui s'est passé pendant la Première Guerre mondiale, alors que l'Espagne n'était pas impliquée.

Conclusion

De cette analyse sur la terminologie de l'épidémiologie, il en ressort que la crise sanitaire de COVID-19 a introduit de nouveaux concepts. Ces termes renvoient à des notions liées à l'environnement social, à la transmissibilité, au virus lui-même, ainsi qu'à l'hôte. Par conséquent, de nouveaux termes ont été actualisés, tandis que d'autres sont apparus au cours de la pandémie de COVID-19.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

Au niveau de la langue française, deux pôles se distinguent dans l'actualisation terminologique au cours de la crise sanitaire. Le premier pôle est européen et français, tandis que le second est québécois et canadien. Cette distinction se traduit par l'utilisation de différents termes pour décrire des concepts similaires. Par exemple, la gestion des soins dans les hôpitaux est exprimée par deux termes, le « délestage » au Québec et la « déprogrammation » en Europe. Par ailleurs, la « déprogrammation » des procédures médicales côtoie la « responsabilité médicale » et les « principes éthiques et juridiques ». Cette stratégie médicale est mise en œuvre dans des contextes spécifiques tels que les pandémies.

Nous pouvons également observer plusieurs procédés de création terminologique en français. Parmi ces procédés, on trouve l'abréviation et l'acronyme [tels que SARS et SARS-CoV-2], le transfert sémantique à une unité lexicale déjà existante, le calque [par exemple, « confined » qui devient « confiné »], l'emprunt intégral ou partiel, de la composition [comme « bêtacoronavirus », « couvre-feu », « couvre-visage »] et la dérivation [comme « déconfinement »]. Les termes « confinement » et « déconfinement » sont un exemple intéressant, car bien que le premier existât déjà, il a dû changer de sens dans les dictionnaires pour refléter les nouvelles significations attribuées par l'épidémiologie. L'antonyme « déconfinement » ne figurait pas dans les dictionnaires et son inclusion a été nécessaire pour décrire les réalités de la pandémie. Par exemple, *Larousse.fr* a adapté la définition de « confinement », et le Dictionnaire *Usito* de l'Université de Sherbrooke a inclus la définition «[…] action de confiner des personnes à l'intérieur d'un édifice public (école, prison, hôpital, etc.) et de restreindre l'accès à cet édifice.»

Enfin, il est essentiel de reconnaître que l'épidémiologie a pu contribuer à l'apparition de nouveaux termes, et il est important que les utilisateurs de la langue fassent preuve de sélectivité dans le choix des termes, car les concepts véhiculés par les termes peuvent considérablement influencer la perception des locuteurs. L'usage de la langue joue un rôle crucial dans notre compréhension des maladies. Par exemple, depuis le début de la crise sanitaire, on observe sur le moteur de recherche «Google» une fréquence élevée du terme «pénurie» associé aux occurrences de «masques» et de «gels hydroalcooliques». De plus, différents hashtags préventifs ont été largement utilisés sur les réseaux sociaux en français, tels que « #RestezChezVous », ainsi que des équivalents en anglais comme « #safeathome », « #togetherathome », et « #selfquarantine », dans le but d'encourager des milliers de personnes à rester chez elles. Ces phénomènes méritent une analyse approfondie, car ces termes suscitent souvent de l'anxiété chez les individus, étant directement liés à la réalité de la pandémie mondiale. L'ignorance de ces nouveaux termes peut limiter l'accès à des informations cruciales, alimentant ainsi l'anxiété. D'autre part, certains termes suscitent la crainte en raison de leur lien avec une réalité objective.

Références Bibliographiques

Alshtaiwi, M. 2016. *Structuration Ontologique et Etude Sémantico-Syntaxique de la Terminologie des Sciences et Techniques Spatiales*. Thèse de Doctorat. Lyon: Université Lumière Lyon 2.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1987 (Article)

Bureau de la traduction — Services publics et Approvisionnement Canada. 1976. *TERMIUM Plus.* [En ligne.] Disponible sur: https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng (dernière consultation: 15 mars 2025)

Dahlberg, I. 1981. Conceptual Definitions for Interconcept. *Knowledge Organization* 8(1): 16-22.

Dictionnaire COVID: le lexique de la pandémie de A–Z. 2021. [En ligne.] Disponible sur: https://fr.africanews.com/2020/07/03/dictionnaire-covid-le-lexique-de-la-pandemie-de-a-z/ (dernière consultation: 15 mars 2025)

Dictionnaire Usito en ligne. https://usito.usherbrooke.ca/ (dernière consultation: 15 mars 2025)

International Federation of Surveyors. Technical Dictionary Commission. 1963. *Dictionnaire multilingue de la Fédération Internationale des Geometres.* Edition trilingue, francais–allemand–anglais. (Argus (Ed.)).

Kocourek, R. 1991. Textes et termes. *Meta* 36(1): 71-76.

OQLF. 2010. *Le Grand Dictionnaire Terminologique (GDT). Gouvernement du Québec*. Disponible sur: http://www.granddictionnaire.com/ (dernière consultation: 15 mars 2025)

Poitou, J. 2022. *Pandémie de Covid-19 (2019– ???): Les Mots pour le Vivre, les Mots pour le dire — Enquête Lexicale, Contextuelle et Diachronique.* Disponible sur: http://j.poitou.free.fr/pro/html/voc/corona.html (dernière consultation: 15 mars 2025)

Wüster, E. 1976. La théorie générale de la terminologie — un domaine interdisciplinaire impliquant la linguistique, la logique, l'ontologie, l'informatique et les sciences des objets. Dupuis, H. (Éd.). 1976. *Essai de Définition de La Terminologie. Actes Du Colloque International de Terminologie, Québec, Manoir du lac Delage, 5–8 octobre 1975*: 49-57. Québec: Régie de la langue française.

Wüster, E. 1979. *Introduction to the General Theory of Terminology and Terminological Lexicography.* Wien: Springer.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

LEXICC: The Design and Development of an Online Dictionary Writing System

Udiluz del Carmen Monsalve Muñoz, *Instituto Caro y Cuervo, Bogotá, Colombia (udiluz.monsalve@caroycuervo.gov.co) (https://orcid.org/0009-0003-5225-3181)*

Johnatan E. Bonilla, *Instituto Caro y Cuervo, Colombia; and Humboldt-Universität zu Berlin, Berlin, Germany (johnatan.bonilla@caroycuervo.gov.co) (https://orcid.org/0000-0002-8166-3548)*

Ruth Yanira Rubio López, *Instituto Caro y Cuervo, Bogotá, Colombia (ruth.rubio@caroycuervo.gov.co) (https://orcid.org/0000-0002-1396-9238)*

and

Andrés Steban Luna Cortés, *Instituto Caro y Cuervo, Bogotá, Colombia (andres.luna@caroycuervo.gov.co) (https://orcid.org/0009-0002-6021-1785)*

Abstract: The Instituto Caro y Cuervo (Caro and Cuervo Institute, ICC) was initially founded to complete Rufino José Cuervo's *Diccionario de Construcción y Régimen* (Dictionary of Construction and Usage) (Cuervo and ICC 1998) and has since expanded its mission to include the research and promotion of Colombia's linguistic heritage. Following this lexicographic tradition, the Institute developed the *Diccionario de Colombianismos* (Dictionary of Colombianisms, DiCol) (ICC 2018) using the proprietary software TshwaneLex, which facilitated the production of its print version but created a dependency on third-party resources, the need for a more flexible and independent solution became apparent. In response, this report introduces *LEXICC — Diccionarios y Lenguajes* (Dictionaries and Languages, LEXICC), a new, tailored online Dictionary Writing System (DWS) developed from scratch as an open-source solution. LEXICC empowers researchers, linguists, lexicographers, and anyone interested in dictionaries to create and manage their lexicographic resources separately. This paper details the design and development process of LEXICC, highlights its main functionalities, and discusses the electronic adaptation of the DiCol, now accessible online through LEXICC.

Keywords: ELECTRONIC DICTIONARIES, DICTIONARY WRITING SYSTEM, COLOMBIAN SPANISH, CARO AND CUERVO INSTITUTE, DICTIONARY OF COLOMBIANISMS, NON-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

FUNCTIONAL REQUIREMENTS, FUNCTIONAL REQUIREMENTS, DEMO DICTIONARY, LEXICOGRAPHER DIRECTOR

Opsomming: LEXICC: Die ontwerp en ontwikkeling van 'n aanlyn woordeboekskryfprogram. Die Instituto Caro y Cuervo (Caro- en Cuervo-Instituut, ICC) is aanvanklik gestig om Rufino José Cuervo se *Diccionario de Construcción y Régimen* (Woordeboek van samestelling en gebruik) te voltooi (Cuervo en ICC 1998). Sedertdien is die Instituut se missie uitgebrei om die navorsing en bevordering van Colombia se linguistiese erfenis in te sluit. In navolging van hierdie leksikografiese tradisie, het die Instituut die *Diccionario de Colombianismos* (Woordeboek van Colombianismes, DiCol) ontwikkel (ICC 2018) met behulp van die kopieregsagteware TshwaneLex, wat die vervaardiging van die gedrukte weergawe daarvan vergemaklik het, maar wat terselfdertyd 'n afhanklikheid van derdeparty-hulpbronne geskep het. Dit het die behoefte aan 'n buigsamer en meer onafhanklike oplossing duidelik laat word. In reaksie hierop stel hierdie verslag *LEXICC — Diccionarios y Lenguajes* (Woordeboeke en Tale, LEXICC) 'n nuwe, pasgemaakte aanlyn woordeboekskryfprogram (WSP) bekend wat reg van die begin af as 'n oopbron-oplossing ontwikkel is. LEXICC bemagtig navorsers, linguiste, leksikograwe, en enigiemand wat in woordeboeke belangstel om hul leksikografiese hulpbronne afsonderlik te skep en te bestuur. In hierdie artikel word die ontwerp en ontwikkelingsproses van LEXICC uitvoerig beskryf, die hooffunksies daarvan word beklemtoon, en die elektroniese aanpassing van die DiCol, wat nou aanlyn via LEXICC toeganklik is, word bespreek.

Sleutelwoorde: ELEKTRONIESE WOORDEBOEKE, WOORDEBOEKSKRYFPROGRAM, COLOMBIAANSE SPAANS, CARO- EN CUERVO-INSTITUUT, WOORDEBOEK VAN COLOMBIANISMES, NIE-FUNKSIONELE VEREISTES, FUNKSIONELE VEREISTES, DEMO-WOORDEBOEK, LEKSIKOGRAAF-BESTUURDER

1. The lexicographical tradition of the Instituto Caro y Cuervo

Founded in 1942 to complete Rufino José Cuervo's *Diccionario de Construcción y Régimen* (DCR, Dictionary of Construction and Usage) (Cuervo and ICC 1998), the Instituto Caro y Cuervo (ICC) has significantly contributed to the development of dictionaries for both Spanish and endangered languages in Colombia. Cuervo began this monumental work in 1872, and the ICC's Lexicography Department published the final volumes in 1994. Concurrently, the ICC enhanced Colombian lexicography by creating a lexicographic glossary (Montes et al. 1986) for the *Atlas Lingüístico-Etnográfico de Colombia* (ALEC, Linguistic and Ethnographic Colombian Atlas) (ICC 1983) and supporting the creation of the *Nuevo Diccionario de Colombianismos* (NDC, New Dictionary of Colombianisms) (Haensch and Werner 1993).

Building on these efforts, ICC researchers have been part of various lexicographic projects. One example is the *Diccionario Básico de la Lengua de Señas Colombiana* (DBLSC, Basic Dictionary of Colombian Sign Language) (INSOR and ICC 2011), the first Colombian sign language dictionary representing dialectal variation as it was developed from corpora collected in two major cities: Cali and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Bogotá. In 2014, ICC researchers were involved in two online publications implementing MediaWiki for dictionary management. The first was the *Diccionario Bilingüe Sáliba–Español*[1] (DBSE, Sáliba–Spanish Bilingual Dictionary) (ICC 2014), developed after extensive fieldwork and documentation in collaboration with indigenous knowledge keepers from the Asociación de Autoridades Indígenas de Orocué (Association of Indigenous Authorities of Orocué, ASAISOC) (Dueñas and Gómez 2015). The second was the *Diccionario Académico de Medicina*[2] (DIACME, Academic Dictionary of Medicine) (Academia Nacional de Medicina de Colombia 2023), primarily advised by the ICC (Bernal et al. 2020). The ICC's lexicographic work also extends into pedagogical and sociocultural projects and research. Notable contributions include the *Glosario de aprendizaje del Español de Colombia* (Glossary for Learning Colombian Spanish) (Nieto 2017) and the *Léxico de la Violencia en Colombia 1948–1970* (Lexicon of Violence in Colombia 1948–1970), based on the literature of the violence canon (Rozo et al. 2020).

However, the most significant work in recent times is the *Diccionario de Colombianismos* (DiCol, Dictionary of Colombianisms) (ICC 2018). This dictionary was created after reviewing various dictionaries of Americanisms and Colombianisms, such as the NDC (Haensch and Werner 1993), the *Diccionario de Americanismos* (Dictionary of Americanisms) (Asociación de Academias de la Lengua Española 2010), and the *Breve Diccionario de Colombianismos* (Brief Dictionary of Colombianisms) (Academia Colombiana de la Lengua 2012). It was further refined through collaboration with researchers from various Colombian regions, resulting in 6,000 entries, 8,000 definitions, and 4,500 examples (ICC 2018: 16).

The DiCol was developed with the commercial Dictionary Writing System (DWS), TshwaneLex (TLex). However, despite TLex's robust capabilities, its proprietary nature poses significant challenges. In the first place, the procurement of a paid license may present a notable obstacle for budget-constrained researchers, small institutions and native communities. Additionally, in the case of DiCol, the team of lexicographers needed the assistance of a developer to build the structure of the entries, which affected the autonomy in the initial phase of the project. Lastly, for the online publication of DiCol it was necessary to devise another system that allowed both its online viewing and constant updating of the entries by the team of lexicographers.

In summary, the experience in lexicographic projects and the approach to electronic lexicography (such as the creation of DiCol) have awakened multiple demands in the Institute: firstly, there is the need for efficient and user-friendly tools for the development of dictionaries to facilitate the work of the Institute's lexicographers and researchers. For example, the *Glosario de Aprendizaje del Español de Colombia* (Glossary for Learning Colombian Spanish) (Nieto 2017) was edited through a word processor which implies difficulties in the organisation and efficient search of entries, limitations in the handling of large volumes of data, and problems integrating and updating information in a collaborative and automated manner.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Secondly, there is the need to ensure the publication and consultation of these resources in an accessible and efficient way for different audiences. Several of the Institute's dictionaries are in printed format or in spaces that do not guarantee the safeguarding and easy consultation of the data. Thirdly, it is important to integrate the functionalities that optimize the processes of creating, updating, and handling lexicographic data. For instance, the migration of the DBSE (ICC 2014) from its MediaWiki format will be vital to ensure the safeguarding of information and the constant updating of the dictionary by the community. Finally, it is vital to highlight the lexicographic projects that are currently under development and that reflect the need for a platform for their publication, access, and consultation: the *Diccionario Español–Andoque* (Spanish–Andoque Dictionary) (Andoque and Landaburu 2023), created with members of the Aduche Reserve in the Amazon (Colombia), and the *Corpus Léxico del Español de Colombia* (CorlexCo, Lexical Corpus of Colombian Spanish) (Nieto 2020), a linguistic corpus composed of lexical combinations used in Colombia and other Spanish-speaking countries, which could also became a dictionary in the future.

The ICC Corpus and Computational Linguistics (LICC) research line initiated the project *LEXICC — Diccionarios y Lenguajes* (Dictionaries and Languages, LEXICC)[3]. This open-source DWS allows the creation, management, publishing, and consultation of the lexicographic products of the Instituto Caro y Cuervo (whether new or existing), and any dictionary created by platform users, specialists, or not in lexicography. This report outlines the construction process of LEXICC, its requirements, its main features (users and interfaces), future improvements, and the digitisation process of the DiCol, the first printed dictionary published on the platform.

2. Building LEXICC

To design and develop the LEXICC platform, a multidisciplinary team was formed, consisting of a full-stack web developer, a graphic designer, and four linguists. The developer handled the technological implementation of the platform and supporting technical requirements; the graphic designer worked on the layout and visualization of the system; and the linguists conducted preliminary research for platform development, proposed requirements, advised on development and design tasks, and tested the platform's usability.

Methodologically, a cascade-hybrid system development was adopted. This approach, based on the six phases of the system development life cycle — feasibility study, analysis, design, coding, testing, and maintenance (Leau et al. 2012), permitted the execution of each phase systematically and interdependently. The construction of LEXICC began in 2019 with a bibliographical inquiry into electronic lexicography, systems or tools for dictionary creation and publication, and Corpus Query Systems (CQS) and how they contribute to lexicographic work. The characteristics of the Institute's dictionaries and various types of electronic dictionaries were also reviewed. Based on this analysis, the platform

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

requirements outlined in Section 3 were established. In the first semester of 2020, the design and development phases commenced with the definition of the technical specifications. During this semester, the database was built, and the back-end services were implemented ensuring traceability through GitLab[4] (these technical specifications are detailed in Section 4).

In 2021, the focus shifted to developing user interfaces presented in Section 5. During this period, interfaces for different user types and specific services for dictionary structures were created. These developments have continually improved to enhance the user experience up to 2023. By then, the dictionary creation functionality was refined, the coordinator lexicographer interface (see Section 5.3) was developed for efficient dictionary management, and the formatting of both print and digital versions of the articles was improved. Finally, as presented in Section 6, the 2018 DiCol was migrated to the new database, so it is now accessible and managed through the platform.

3. Requirements analysis

To identify and fulfil the specific needs for LEXICC, a comprehensive three-part requirements analysis was conducted. Firstly, key aspects of electronic lexicography and features from renowned dictionary writing systems (DWS) were analysed. This included systems such as TLex (TshwaneDJe HLT 2023), EELex (Institute of the Estonian Language 2023), Dictionary Production System or DPS (IDM France n.d.), and Lexonomy (Měchura 2017). Secondly, existing ICC dictionaries (mentioned in Section 1) were reviewed in detail, and consultations were held with their authors to gain deeper insights into the need for a customised platform for the institute. Finally, the experience of the LICC group in developing linguistic information management systems was leveraged. This includes projects like the *Corpus Lingüísticos del Instituto Caro y Cuervo (CLICC)*[5] (CLICC, Linguistic Corpora of the ICC) (Rubio et al. 2023; Rubio and Bernal 2019) and the *Sistema de Información Geográfica del ALEC*[6] (SIGALEC, ALEC's Geographic Information System) (Bonilla et al. 2020; Bonilla and Bernal 2020), particularly their collaborative methodologies in interdisciplinary processes and project management.

Based on this analysis, three fundamental components of a DWS were identified: a database, a dictionary entry editing interface, and administrative tools for user registration and dictionary tracking. However, this structure was refined for current development with: (1) a database (DB) (Section 4) to store and manage all dictionary data; (2) a General Interface (GI) (Section 5.1) for the presentation of the project to anonymous users, and options for registering or logging in; (3) a Visualization Interface (VI) (Section 5.3.2) for the query of public dictionaries; (4) an Administrative Interface (AdI) (Section 5.2) for user management: roles, projects, and editable aspects of website interfaces; (5) a Demo Interface (DI) (Section 5.3.1) that allows users to create and consult demo dictionaries for educational purposes; and (6) a Lexicographer Interface (LI) (Section 5.3) for expert

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

researchers and their collaborators to manage their dictionary projects comprehensively. This refined structure ensures a customised interface to the specific needs of its users, from public interaction to specialised research and project management.

To achieve this, a list of requirements was gathered. These requirements were divided into three categories: non-functional (Table 1), functional (Table 2), and users and use cases. According to Sommerville (2011), functional requirements refer to what the system can do; therefore, they describe the functions or services that the system should offer. In contrast, non-functional requirements relate to the properties of the system that enable it to adequately provide those functions or services, such as performance, response time, and storage capacity, among others. Finally, use cases identify the different scenarios in which users can interact with the system.

Regarding LEXICC, non-functional requirements ensure the platform's reliability and user experience, while functional requirements and use cases specify the tasks each user role can perform within the system. Each requirement was organised in tables with the name and description of the requirement, the requirement source, and the priority level (High/Essential, Medium/Desired, Low/ Optional).

Table 1: Non-functional requirements of LEXICC

No.	Requirement Name	Description	Priority
RNF01	Flexibility	Store various types of information and create new dictionaries.	High/Essential
RNF03	Performance	Queries responded to in less than 5 seconds.	Medium/Desired
RNF04	Information Update	Constant updates and immediate saving.	High/Essential
RNF05	Security	Protection against unauthorised access.	High/Essential
RNF07	Backup Copies	Weekly backups of data and developments.	High/Essential
RNF08	Support and Maintenance	Documentation for future adjustments.	High/Essential
RNF09	Graphical Interface Policies	Compliance with design policies.	High/Essential
RNF10	Query Browsers	Accessible from the most used browsers.	High/Essential
RNF11	Applications	Offline use as a downloadable application.	Medium/Desired

The non-functional requirements of LEXICC (Table 1) focused on key aspects such as flexibility in storing diverse information types, quick response times for

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

queries, and constant updates with immediate saving. Security measures protect against unauthorised access, with weekly data backups and comprehensive support documentation ensuring system integrity and maintainability. It was required that the graphical interface comply with design policies and be accessible from major browsers, with offline use available as an optional feature.

Table 2: Functional requirements of LEXICC

Interface	General Description of Requirements	Reference Platforms or Dictionaries
General Interface	Project data, menu with resources and dictionaries, options to login and sign up, it allows to visit available dictionaries with Visual Interface and to access the Demo Interface.	All
Administrative Interface	The administrator can assign roles, consult, and export query statistics, and edit news and content. The system will log changes and the users who made them.	CLICC, TLEX
Lexicographer Interface	Tools for configuring dictionaries, editing, and modifying information, viewing entries in three formats, adding lexical relations, and managing editing windows for bilingual dictionaries. Integration with CQS.	Sáliba–Spanish Dictionary, CLICC, TLEX, Lexonomy, TERMSTAR, Golden Dict, Termania
Demo Interface	Menu to customise searches, create DEMO dictionaries/glossaries, download in Word, PDF, XML, save changes in the system, and participate in consultation forums.	Wordreference, DEBWrite
Visualization Interface	Search box with various search types: approximate, exact, alphabetical, begins with, contains, ends with, anagrams, random, manual configuration, region, semantic field, marks, visual dictionary.	RAE, INSOR Dictionary

On the other hand, functional requirements (Table 2) defined the specific capabilities needed for each interface within LEXICC. For instance, The AdI allows administrators to manage user registrations, assign roles, export query statistics, and edit content, with all changes logged in. The LI provides tools for configuring and editing dictionaries, viewing entries in different formats, adding lexical relations, and managing bilingual dictionaries, integrated with Corpus Query Systems (CQS). The DI enables users to customise searches, create demo dictionaries, download content, save changes, and participate in forums. Finally,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

VI offers various search options within the dictionaries.

In terms of user experience, the requirements imply dedicated access and features for distinct types of users. The requirements defined per user experience were as follows:

- **Administrator User (AU)**
  - Manage the entire system, including user and dictionary management.
  - Register, modify, search, or delete users.
  - Activate or deactivate roles.
  - Add, delete, and manage the display of dictionaries.
  - Edit content viewed by users.
  - Obtain statistical reports on user activity and general queries.
- **Guest or Anonymous Users (GU)**
  - Navigate the general webpage.
  - Explore available dictionaries through the VI.
  - Create dictionaries using the DI but cannot save them.
  - Access sign-up and login interfaces.
- **Registered User (RU)**
  - All Anonymous Users capabilities
  - Save and download information from created and consulted dictionaries.
  - Participate in collaborative tools.
  - Choose query preferences.
  - Review recent searches.
  - Select specific dictionaries to consult.
  - Save consulted words.
  - Forums participation.
- **Lexicographer Director (LD)**
  - Define the type of dictionary and its general structure.
  - Define styles and fonts of dictionary.
  - Grant Coordinator and Editor roles to users.
  - Supervise the team and indicate progress states of an entry (unassigned, in editing, for review, approved).
  - Grand consultation and download permissions for their dictionary.
  - View statistics of the dictionary's development.
  - Create, edit, and delete entries.
  - Manage forums, comments, and dictionary suggestions.
  - Include notes in entries.
- **Lexicographer Coordinator (LC)**
  - Create, edit, and delete entries.
  - Manage forums, comments, and dictionary suggestions.
  - Include notes in entries.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

- **Lexicographer Editor (LE)**
  - Create, edit, and delete entries.
  - Include notes in entries.

To summarise, the user experience requirements are tailored to provide dedicated access and features for distinct types of users, ensuring each role can efficiently perform its designated tasks. AUs manage the entire system and oversee user activities, while GUs have limited navigation and dictionary exploration capabilities. RUs gain additional privileges, such as saving and downloading information and participating in collaborative tools. LDs and LCs have more specialised functions related to the creation and management of dictionaries, with the LD also overseeing team roles and progress. LEs focus on entry creation and editing. Each role's specific capabilities ensure a structured and efficient workflow, promoting effective dictionary management and user interaction.

4. Technical implementation and development

The technical implementation for the LEXICC project aims to address challenges identified from previous developments by the LICC group and to remain at the forefront of current software development. While the CLICC and SIGICC platforms (Rubio et al. 2023; Rubio and Bernal 2019; Bonilla et al. 2020; Bonilla and Bernal 2020) utilised relational databases constructed in MySQL and PostgreSQL respectively, and PHP, LEXICC has adopted the MERN stack (MongoDB, ExpressJS, React, and NodeJS) (Aggarwal and Verma 2018). These technologies are widely used, constantly evolving, and provide dynamic and efficient interfaces, significantly enhancing user experience and data visualization fluency.

Firstly, MongoDB is a NoSQL database designed for scalability, flexibility, and storing large volumes of data. Unlike traditional table-based relational databases, MongoDB uses collections and documents, facilitating the configuration of a dynamic and adaptable structure (Győrödi et al. 2022). This allows the registration of dictionaries without a predefined structure. Users can create their own configurations for their dictionaries without requiring specialist intervention, enabling the storage of attributes not defined during the requirements phase based on a relational model.

As described in Section 3, LEXICC's database allows for the storage of information regarding dictionaries, users, and queries. It is essential for registering and systematising lemmas, definitions, parts of speech, and other entry components, as well as for administrative actions such as user activation and deactivation, recording entry progress states, organising lexicographer teams and task assignments, and generating statistics, among others. Additionally, the database facilitates the import of dictionaries created in other systems, which is crucial for populating the platform. The LEXICC database is flexible for storing any type of information (documents, videos, audios, texts, images, etc.), and its per-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

formance supports the continuous growth of system data. This solution effectively resolves a challenge encountered during the development of CLICC. Specifically, the integration of new corpora and metadata with differing attributes necessitated normalisation to the sixth normal form. This process involved identifying and eliminating all multivariable relationships within the database, allowing for the registration of a table to store attributes. Subsequently, information was recorded in another table where responses to these attributes were stored and linked to the newly registered corpora entity (Rubio et al. 2023).

Secondly, ExpressJS is a web framework for NodeJS used to create APIs simply and efficiently. This tool was used for data persistence and to meet the group's need to share information between applications and even entities that required it. Thirdly, React is a JavaScript library widely used in modern web development due to its reactive interfaces, allowing users to have a smooth and fast experience when querying or configuring dictionaries. This implementation was based on an atomic design methodology, which is grounded in the creation of modular and scalable systems, making each created component reusable. Lastly, Node.js, the server-side JavaScript runtime environment, integrates seamlessly with Express.js and React to provide a comprehensive and efficient solution for web application development (Saundariya et al. 2021).

Finally, the development process was supported by traceability management in GitLab, a tool that allows for managing the LEXICC repository and controlling all platform versions. GitLab has also been very useful for tracking the development process, as there is a dedicated space available to help plan, organise, review, and approve workflow tasks until the development project is released to production.

5. LEXICC interfaces

Following rigorous research and requirements gathering, five key interfaces that power LEXICC were developed: General Interface (GI), Administrative Interface (AdI), Demo Interface (DI), Lexicographer Interface (LI), and Visualization Interface (VI). In the following sections, their main characteristics will be explained.

5.1 General Interface (GI)

The GI, depicted in Figure 1, serves as the homepage to navigate dictionaries and familiarise oneself with LEXICC. The numbered modules in this figure are described below.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Figure 1: Homepage in the GI

The main menu (Module 1) consists of 6 hyperlinks: (1) *Inicio* (Home), which appears by default when entering the platform to welcome the users; (2) *¿Qué es LEXICC?* (What is LEXICC?) presents a description of the objectives and characteristics of the platform; (3) *Diccionarios* (Dictionaries) allows access to the available dictionaries; (4) *Guía de uso* (User Guide) is a dropdown menu containing the platform's usage manuals, including the RU manual and the LD manual, which provide detailed information about the system tools; (5) *Equipo* (Team) displays the names and contact information of the research group responsible for creating LEXICC, including linguists, designers, and developers; finally,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

(6) *Iniciar sesión* (Log in) enables users to register in LEXICC and access the platform as a RU. If they do not have an account in the system, they fill out the registration form that appears at the link *Crea una cuenta ahora* (Create an account now) (Figure 2). However, if they already have an account, they can log in by entering their email and password.

Figure 2: Interface to log in

Referring back to Figure 1, users can also access the dictionaries of the system in Module 2 via the *Inicio* (Home) hyperlink. By clicking on *Ir al diccionario* (Go to Dictionary) users are directed to the VI of the DiCol, which is currently the only available dictionary (see Section 5.3.2). Notably, registration on the platform is not required to access the dictionaries, if they have been enabled by the AU. Additionally, in Module 3, users can access the DI without registration. They simply need to click on the *Ir al demo para creación de diccionarios* (Go to dictionary creation demo) option and they will immediately be directed to the DI (presented in Section 5.3).

5.2 Administrative Interface (AdI)

To access the AdI, users must log in and select the AU role. This selection is made by clicking the blue arrows next to the role, which will display a window where the AU role can be chosen (Figure 3). The AdI enables the management users and dictionaries within the system.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

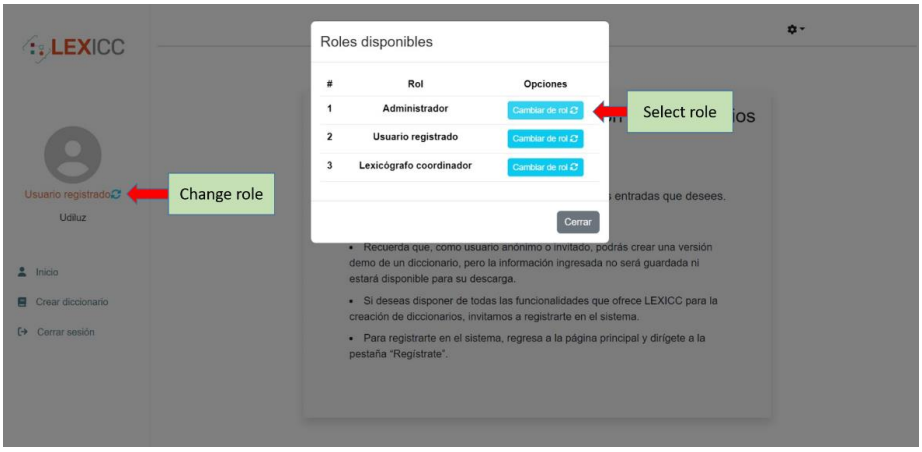


Figure 3: Selection of the AU role by a RU

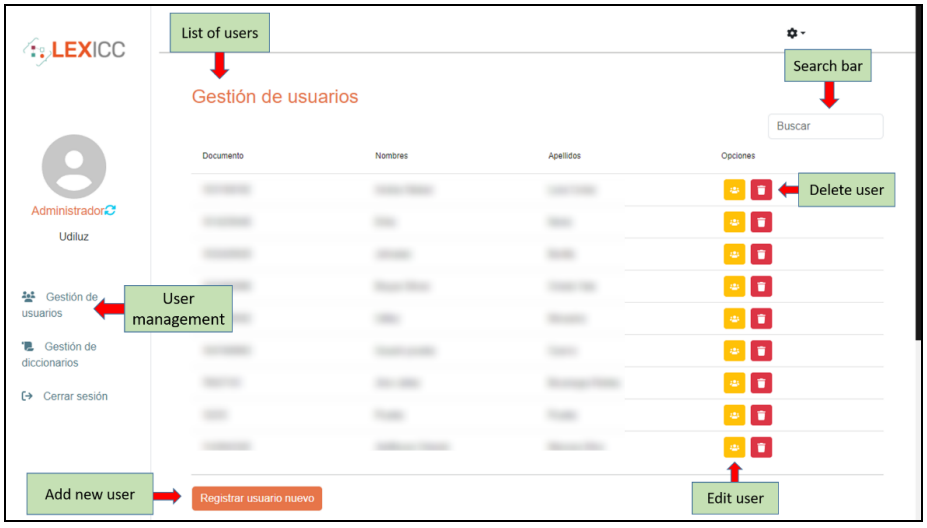


Figure 4: User management in the AdI

For user management, the AU must select the *Gestión de usuarios* (User management) option in the left menu of the interface (see Figure 4). This will display a list of registered users within the system. Users can be deleted or edited using the options on the right-hand margin. Additionally, specific users can be searched using the search bar in the top-left corner. New users can be added by clicking the *Registrar usuario nuevo* (Register new user) button at the bottom of the interface.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Editing a user involves modifying their personal data or assigning permissions within the system. To edit a user, the AU must click on the yellow symbol in the right margin of the interface, which will display an editing window (see Figure 5). In this window, the AU can change the roles of the RU and enable the dictionaries in which they can participate as directors or editors. The image indicates that the user has the DL role deactivated and two assigned dictionaries. The AU can modify these permissions at any time.

Figure 5: User editing window in the in the AdI

5.3 Lexicographer Interface (LI)

The LI is used by the three types of lexicographer users described in Section 3 (LD, LC, and LE). To access this interface, users must log in with their e-mail and password (Figure 2, Section 5.1), and then select their assigned role from the role change window (Figure 3, Section 5.2). The LI currently includes a user management menu, an option to create new dictionaries, a menu to select the dictionary where lexicographers wish to work, and the DI for creating and editing dictionaries. Figure 6 shows the LI used from the perspective of an LD. On the left side, DiCol is shown as the selected dictionary for editing, and a new dictionary can be created using the option below. After this, the LD can manage the users of the dictionaries under their responsibility, similar to the AU's user management (Figure 4, Section 5.2). To access the DI, the LD clicks on *Mi diccionario* (My dictionary), where the structure and entries configuration of the chosen dictionary are managed. Figure 6 also displays the General module of the DI, where the general description of the chosen dictionary (DiCol in this case)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

is entered. For further details, Section 5.3.1 will explain what the DI consists of and the process of creating a dictionary.

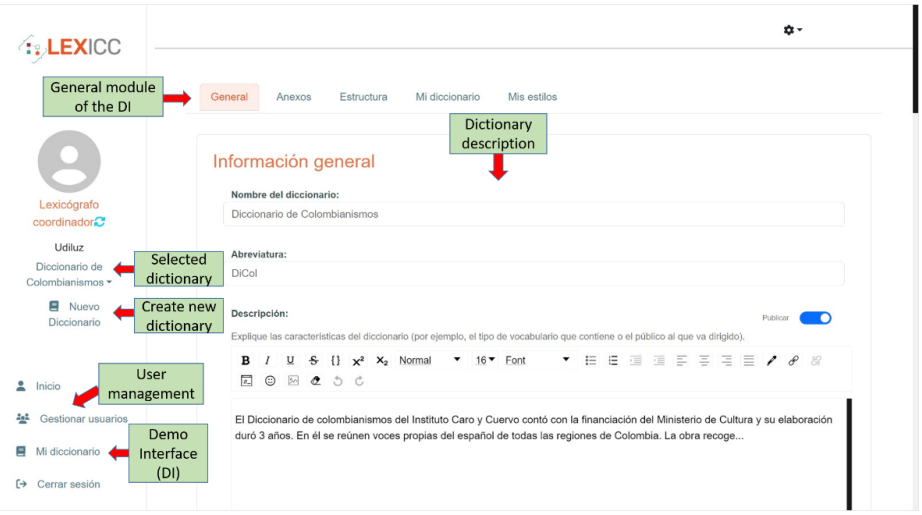


Figure 6: General module of the DI viewed from the LI

5.3.1 Demo Interface (DI)

The DI comprises four modules:

(1) *General* (General module): Used to enter general information about the dictionary, such as its name, description, abbreviation, team involved, information about the lexicographical project, and copyright information (Figure 6, Section 5.3).
(2) *Anexos* (Annexes module): Allows including additional information such as the dictionary's usage guide, abbreviations, symbols, and any other relevant documents (Figure 7).
(3) *Estructura* (Structure module): Allows to organise the structure of the dictionary entries, including the headword, definition(s), example(s), etc. (Figure 9).
(4) *Mi diccionario* (My Dictionary module): Facilitates the creation of dictionary entries following the predefined structure (Figure 16).
(5) *Estilos* (Styles module): Allows users to edit the style of the entry structure and the style of the VI (Figures 17 and 18).

For GU and RU, only modules (1), (3) and (4) are available, while lexicographer users (LD, LC, and LE) can use all of them. This section of the article showcases the modules of the DI using the interface of an LD, as this user has the most comprehensive editing functionalities. To explain the entry structure, also known as

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

microstructure (Atkins and Rundell 2008), the term *entry components* will often be used, referring to "the separate pieces which go to make up the dictionary entry" (Atkins and Rundell 2008: 202). These pieces or elements vary depending on the type of dictionary (monolingual or bilingual), its purposes, and the intended audience. Generally, entries in a monolingual dictionary will consist of a headword, a definition, grammatical information, and other linguistic labels regarding usage (dialect, register, attitudes).

To create a dictionary in LEXICC, users first select the *Nuevo diccionario* (New dictionary) option from the left menu of the DI (Figure 6, Section 5.3), which will open the form that needs to be filled out. The first part of the form is the General module, presented in Figure 6 (Section 5.3), where users complete the dictionary description and click the *Crear diccionario* (Create dictionary) option at the end of the module (Figure 7).

Figure 7: Button to create a dictionary

After this, they should proceed to the Annexes module to enter additional information about the dictionary (Figure 8). Both modules have WYSIWYG editors (What You See Is What You Get) for inserting and editing information, and users can decide which texts or documents appear published in the VI using a toggle button. The general description of the dictionary will appear at the top of the VI, and its annexes will appear on the left side (Figure 19, Section 5.3.2).

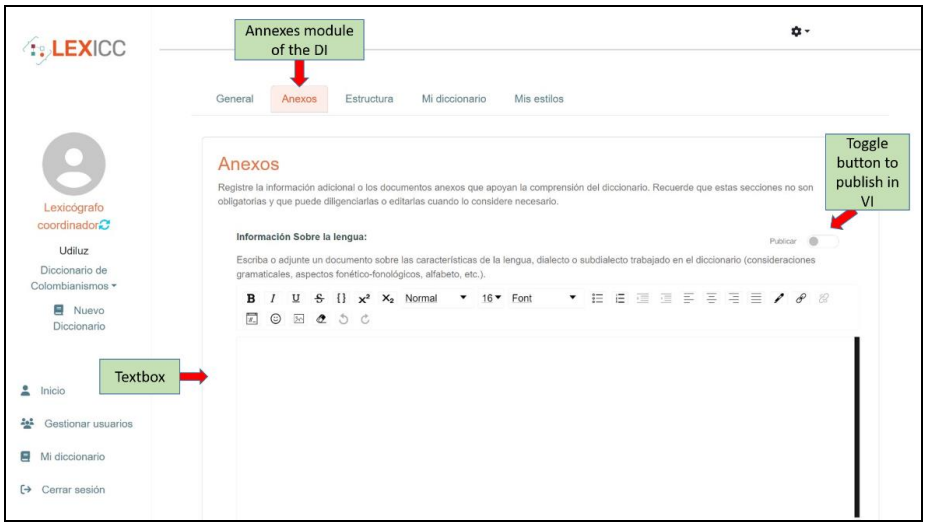


Figure 8: Annexes module in the DI

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Next, users establish the entry structure in the Structure module illustrated in Figure 9. To add entry components, the first step is to click on the blue button with the plus sign (+), indicated in the same figure. This action will display the window presented in Figure 10, where users must choose the component type and the symbols accompanying the component.

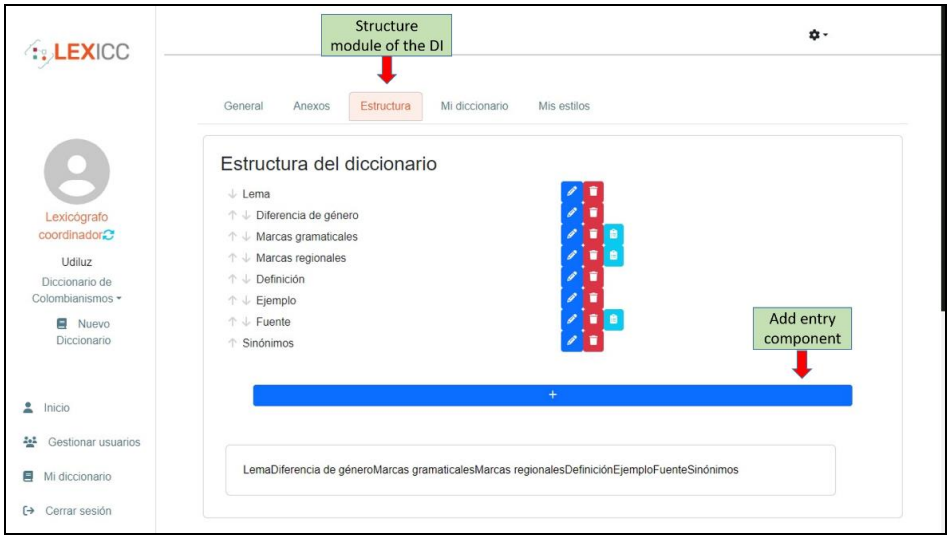


Figure 9: Structure module in the DI: Button to add entry component

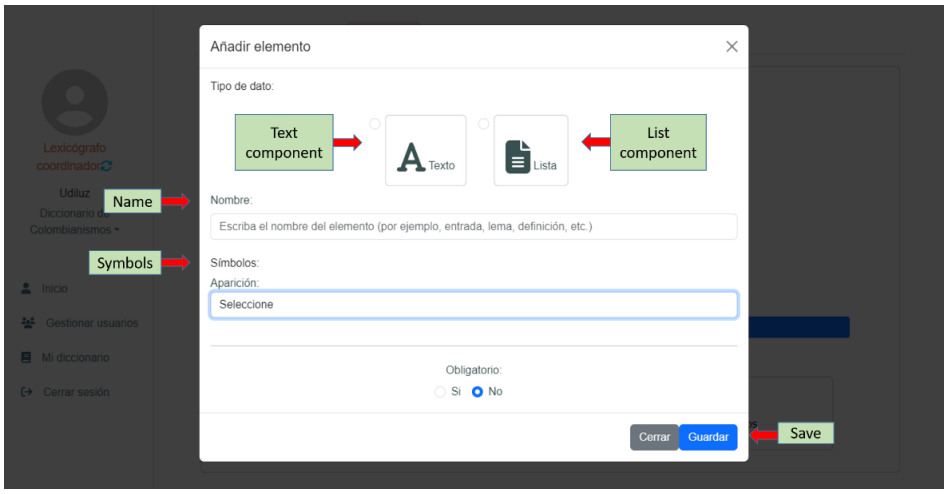


Figure 10: Menu to add entry components in the DI

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

An entry in LEXICC can include text components (e.g., definitions) and list components where a specific option is selected (e.g. parts of speech). If users want to create a text component named *Lema* (headword), they should click on the circle corresponding to text data (Figure 11, step 1) and specify the name of the entry component (Figure 11, step 2). Afterward, the user designates the placement of the symbols accompanying the component (left, right, or both), and the type of symbol (e.g., period, comma, colon, bars, etc.) (Figure 11, steps 3 and 4).

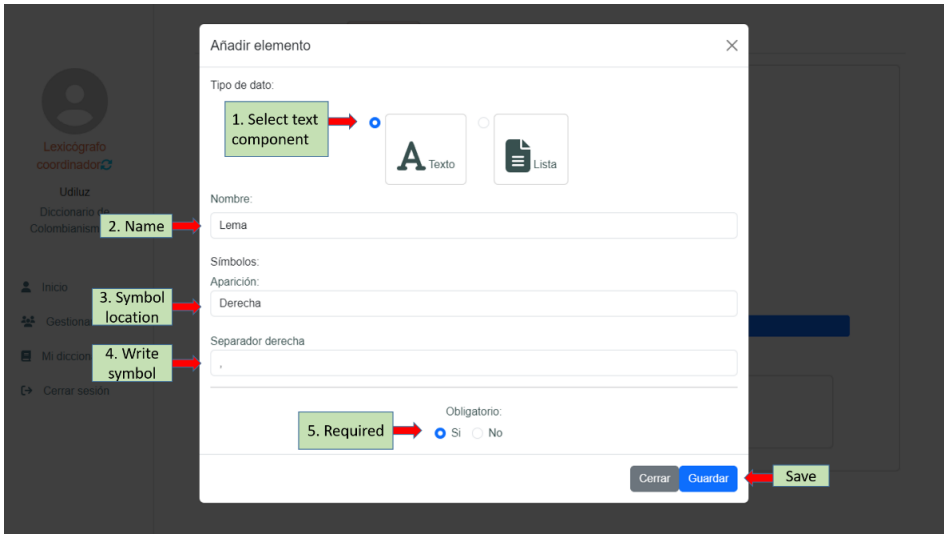


Figure 11: Menu to add entry components: Example of text component

In the example shown in Figure 11, the headword will be accompanied by a comma on the right side, and it will be a required component in the dictionary (step 5), meaning it will appear in all registered entries. If users wish to include a list component in the entry structure, the procedure described is repeated, but users must click on the circle corresponding to list data. Figure 12 illustrates all the steps to add the list component named *Categoría gramatical* (part of speech), which will appear in parentheses in the dictionary.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

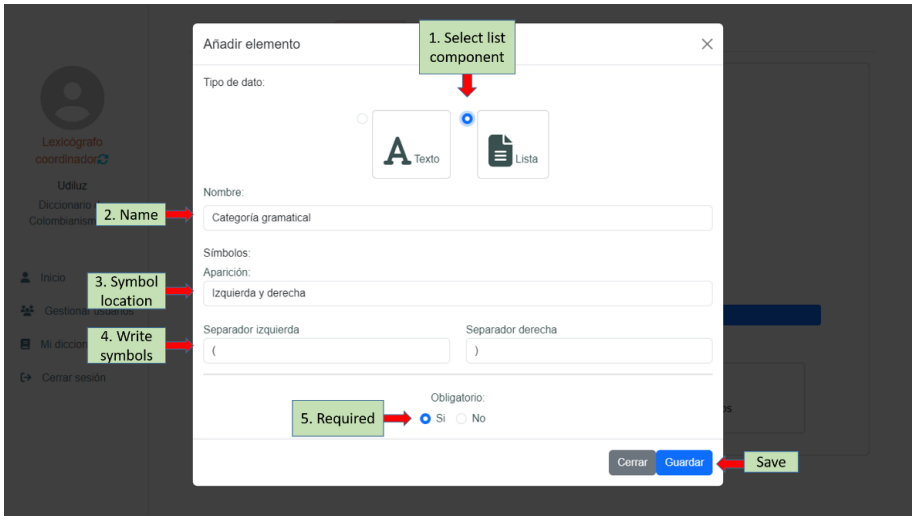


Figure 12: Menu to add entry components: Example of list component

Considering that a list component consists of different attributes, the corresponding menu must be filled out. To access this menu, users must return to the main interface of the Structure module and click on the blue icon located next to the list component that was previously added (Figure 13). This action will display the menu shown in Figure 14, where users can enter the attribute information that makes up the component.

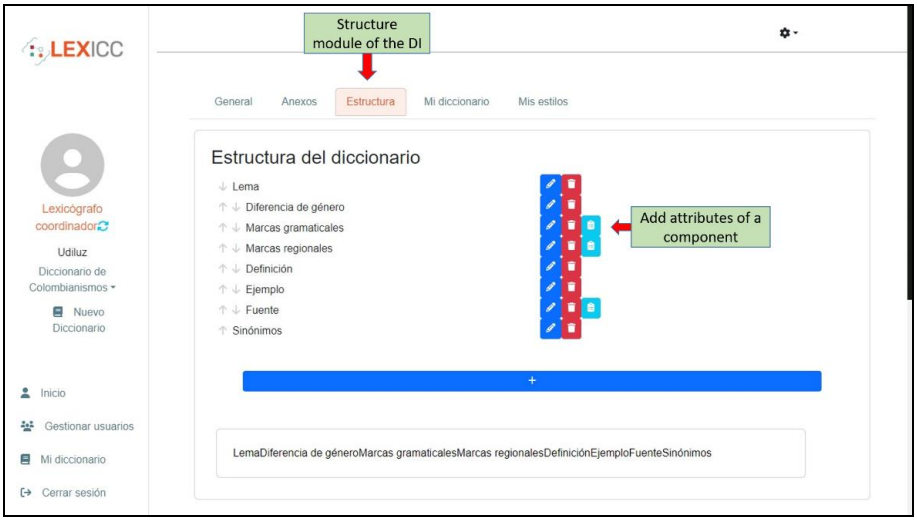


Figure 13: Structure module in the DI: Button to add attributes of a component

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

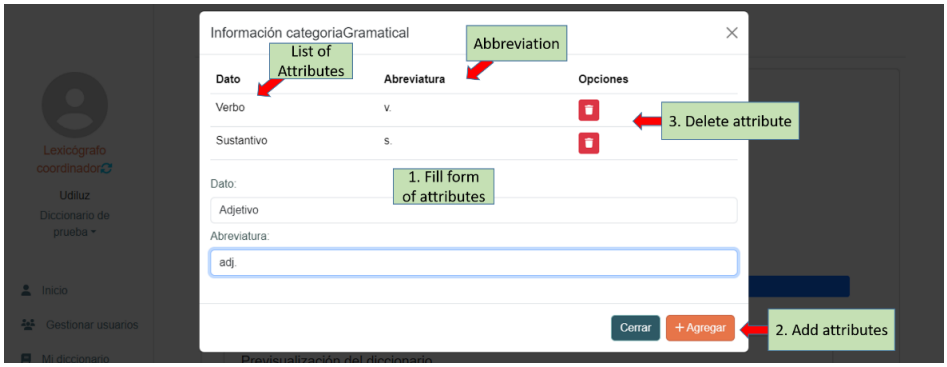


Figure 14: Menu to add attributes of an entry component in the DI

The first step is to include the name of the attribute and its abbreviation. For example, the *Adjetivo* (Adjective) is an attribute for a parts of speech component, and the form *adj.* is its abbreviation (Figure 14, Step 1). The second step is to click the *Agregar* (Add) button (Figure 14, Step 2), and the new attribute will appear at the top of the window. If the attribute is no longer needed, it can be deleted by clicking on the trash can icon (Figure 14, Step 3).

The main interface of the Structure module also allows for other actions as indicated in Figure 15: delete components that are no longer needed by clicking on the trash can icon, edit components as required by selecting the pencil icon, and reorder components by clicking on the gray arrows on the left side.

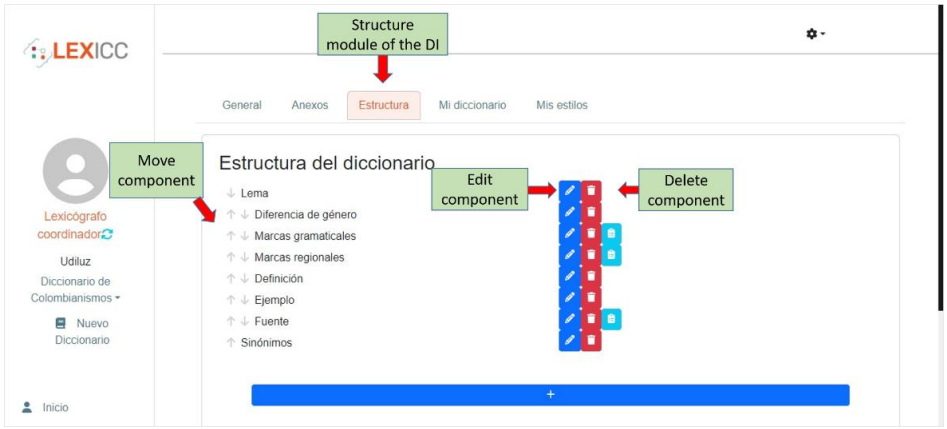


Figure 15: Structure module in the DI: Buttons to move, edit and delete a component

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Once the dictionary entries are structured, the user should navigate to *Mi diccionario* (My dictionary module) to fill out the proposed form (Figure 16). In this same window, entries can be searched, previsualized, and downloaded in their current state, whether draft or finalised.

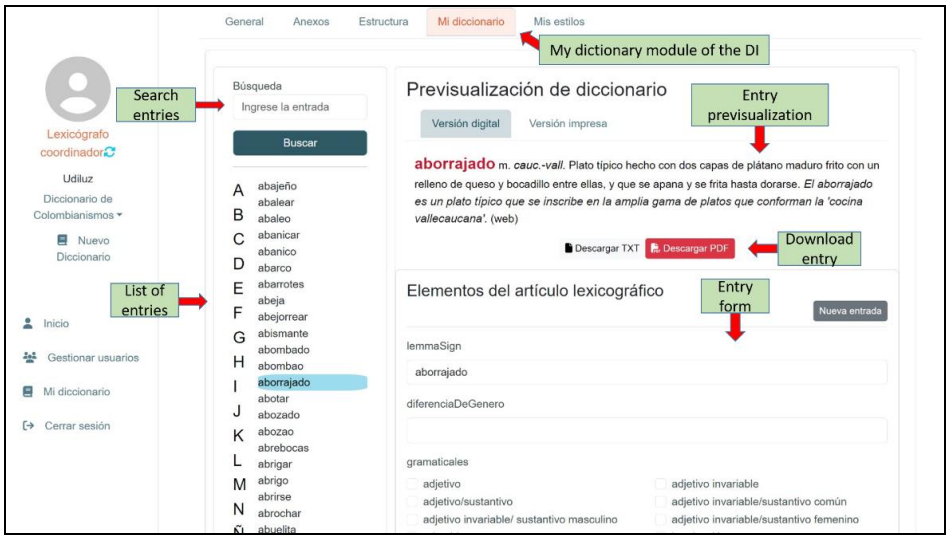


Figure 16: My dictionary module of the DI viewed from the LI. This figure shows the form for an entry in the DiCol

Each of the entry components can have the style assigned by the LD. Modifications can be made in the *Estilos* module (Styles), which is only available in the LI. Figure 17 shows the entry styles interface, where the LD can choose font colour, types of fonts, and font size, among other options. On the other hand, Figure 18 depicts the dictionary styles interface, where the LD can select the main and secondary colours that users will see in the VI. Figure 19 in Section 6 shows the VI of the DiCol, characterised by red as the main colour and black as the secondary colour. However, for other dictionaries, these colours may vary according to the preference of the lexicography team in charge.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Figure 17: Styles module of the DI: Entry styles interface

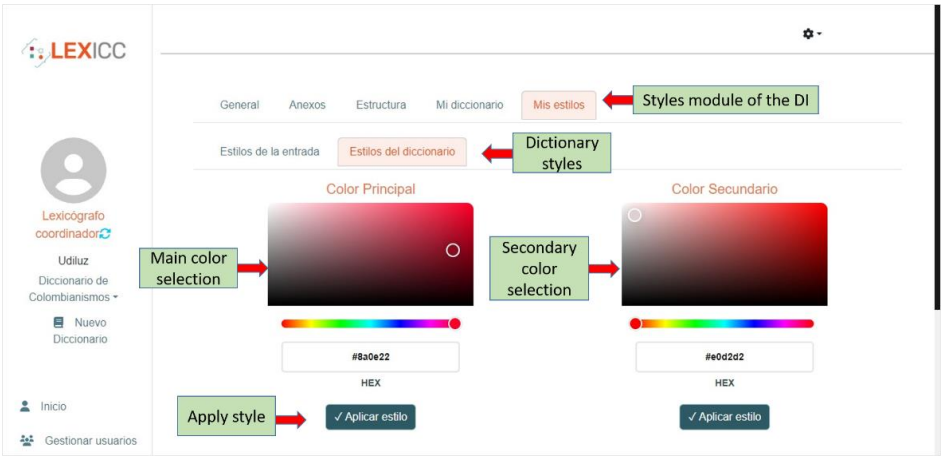


Figure 18: Styles module of the DI: Dictionary styles interface

5.3.2 Dictionary Visualization Interface (VI)

This space is where any user can consult available dictionaries, their features, project information, and contact details of the involved team. Currently, only the DiCol is available, as shown in Figure 19.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Figure 19: VI of the DiCol

The online version of DiCol in LEXICC consists of 5,915 entries. To achieve this transition to LEXICC, the team developer exported the DiCol data from TLex as a MySQL extension file, which was then transformed into JSON format using a script for better data handling. This JSON file loaded successfully into the LEXICC database in MongoDB, where data structure configuration was fundamental to ensure correct visualization. During this configuration, the ID naming the entry components in TLex was identified and collected in MongoDB, with other tags related to position, symbols, and styles (font, colour, text-alignment). These characteristics were implemented in the VI, giving the electronic version of the DiCol its resemblance to the printed version. Also, at the top of this interface, is an alphabetical search and the search box for word lookup, as depicted in Figure 19. Based on Almind's suggestions (2005), electronic dictionaries should display a visible search box, preferably at the top, and legible text, such as in DiCol electronic version. Furthermore, this new format aims to empower our lexicography team to draft entries separately without the constant assistance of a systems engineer. The DiCol editing interface allows lexicographer users to modify the entries' structure, edit existing entries, and create new ones, as shown in Figure 16.

6. Conclusions

The research work of the Instituto Caro y Cuervo has enabled the electronic and printed publication of various lexicographical projects on the languages and varieties of Spanish spoken in Colombia. Although they are meaningful to Colombian linguistic diversity recognition, their dispersion in different formats hampers their access and visibility for interested users, such as students, teachers,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

and researchers from the same institution. Therefore, LEXICC emerges as an innovative platform whose primary purpose is to host and manage the lexicographic projects of the ICC. One such venture is the DiCol, a project initially published in print in 2018, and its electronic version hosted on LEXICC promises to become a more accessible resource for different kinds of users, offering the opportunity to delve into the diverse regional varieties of Spanish-spoken in Colombia.

The creation of LEXICC adhered to the software development lifecycle methodology, which includes the planning and organisation of requirements, the design and development of the platform, functionality testing, and eventual deployment. This rigorous process culminated in an alpha version of LEXICC, incorporating the DiCol's database migrated from TLex. LEXICC, with its varied user interfaces and types, promises a user-friendly navigation experience and easy handling of its functionalities. For non-specialist users, LEXICC provides a demo dictionary, where can be created the structure of a basic dictionary. For specialist users, LEXICC offers the Lexicographer Interface, which allows the creation of a dictionary with different styles.

In the long term, the optimization of the platform will allow the Lexicographer Director to work collaboratively with a team by assigning tasks and drafting notes in entries to assess the progress of the lexicographic project. LEXICC users also can make suggestions to the available dictionaries through forums that will be available soon. Moreover, search capabilities will significantly improve within DiCol's electronic version with searches by linguistic labels: part of speech, dialect, register, domain, or attitude. Additionally, the system will offer integrated geolocation services in partnership with ICC's Geographic Information System platform. This integration aims to transform LEXICC into a multifunctional platform (cf. Kruyt 2003) capable of interacting with other electronic linguistic resources. Finally, a comprehensive methodology will be implemented to test the platform with external stakeholders, including ICC staff, DiCol coordinators, and other institutional researchers. The main goal is to ensure that LEXICC will provide a valuable, dynamic resource for Colombian Spanish exploration and comprehension.

Acknowledgments

This work is financed by the Instituto Caro y Cuervo (ICC), an institution adjunct to the Colombian Ministry of Cultures, under the project "ICC 2.0: Technological Integration for Linguistic Preservation and Diffusion Phase 2024". This project is being developed by the line of research in corpus and computational linguistics.

Endnotes

1. https://saliba.caroycuervo.gov.co/
2. https://www.idiomamedico.net/

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

3. https://lexicc.caroycuervo.gov.co/
4. GitLab is a web-based tool that helps developers manage their code, track issues, and automate the process of testing and deploying applications. It integrates various aspects of software development and operations (commonly referred to as DevOps) into a single platform. For more information, visit https://about.gitlab.com.
5. https://clicc.caroycuervo.gov.co/
6. https://alec.caroycuervo.gov.co/sig-alec.php

References

Academia Colombiana de la Lengua. 2012. *Breve diccionario de colombianismos.* Fourth Edition. Bogotá: Academia Colombiana de la Lengua.

Academia Nacional de Medicina de Colombia. 2023. *Diccionario Académico de Medicina*. https://www.idiomamedico.net/

Aggarwal, S. and J. Verma. 2018. Comparative Analysis of MEAN Stack and MERN Stack. *International Journal of Recent Research Aspects* 5(1): 133-137.

Almind, R. 2005. Designing Internet Dictionaries. *Hermes* 18(34): 37-54.

Andoque, H. and J. Landaburu. 2023. *Diccionario preliminar de la lengua del pueblo gente de Hacha-Andoque del Amazonas.* Bogotá.

Asociación de Academias de la Lengua Española (ASALE). 2010. *Diccionario de americanismos*. Madrid: Santillana.

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Bernal, J., W. López, D. Moreno and F. Mendieta. 2020. *Lexicografía electrónica especializada: el caso del diccionario académico de medicina — DIACME.* Bogotá: Instituto Caro y Cuervo.

Bonilla, J.E. and J. Bernal. 2020. Modelamiento de una base de datos espacial para el Atlas Lingüístico-Etnográfico de Colombia. *Revista signos* 53(103): 346-368.

Bonilla, J.E., R. Rubio, A. Llanos, D. Bejarano and J. Bernal. 2020. Proyecto de digitalización y nuevas perspectivas tecnológicas del "Atlas Lingüístico-Etnográfico de Colombia". Gallego, Á. and F. Roca (Coords.). *Dialectología digital del español*: 13-28. Santiago de Compostela: Universidad Santiago de Compostela, Servicio de Publicaciones = Servizo de Publicacións.

Cuervo, R.J. and Instituto Caro y Cuervo. 1998. *Diccionario de Construcción y Régimen de la Lengua Castellana*. Barcelona: Herder.

Dueñas, G. and D. Gómez. 2015. Diccionario electrónico sáliba–español: una herramienta interactiva para la documentación de la lengua y de la cultura sáliba. *Forma y Función* 28(2): 49-61.

Győrödi, C.A., D.V. Dumşe-Burescu, D.R. Zmaranda and R.Ş. Győrödi. 2022. A Comparative Study of MongoDB and Document-based MySQL for Big Data Application Data Management. *Big Data and Cognitive Computing* 6(2). https://doi.org/10.3390/bdcc6020049

Haensch, G. and R. Werner. 1993. *Nuevo diccionario de Americanismos. Nuevo diccionario de colombianismos. Vol. 5.* Bogotá: Instituto Caro y Cuervo.

IDM France. n.d. *DPS User Manual*. https://dps.cw.idm.fr/

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-1989 (Project)

Instituto Caro y Cuervo. 1983. *Atlas Lingüístico-Etnográfico de Colombia (ALEC)*. Bogotá: Instituto Caro y Cuervo.

Instituto Caro y Cuervo. 2014. *Diccionario bilingüe sáliba–español*. https://saliba.caroycuervo.gov.co/

Instituto Caro y Cuervo. 2018. *Diccionario de colombianismos*. Bogotá: Instituto Caro y Cuervo.

Instituto Nacional para Sordos (INSOR) and Instituto Caro y Cuervo. 2011. *Diccionario básico de la lengua de señas colombiana*. Bogotá: Imprenta Nacional de Colombia. http://www.insor.gov.co/descargar/diccionario_basico_completo.pdf

Institute of the Estonian Language. 2023. *EElex* (version 3.4) [software]. https://eelex.eki.ee/

Kruyt, T. 2003. Multifunctional Linguistic Databases: Their Multiple Use. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 194-203. Amsterdam: John Benjamins.

Leau, Y.B., W.K. Loo, W.Y. Tham and S.F. Tan. 2012. Software Development Life Cycle AGILE vs Traditional Approaches. *International Conference on Information and Network Technology* 37(1): 162-167.

Měchura, M. 2017. Introducing Lexonomy: An Open-Source Dictionary Writing and Publishing System. Kosem, I., C. Tiberius, M. Jakubíček, J. Kallas, S. Krek and V. Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century: Proceedings of eLex 2017 Conference, Leiden, The Netherlands, 19–21 September 2017*: 662-679. Brno, Czech Republic: Lexical Computing CZ s.r.o.

Montes, J.J., J. Figueroa, S. Mora and M. Lozano. 1986. *Glosario Lexicográfico del Atlas Lingüístico-Etnográfico de Colombia (ALEC)*. Bogotá: Instituto Caro y Cuervo.

Nieto, G. 2017. *Glosario de Aprendizaje del Español de Colombia.* Bogotá: Instituto Caro y Cuervo. https://spanishincolombia.caroycuervo.gov.co/documentos/imagenes/Hecho%20en%20Colombia-glosario.pdf

Nieto, G. 2020. *Corpus léxico del español de Colombia* (CorlexCo). https://clicc.caroycuervo.gov.co/corpus/CorlexCo [29 may 2024]

Rozo, N., M.B. Espejo, G. Duarte, D. Guevara and S. Lamprea. 2020. *Léxico de la Violencia en Colombia 1948–1970*. Bogotá: Instituto Caro y Cuervo.

Rubio, R. and J. Bernal. 2019. Corpus Oral del Instituto Caro y Cuervo: reestructuración, diseño y construcción. *Lexis* 43(1): 195-219.

Rubio, R., A. Luna and N. Solano. 2023. Corpus Lingüísticos del Instituto Caro y Cuervo (CLICC): una plataforma en línea para el almacenamiento, sistematización y consulta de corpus. *Linguamática* 15(2): 89-96. https://doi.org/10.21814/lm.15.2.407

Saundariya, K., M. Abirami, K.R. Senthil, D. Prabakaran, B. Srimathi and G. Nagarajan. 2021. Webapp Service for Booking Handyman Using MongoDB, express JS, React JS, node JS. *2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 13–14 May 2021*: 180-183. Danvers, MA: Institute of Electrical and Electronics Engineers (IEEE).

Sommerville, I. 2011. *Ingeniería de Software*. Ninth edition. Mexico City: Pearson Education.

TshwaneDJe HLT. 2023. *TshwaneLex Suite: Dictionary Compilation Software* (version 15). http://tshwanedje.com/tshwanelex/

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Digital Lexicography in ESP: Enhancing Vocabulary Acquisition and Reading Comprehension in Students of Commerce Degree

Beatriz Chaves-Yuste, *Department of English Studies, Complutense University of Madrid, Madrid, Spain (bchave01@ucm.es) (https://orcid.org/0000-0002-4442-9138)*
and
Manuel Márquez-Cruz, *Department of Linguistics, Arabic, Hebrew, Basque and East Asian Studies, Complutense University of Madrid, Madrid, Spain (manmarqu@ucm.es) (https://orcid.org/0000-0001-9040-169X)*

Abstract: This study investigates the impact of Sketch Engine and Lexonomy on vocabulary acquisition and reading comprehension in English for Specific Purposes (ESP), focusing on international commerce among undergraduate students. Conducted over thirteen weeks, the research involved 90 first-year Commerce students from a public university in Madrid. A pre-experimental design was implemented, dividing participants into control and experimental groups. The control group used traditional analogue dictionaries, while the experimental group utilized digital tools to compile specialized dictionaries. Performance was assessed through pre- and post-intervention reading comprehension tests and a questionnaire assessed the use of lexicographical resources by ESP undergraduate students. On the one hand, results revealed that the use of these two lexicographical digital tools implies significant improvements regarding vocabulary acquisition and reading comprehension. On the other hand, digital tools emerged as the preferred resources, emphasizing translation, contextual meaning, and voice search functionalities, though challenges such as navigation difficulties and interpreting phonetic transcriptions were noted. The findings underscore the pedagogical potential of digital lexicography to enhance ESP instruction, foster autonomous learning, and equip students with essential professional skills. This research provides a foundation for integrating innovative methodologies into higher education, contributing to more effective and equitable teaching practices.

Keywords: SKETCH ENGINE, LEXONOMY, DIDACTIC LEXICOGRAPHY, ENGLISH FOR SPECIFIC PURPOSES, READING COMPREHENSION, VOCABULARY ACQUISITION

Opsomming: Digitale leksikografie in ESD: Die verbetering van woordeskatverwerwing en leesbegrip by handelstudente. In hierdie studie word die impak van Sketch Engine en Lexonomy op woordeskatverwerwing en leesbegrip in Engels vir Spesifieke

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Doeleindes (ESD) ondersoek. Daar word spesifiek gefokus op internasionale handel onder voorgraadse studente. Die navorsing wat oor dertien weke uitgevoer is, het 90 eerstejaar-handelstudente van 'n openbare universiteit in Madrid betrek. 'n Pre-eksperimentele ontwerp wat deelnemers in kontrole- en eksperimentele groepe verdeel, is geïmplementeer. Die kontrolegroep het tradisionele analoogwoordeboeke gebruik, terwyl die eksperimentele groep digitale hulpmiddels gebruik het om gespesialiseerde woordeboeke saam te stel. Prestasie is aan die hand van pre- en postintervensieleesbegripstoetse beoordeel en met behulp van 'n vraelys is die gebruik van leksikografiese hulpbronne deur ESD-voorgraadse studente beoordeel. Aan die een kant het die resultate aangedui dat die gebruik van hierdie twee leksikografiese digitale hulpmiddels beduidende verbeterings rakende woordeskatverwerwing en leesbegrip inhou. Aan die ander kant het digitale hulpmiddels na vore getree as die voorkeurhulpbronne met die fokus op vertaling, kontekstuele betekenis, en stemsoektogfunksies. Uitdagings soos navigasieprobleme en die interpretasie van fonetiese transkripsies is egter ook aangeteken. Die bevindings beklemtoon die opvoedkundige potensiaal van die digitale leksikografie om ESD-onderrig te verbeter, outonome leer te bevorder en studente met noodsaaklike professionele vaardighede toe te rus. Hierdie navorsing vorm 'n basis vir die integrering van innovorende metodologieë in die hoër onderwys, wat sodoende bydra tot meer effektiewe en billike onderrigpraktyke.

Sleutelwoorde: SKETCH ENGINE, LEXONOMY, DIDAKTIESE LEKSIKOGRAFIE, ENGELS VIR SPESIFIEKE DOELEINDES, LEESBEGRIP, WOORDESKATVERWERWING

1. Introduction

Lexicography is a millennia-old discipline which, in its practical aspect, has been focused on the compilation of comprehensive or partial vocabularies of one or several languages. In its theoretical aspect, it has focused on the study of the construction of these vocabularies throughout history from a prescriptive, descriptive and historical perspective, or on the methodologies used in the compilation of such works. In this context, the dictionary has emerged as the flagship product, though early lexicographic manifestations in Europe and Asia often took the form of word lists or glosses added to specific works (cf. Tarp and Gouws 2023: 429-430). These glosses facilitated the comprehension of terms that might prove complex for a community, due to the antiquity of the vocabulary, or its technical, religious, scientific, or literary nature, among others (Boisson et al. 1991: 261-262; Béjoint 2016: 8). Such was the case with the Homeric glosses written by ancient scribes in the works of Homer (Tarp 2019: 11), or with the *Cangjiepian*, a textbook dated around 220 B.C. used to teach children Chinese writing, which could only be read with the help of specialized annotations (Yong and Peng 2008: 16). These glosses and annotations were later compiled into independent books, whose diffusion reached significant levels thanks to the technological revolution of the printing press, and more recently, through their publication in digital format.

From the points outlined above, it is evident that one of the primary purposes of a language dictionary — among others — is pedagogical. This pedagogical function is even more apparent in dictionaries specifically designed to

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

enhance the process of language learning, whether for the native language or a second or foreign language (Márquez Cruz 2024: 64). This field is known as pedagogical lexicography, and the works derived from its practice are typically categorized, depending on the type of user, into school dictionaries and learner's dictionaries. These works aim to meet the cognitive and communicative needs of learners, either of their native language (school dictionaries) or of a second or foreign language (learner's dictionaries). The lexicographic structures of both types of dictionaries are determined by the quantity, quality, and organization of data provided in the paratexts and nomenclature, in the headwords (macrostructure), in the lexicographic entries (microstructure), in the relationships between different parts and elements of the dictionary (mediostructure), and in the images and/or multimedia elements (iconostructure; cf. Camacho 2017: 2). Additionally, the vocabulary used in the definitions and the typography of the textual conglomerate that makes up the lexicographic article should also be considered. Regarding the type of information provided in these dictionary structures, Heuberger (2016: 30-40) identified a series of features that should be present in learner's dictionaries, which we also believe should be part of school dictionaries: (i) the use of controlled defining vocabularies, (ii) definitions, (iii) examples, (iv) grammatical information, (v) usage information (pragmatics), (vi) collocations, (vii) pronunciation, (viii) data accessibility, (ix) preliminary texts, and (x) appendices (final materials).

Regarding school dictionaries, it is crucial to examine the outer texts, both the preliminary and concluding sections of the dictionary in conjunction with its macrostructure, microstructure, mediostructure, iconostructure, and target users. Many of these dictionaries are published — in analogue format — in lightweight, manageable volumes (especially when the target user group is primary or secondary school students). They use italics, different font sizes, and various colours to differentiate content within the dictionary entries. They also feature a selective and appropriate lexicon for the school vocabulary, including non-standard terms, such as some neologisms. In addition, they employ controlled vocabularies (between one thousand and two thousand words) to ensure that the definitions are simple and provide ample examples, as well as illustrations, grammatical indications, synonym and antonym relationships, lexical families, semantic fields, idiomatic expressions, and collocations. Finally, they usually contain grammatical appendices, and the prefaces often provide guidance on dictionary usage, including explanations of the various components of dictionary entries based on word type, abbreviations, etc. for users who may not be familiar with them.

Learner's dictionaries differ from school dictionaries — those created for language learning by native speakers — in that they tend to be slightly larger in size. They use different colours and font sizes to distinguish the various types of data provided in dictionary entries. The order of definitions generally follows usage frequency, with detailed descriptions of phraseology, idioms, and collocations, and synonyms and antonyms. This information is followed by the use of descriptors and guide words, syntagmatic or sentence-based definitions through

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

equivalence or examples (Climent 2005: 363-366), phonetic information, grammatical information, usage labels, grammatical appendices, and annexes with varied information on culture, economics, geography, or encyclopaedic content (currencies, weights, places, famous figures, etc.). The iconostructure is usually more extensive in bilingual dictionaries than in monolingual ones in order to visually convey the relationship between the target language and the mother tongue.

The described features provide a concise theoretical framework on the essence of the dictionary models used in language learning classrooms. This may seem unusual given the learning context of the present study, which focuses on English for Specific Purposes (henceforth ESP), specifically English in the field of Commerce. We should argue that the lexicographic proposal we present should be underpinned by theoretical principles and practical developments specific to specialized lexicography, terminology, and terminography. It is true that the creation of a commerce dictionary should adhere to this branch of lexicography and that, therefore, we would need to describe a theoretical framework more aligned with specialized dictionaries. However, the primary objective of our lexicographic proposal is to enhance the acquisition of grammatical and lexical knowledge of the English language, using commerce as a thematic area, but without intending to create a dictionary whose purpose is to provide definitions of commerce-specific terms. In other words, the goal is not to create a dictionary of commerce in English, but to improve linguistic knowledge of English, especially, the reading comprehension, through the lexicographic coding of lexical and grammatical structures that are characteristic of, but not exclusive to, the English language in the context of commerce. For this purpose, first-year Commerce students at a public university in Madrid developed a monolingual English dictionary, with its nomenclature based on the compilation and analysis of a scientific corpus focused on international commerce.

2. Corpus Linguistics

Of the many different definitions of Corpus Linguistics (hereafter, CL), one of the most current and, in our opinion, one that clearly and precisely captures the essence of this very eclectic discipline is the definition provided by Rojo (2021: 48). He (ibid.) described it as an approach to the study of linguistic phenomena with an empirical orientation, based on the detailed analysis of large amounts of data (corpora). As Rojo (2021: 48) argued, CL is characterized by its empirical nature, deriving data from naturally produced texts collected in the form of text corpora. In this task, the use of computers is essential to store and manage the corpora, employing both qualitative and quantitative analytical techniques, and conducting systematic and exhaustive analyses of the data provided by corpora. It is worth noting that working with computers has not always been a must for corpus analysis, as Shaxrizoda (2024: 104-105) pointed out, given that corpus-based research existed prior to the use of computers for analysing linguistic data.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Regarding the concept of a corpus, it is an organized collection of oral and/or written texts in a language, compiled to exemplify its use. The coding of the corpus should adhere to a standard that allows for the retrieval of information using different systems based on various linguistic criteria. Therefore, when compiling a corpus, it is essential to consider the language(s) or variety(ies) to ensure representativeness, as well as the text's mode (oral or written), type (e.g., book, magazine, advertisement, etc.), domain (academic or popular), location, and date (Sinclair 2005: 8). The size of the corpus will depend on several factors, including its intended use; however, whenever possible, it should consist of complete documents to avoid any bias.

Corpus-based research follows two main approaches: starting from corpus data to develop theories (corpus-driven) or consulting corpora to verify theories (corpus-based). In either case, computational tools are essential for compiling and analysing corpora automatically or semi-automatically, reducing the noise caused by irrelevant or questionable-quality information in the corpus (Costa et al. 2015: 74). Tools like Babouk, BootCat, Corpógrafo, and Terminus (Toledo and Camero 2021: 20) are used for terminological extraction and corpus compilation and analysis. Among these, BootCat stands out for extracting large amounts of digital textual material using seed words representative of a specific semantic field. BootCat randomly combines these seed words to generate sequences called tuples, which are then used as search strings in Google to retrieve web pages. The retrieved HTML documents are cleaned and converted into a plain text (Flores 2022: 174). An extension of BootCat is WebBootCat, a tool based on the seed words technique used by Sketch Engine.

3. Integrating corpus compilation and digital lexicography in ESP vocabulary acquisition

With the aim of enhancing the effectiveness of the vocabulary acquisition process in the domain of ESP — specifically, Business English — through lexicographic tasks, two tools were made available to the students: Sketch Engine (https://www.sketchengine.eu/) and Lexonomy (https://www.lexonomy.eu/). The former was employed to compile and manage a corpus that served as the foundational resource for the dictionary developed by the students. The latter was used, first, to design the macrostructure and microstructure of the dictionary and, subsequently, to input lexicographic data, publish the entries, and share them among all participants.

The corpus was compiled from nine business-themed English texts. The Word Sketch tool was used to explore collocations and word combinations; Wordlist, to identify the most frequent lexical items; Keywords, to extract core terminology specific to the commercial domain; and Concordances, to analyse the contextual usage of high-frequency keywords and obtain authentic usage examples.

Regarding Lexonomy, the microstructure of the dictionary entries was

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

designed according to the following core lexicographic information: lemma, part of speech, British and American English pronunciation, definition, and examples of use.

The theoretical principles of corpus linguistics and the use of corpus compilers and analysers, as well as Dictionary Writing Systems,[1] have already been successfully implemented in various teaching practices aimed at promoting incidental lexical learning in foreign languages classrooms. Chaves-Yuste and Márquez-Cruz (2024: 183; 2022: 364) demonstrated that Lexonomy, used to create a cooperative digital dictionary, helped to improve reading comprehension, vocabulary and linguistic competence of secondary education students. They demonstrated that Lexonomy was an efficient tool in a b-learning context and facilitated vocabulary acquisition. However, in any of these studies they used it in combination with Sketch Engine to compile the corpus they were working with.

Srdanović (2020: 1059-1060) implemented a classroom project in the field of Tourism in which instructors and students collaboratively utilized Sketch Engine and Lexonomy to create a bilingual Japanese-Croatian dictionary. The study showed two main findings: (i) the use of these advanced tools fosters students' proficiency in leveraging technologies for linguistic analysis and lexicographic resource creation; (ii) the pressing need to train professionals capable of applying Japanese language skills within the tourism sector (Mavrić et al. 2023: 51). Furthermore, the benefits of these tools extend beyond language acquisition. They impart crucial knowledge and skills to dictionary users, enabling them to maximize the utility of these resources. As argued by Caruso (2024: 139), when students engage in the process of creating dictionaries in one or more second languages, they achieve multifaceted gains. By compiling corpora using Sketch Engine and designing dictionary structures with Lexonomy, students not only enhance their passive and active knowledge of the target language but also deepen their understanding of the microstructural organization of lexicographic entries in learner dictionaries.

However, it is essential to acknowledge that these tools comprise certain limitations, and their implementation should be carefully evaluated. First, Sketch Engine requires the purchase of a license, which may limit its accessibility for some educational institutions. Second, while Lexonomy is freely available, it has constraints concerning data storage capacity, which could impact the scalability of certain projects (Kovalevskaitė et al. 2020: 246). Thus, the use of Sketch Engine and Lexonomy presents a valuable opportunity to integrate technology into linguistic and lexicographic education, equipping students with practical and theoretical expertise. Nonetheless, educators must weigh the tools' benefits against their financial and technical limitations to ensure optimal application in educational contexts. Hence, this work aims to explore the benefits of the compilation of a digital learner's dictionary using Sketch Engine and Lexonomy in students' acquisition of commerce vocabulary. The objective is twofold: On the one hand, to measure the effect of the compilation of a corpus and a dictionary on students' acquisition of specific vocabulary of commerce (especially, inter-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

national negotiation) which helps students develop their reading comprehension of business and trade texts, and on the other hand, to know the real use of the available lexicographic tools for ESP students who need to foster English for international commerce firstly, for academic purposes, and secondly, for their future professional endeavours. Thus, the research questions raised are:

1. To what extent do the corpus compiler software Sketch Engine and the Dictionary Writing System Lexonomy influence commerce students' acquisition of specific commerce vocabulary and development of their reading comprehension?
2. Which is the real use of commerce students of the available lexicographic tools in their English for commerce learning process?

Based on the scientific literature and the proposed research questions, the following hypotheses are formulated:

— Hypothesis 1 (H1): The use of the corpus compiler software Sketch Engine and the Dictionary Writing System Lexonomy positively impact on the acquisition of commerce vocabulary and on the development of reading comprehension by commerce undergraduate students.
— Hypothesis 2 (H2): Students use more digital than analogue lexicographic tools in the English for commerce learning process.

This study aims to provide a preliminary contribution to the scientific understanding of vocabulary acquisition through the creation and use of lexicographic tools by students. We believe that through an incidental, intentional, and lifelong learning experience, students can acquire the target specialized vocabulary necessary to develop their professional skills successfully.

4. Design

This research follows a quantitative approach, gathering data through a quasi-experimental design. On the one hand, it collects quantitative data by measuring academic performance before and after an intervention in which the creation of either an analogue or digital dictionary was employed in two groups of university students randomly assigned as a control group and an experimental group. On the other hand, it also gathers quantitative data concerning the use and evaluation of lexicographic tools, particularly dictionaries, for the learning of English for professional purposes. The research spanned thirteen weeks, from February 6 to May 10 of the 2023–2024 academic year.

Initially, the academic scores obtained in the PRE reading comprehension test in the field of commerce were collected, administered to all the students before the intervention began. After thirteen weeks of instruction a similar test was administered (POST), providing a text which gathered some of the terms collected in the different dictionaries. During the instruction period, the stu-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

dents of the experimental group worked in heterogeneous cooperative teams and compiled and used dictionaries using Sketch Engine and Lexonomy. Meanwhile, the control group worked with analogue traditional glossaries. The dependent variable in this study is the acquisition of a specialized professional lexicon, measured through the scores of these two tests (PRE and POST), administered before and after compiling the various dictionaries in both the control group and the experimental group. The independent variable is the tool used and the procedure followed, specifically the use of digital lexicographic tools such as Sketch Engine and Lexonomy.

4.1 Participants

The sample was selected based on accessibility, using a non-probabilistic purposive sampling of 90 first-year students enrolled in the Commerce degree at a public university of Madrid. These students were recruited through purposing samplings since one of the researchers was the instructor of two pre-existing groups (the control group and the experimental group) which were naturally split up. The criteria established for selecting the sample were: being enrolled in the first year of the Commerce degree, attending classes regularly, completing all the tests and providing signed consent to participate in the research. Of all participants, 54.4% were male (n=49) and 45.6% were female (n=41). The average age of the participants was 19.58 years, with a standard deviation (SD) of 1.83 years. None of the students had prior experience using Sketch Engine or Lexonomy. However, all participants demonstrated familiarity with technological devices and exhibited a moderate level of digital competence, particularly when operating their mobile phones.

4.2 Instruments

In order to assess the usefulness of these digital lexicographic tools in the development of our students' professional skills, two custom-designed tests were used for data collection. The first test was administered in February (February 6), prior to the experimental use of digital lexicographic tools in the experimental group, and the second test in June (June 19), after the classes had ended. The second test was administered after the conclusion of the semester and at the onset of the examination period. During this five-week interval, students were assessed in various subjects within their degree program, completed internships at local companies, and presented their end-of-year assignments. Therefore, despite the time gap, no significant external factors are considered to have influenced the results obtained in the post-test. Both tests were two reading comprehension exercises, part four of the official Business English Certificate (BEC) Vantage (level B2, according to the Common European Framework of Reference for Languages, Council of Europe, 2001) test (University of Cambridge ESOL

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Examinations 2015). Students needed to read a text and choose the best word to fill each gap (fifteen multiple choice items with four options each). All students took the tests, which were scored from 0 to 10 (0.66 points each correct answer), with the following scale: 0-4.99 = fail, 5-6.99 = pass, 7-8.99 = good, and 9-10 = excellent. After the intervention, a questionnaire, based on Knežević et al.'s (2021), containing 51 Likert scale questions was administered to assess students' evaluations regarding the actual use of lexicographic tools, particularly dictionaries, in the learning of ESP. The content validity of this questionnaire was analysed by displaying the content of this instrument to a panel of four experts in lexicography and foreign language instruction. They validated its content by guaranteeing that the items addressed all related aspects of the objectives of the study (Roebianto et al. 2023) and they agreed on their coherence and relevance. The KMO value of this questionnaire was 0.641 and the Bartlett's Test of Sphericity value was significant (p=0.001), so factor analysis was appropriate. The first fifteen components explained 78.126% of the accumulated variance, which is a valid and acceptable percentage since it exceeds the 60% threshold. This implies that the questionnaire captured a significant amount of information and that the initial eigenvalues of the extracted squared loadings corroborated that the retained components of the questionnaire were relevant. Its reliability was assessed using the rank-biserial correlation (r) with a reliability coefficient of 0.575, which indicated a moderate effect size.

4.3 Procedure

The study was first introduced to the students, and informed consent was obtained, allowing them the option to withdraw from the study at any time. The didactic experience was then conducted from February through mid-May, with a total of 32 hours of instruction, scheduled from 1:00 p.m. to 3:00 p.m. To avoid bias, the same instructor taught both the control and experimental groups. The tests were administered in the regular classroom under proper conditions of sound, lighting, and ventilation. The participation rate was 100%, meaning that all 90 students who took the PRE test also completed the POST test and the questionnaire on dictionary usage. The data collected were treated following the ethical guidelines of the Declaration of Helsinki, ensuring confidentiality.

4.4 Data analysis

Quantitative data, both for descriptive and inferential analysis, were examined using SPSS software, version 29 for Windows (IBM 2016). First, the normality of the variables was checked using the Kolmogorov-Smirnov test, considering the sample size, which indicated that neither test met the normality criteria (PRE p=0.001; POST p=0.001). Second, descriptive statistics such as mean and standard deviation were used for each variable, and non-parametric inferential

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

statistics such as the Mann-Whitney U test (for independent samples) were employed to compare the PRE and POST test results in the control and experimental groups, thus verifying the research objective and hypothesis. Third, the PRE and POST test results were compared within the two groups employing the Wilcoxon test. Fourth, the effect size of the questionnaire was calculated and interpreted using ranks-biserial correlation (r), and its control validity was explained through a factor analysis. All analyses were conducted with a significance level of $p < 0.05$. Finally, we did a descriptive analysis of the means and standard deviation obtained in the different items of the questionnaire to analyse the answers provided by students.

5. Results

The findings presented here suggest a potential causal link between the use of teaching dictionaries and the enhancement of specialized lexicon knowledge among students of Commerce. However, it is important to interpret these results with caution, acknowledging possible limitations and alternative explanations. The descriptive analysis is presented in Table 1, showing the mean and standard deviation of both tests, as well as the inferential analysis results that address the research objectives and hypotheses. In the PRE test, the average scores in both groups exhibited comparable levels of lexicon knowledge, with no significant differences (p=0.35; r=0.13). However, in the POST test, the scores of the control group were significantly lower than the average scores obtained by the experimental group, with a difference of 4.68 points. The Mann-Whitney U test indicated significant differences between the scores (p=0.001; r=0.49), which implies that the experimental group benefited more from the intervention, namely, the creation and use of digital learner's dictionaries created by the students. While these results appear to support the alternative hypothesis and reject the null hypothesis, we should be cautious when attributing the improvement solely to the intervention since other factors, such as their individual learning preferences, may have influenced the intervention. Notwithstanding, the findings suggest that the use of digital learner's dictionaries may have contributed to improve lexicon acquisition and professional skills development in the field of international trade.

Table 1: Descriptive results of control and experimental groups

	Control group		Experimental group	
Variables	M	SD	M	SD
PRE test	7.78	1.63	7.62	1.50
POST test	4.36	2.72	9.04	1.41

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

These results corroborate the effectiveness of creating and utilizing digital lexicographic tools to succeed in using English in professional contexts with sufficient fluency and accuracy, enabling students to perform effectively in the labour activities demanded by the commercial sector.

To complement the analyses and address the effect of confounding variables such as gender or age, the non-parametric Mann-Whitney U test was employed. As observed in Table 2, neither the age nor the gender significantly influenced the academic performance in relation to dictionary use. This strengthens the claim that the observed differences were likely due to the intervention. Both men and women do not differ in academic performance in creating and using analogic dictionaries or digital dictionaries.

Table 2: Complementary inferential results by gender and age

Variables	Genre		Age	
	U	*P*	*U*	*P*
Analogue dictionaries	2.81	0.356	2.38	0.919
Sketch Engine + Lexonomy digital dictionary	2.82	0.207	2.02	0.553

Finally, an additional quantitative study was also conducted to analyse the actual use of dictionaries in students' learning process of English for commerce. The high reliability of the administered questionnaire (Cronbach's alpha= 0.946) adds credibility to these findings. This high score confirms that the instrument is both valid and reliable for the purposes of this study. A descriptive analysis was subsequently conducted, focusing on the mean and standard deviation of the variables under study. Particular attention was given to the types of dictionaries used, the reasons for their consultation, the nature of the queries, and the difficulties encountered which reveal nuanced patterns in dictionary usage.

As illustrated in the accompanying figure (Figure 1), digital dictionaries, related applications and Spanish–English bilingual dictionaries emerged as the most frequently used resources, which reflect broader trends in technology adoption rather than the unique advantage of digital learner's dictionaries. Conversely, paper dictionaries, monolingual dictionaries, and specialized bilingual dictionaries for professional purposes were the least utilized. For a better interpretation of the figures, the most frequent answers have been presented in green whereas the most infrequent ones are coloured in red.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

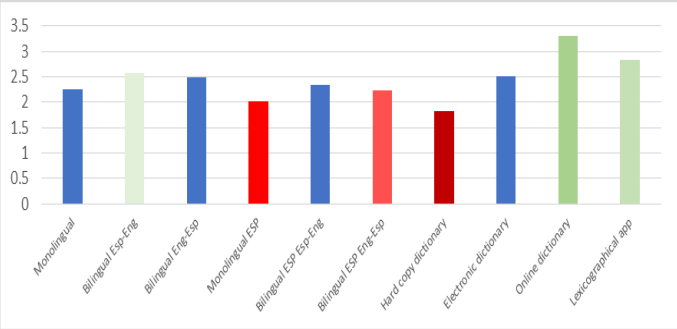


Figure 1: Types of dictionaries

Regarding the reasons for dictionary use displayed in Figure 2, the most common motivations included determining the meaning of unfamiliar words, completing activities that explicitly required dictionary use, and verifying the accuracy of an inferred meaning. However, it is infrequent to use dictionaries for speaking and listening tasks in English or for improving pronunciation skills. This suggests that their impact may be domain-specific rather than comprehensive.

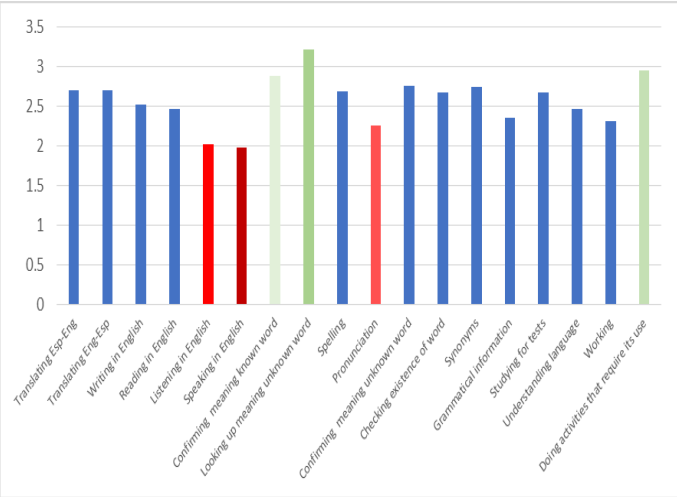


Figure 2: Reasons for dictionary use

As can be observed in Figure 3, in terms of the type of information sought, the most frequent queries were related to accessing the meaning of terms within specific contexts, finding equivalent translations, and using voice search functions. Features such as appendices, images, and usage guides were notably less popular among users.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

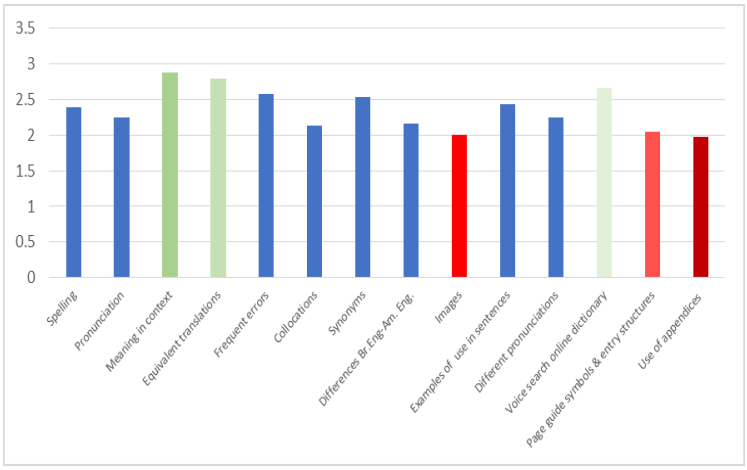


Figure 3: Types of information sought

Finally, regarding the primary difficulties reported in using dictionaries summarized in Figure 4, participants included challenges in navigating the tools, misunderstanding the information provided, and the need to consult multiple references depending on specific needs. These difficulties highlight areas for improvement in integrating digital dictionaries into pedagogical strategies. The least frequent issues included unfamiliarity with abbreviations, difficulty assessing the quality of a dictionary, and challenges in interpreting phonetic transcriptions.

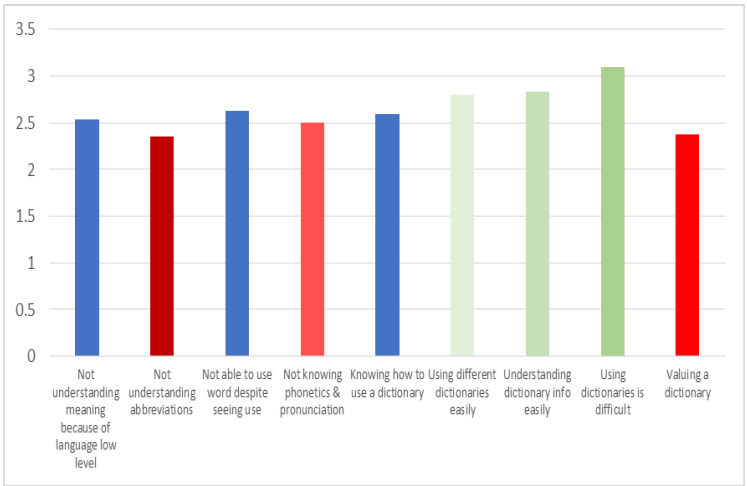


Figure 4: Difficulties encountered when using a dictionary

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

6. Discussion

This study investigates whether incorporating digital lexicographic tools, specifically the corpora compiler Sketch Engine and the DWS Lexonomy, enhances vocabulary acquisition and reading comprehension of scientific texts in the field of international commerce among undergraduate students studying English for Specific Purposes . Conducted in a face-to-face educational context, the research involved 90 students randomly assigned to either a control or an experimental group. Over an equal duration of instruction, the control group relied on traditional teaching methods, including the creation and use of analogue paper dictionaries or glossaries, while the experimental group utilized the aforementioned digital lexicographic tools. To assess the impact of these instructional approaches, participants completed two standardized reading comprehension tests (BEC Vantage, Cambridge Exam). These tests were based on business and international commerce texts, incorporating the vocabulary featured in their respective dictionaries. The findings indicate significant improvements in the experimental group's reading comprehension, attributed to their enhanced vocabulary acquisition. The target vocabulary was acquired through both incidental and intentional learning approaches. As a result, their reading comprehension improved substantially, particularly in the use of specialized vocabulary — an essential skill for success in their professional careers within the commercial sector. Moreover, the practical component of the study enabled students to not only create their own dictionaries but also to navigate and utilize lexicographic resources with proficiency, collaboratively managing entries produced by their peers. These results validate Hypothesis 1, confirming that the integration of Sketch Engine and Lexonomy in ESP instruction fosters vocabulary development and, consequently, reading comprehension skills. The study underscores the potential for these tools to optimize ESP teaching and learning at the university level and in other educational contexts, such as Vocational Education and Training (VET), where specialized English proficiency is essential for professional success. To ensure that gender and age did not bias the results, an inferential analysis was conducted, revealing no significant differences between the groups. These findings strengthen the reliability of the study's conclusions, supporting the idea that digital lexicographic resources can be broadly implemented without demographic limitations. Descriptive analyses reveal a compelling need to optimize university students' academic performance through a balanced approach combining traditional teaching and digital resources. These results align with Bartolomé-Díaz and Frontini (2020: 67), who examined the use of Sketch Engine and Lexonomy in creating a bilingual French–Spanish architectural dictionary. Their findings highlighted the innovative potential of digital lexicography while emphasizing the need for further development to achieve full operational capability. This study supports prior research emphasizing the critical role of vocabulary acquisition in enhancing academic performance. For instance, Shahjahan et al. (2021: 1145) argued that well-prepared graduates exhibit stronger academic and professional competen-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

cies, while Atik and Çelik (2021: 124-125) demonstrated that academic achievement yields broader personal and societal benefits, including motivation, engagement, and life-long learning. To this end, universities must prioritize pedagogical innovation, as suggested by Van Laar (Van Laar et al. 2020: 10), by integrating active methodologies and digital tools into the curriculum. Such practices position students as active participants in their learning processes, fostering motivation and engagement (Snezhko et al. 2022: 88). The study further validates Hypothesis 2, revealing that students favour digital lexicographic tools over analogue methods in their ESP learning. This finding aligns with contemporary research (such as Alshammary and Alhalafawy 2022: 2294; Jia et al. 2024: 13), which highlights the efficacy of digital tools in enhancing vocabulary acquisition and academic performance. Digital resources emphasize active, student-centred learning, personalized instruction, and innovation (Strelan et al. 2020: 16). The increasing adoption of digital ecosystems in higher education, as advocated by Nguyen and Tuamsuk (2022: 2), necessitates the alignment of teaching practices with frameworks such as DigCompEdu (Redecker and Punie 2017: 20). These frameworks emphasize the development of educators' digital competence and provide practical tools to integrate digital strategies effectively. For students, the use of such resources not only enhances learning outcomes but also equips them with essential 21st-century skills, promoting lifelong learning (Molnár et al. 2022: 274).

This study has certain limitations. The sample size restricts the generalizability of findings, necessitating cautious interpretation. Additionally, while efforts were made to control for sociodemographic variables, participants were selected based on convenience and accessibility rather than through randomized sampling. Despite these constraints, the study provides valuable insights into the pedagogical application of digital lexicographic tools, offering a foundation for future research. Future studies could extend these findings to online university settings, other ESP contexts, and additional languages or disciplines. Longitudinal research would also be valuable, providing a deeper understanding of the long-term impacts of digital lexicography on vocabulary acquisition and reading comprehension. Expanding the scope of analysis to include other variables, such as professional performance, could further elucidate the broader implications of this approach. Investigating the relationship between digital resource use and other professional competencies, such as oral communication and writing skills, could further clarify their role in holistic language development. By building on the current findings and addressing existing limitations, future studies can continue to advance the integration of digital lexicography in ESP instruction, fostering more effective and equitable educational practices in a rapidly evolving academic and professional landscape.

7. Conclusion

This study demonstrates the potential of digital lexicographic tools, such as Sketch Engine and Lexonomy, to revolutionize vocabulary acquisition and reading

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

comprehension in ESP education. By placing students at the centre of the learning process and equipping them with specialized language skills, these tools prepare undergraduates to meet the demands of their professional fields. This research provides educators with a roadmap for enhancing both student learning outcomes and professional preparedness. The results emphasize the necessity of adopting innovative teaching methodologies that leverage digital resources. Such tools not only optimize students' linguistic competence but also foster their ability to meet the demands of professional environments requiring specialized English skills. Specifically, the creation and use of digital learning dictionaries empower students to develop autonomy in learning, encouraging active participation and critical engagement with language content. This pedagogical approach aligns with contemporary trends in higher education, which prioritize student-centred and technology-enhanced learning. Furthermore, the study reveals the importance of addressing common challenges faced by students in using digital dictionaries, such as navigation difficulties and the need for targeted guidance on interpreting lexicographic data. These insights suggest that educators should incorporate explicit instruction on the effective use of such tools, tailoring support to diverse learning styles and levels of digital literacy. The findings also highlight the role of dictionary use in promoting incidental and intentional vocabulary learning. By integrating tasks that require students to infer word meanings, verify accuracy, and apply vocabulary in context, educators can create a learning environment that mirrors real-world professional scenarios. Such practices not only strengthen students' lexical competence but also enhance their confidence in using English for specialized purposes.

The significant improvements observed in the experimental group's reading comprehension scores provide robust evidence that the use of Sketch Engine and Lexonomy supports the development of key linguistic and professional skills. These results validate the hypothesis that digital dictionaries are instrumental in equipping students with the fluency and accuracy required for success in the commercial sector. The descriptive and inferential analyses further illustrate the prevalence of digital dictionaries over traditional analogue resources in the learning process. Students reported a strong preference for tools that facilitate contextual understanding, translation, and voice search functions, reflecting a shift towards more dynamic, interactive modes of language learning. However, challenges such as tool navigation, and the interpretation of phonetic transcriptions indicate areas where additional instructional support is needed. Importantly, the findings reveal no significant differences in outcomes based on gender or age, suggesting that the benefits of digital lexicography are broadly applicable across diverse student demographics. This universality underscores the potential of these tools to enhance educational equity and inclusivity in ESP instruction.

Despite the positive results of this intervention, further studies are recommended to confirm these results across diverse educational contexts. In addition, additional factors influencing the effectiveness of digital lexicographic tools,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

such as the students' different learning styles or affective factors that may influence their learning process could also be analysed. Moreover, apart from the quantitative methodology employed in this study, further studies would also benefit from qualitative methods, such as semi-structured interviews with students or open-ended surveys. They would provide new insights about how students experience and struggle with digital lexicographic tools and which challenges need to be met. Nevertheless, these results provide strong evidence supporting the efficacy of digital learner's dictionaries in enhancing specialized lexicon knowledge taking into account the domain-specific limitations of this study.

As digital innovation continues to reshape higher education, integrating such resources represents a critical step toward fostering autonomous, meaningful learning experiences that align with the needs of contemporary society. Furthermore, the administered questionnaire provided valuable insights about the real use of lexicographic tools by ESP Spanish undergraduate students, enabling targeted interventions to address these gaps in classroom settings and thereby enhance students' learning processes.

Endnote

1. These systems are referred to by various names. Tiberius and Heylen (2022) document terms such as *dictionary editing systems*, *dictionary compilation software*, *lexicography software or dictionary production software*, *lexicographic workbench*, *dictionary management system or lexicographer's workbench*, *dictionary editing tool*, *dictionary building software*, or simply *editorial system*. Similarly, Rubio et al. (2021: 342) highlighted additional terminology found in the scientific literature: *dictionary building platform* (Mangeot and Chalvin 2006: 1666), *dictionary content management system* (Alegría et al. 2006: 2), *dictionary editor and browser* or *DEB platform* (Horák and Rambousek 2007: 129), and *dictionary writing and publishing systems* (Měchura 2017: 1).

References

Alegría, Iñaki, Xabier Arregi, Xabier Artola, Mikel Astiz and Leonel Ruiz Miyares. 2006. A Dictionary Content Management System. Corino, Elisa, Carla Marello and Cristina Onesti (Eds.). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Turin, Italy, 6–9 September 2006. Vol. 1:* 105-109. Alessandria: Edizioni dell'Orso.

Alshammary, F.M. and W.S. Alhalafawy. 2022. Sustaining Enhancement of Learning Outcomes across Digital Platforms during the COVID-19 Pandemic: A Systematic Review. *Journal of Positive School Psychology* 6(9): 2279-2301. https://journalppw.com/index.php/jpsp/article/view/12650

Atik, S. and O.T. Çelik. 2021. Analysis of the Relationships between Academic Motivation, Engagement, Burnout and Academic Achievement with Structural Equation Modelling. *International Journal of Contemporary Educational Research* 8(2): 118-130. https://doi.org/10.33200/ijcer.826088

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Bartolomé-Díaz, Zaida and Francesca Frontini. 2020. Building a Domain-specific Bilingual Lexicon Resource with Sketchengine and Lexonomy: Taking Ownership of the Issues. Kernerman, Ilan, Simon Krek, John P. McCrae, Jorge Gracia, Sina Ahmadi and Besim Kabashi (Eds.). 2020. *Proceedings of the 2020 Globalex Workshop on Linked Lexicography, Marseille, France, 11–16 May 2020*: 62-68. Paris: European Language Resources Association.

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.

Boisson, C., P. Kirtchuk and H. Béjoint. 1991. Aux origines de la lexicographie: les premiers dictionnaires monolingues et bilingues. *International Journal of Lexicography* 4(4): 261-315. https://doi.org/10.1093/ijl/4.4.261

Camacho, J. 2017. Aproximación al concepto de medioestructura lexicográfica y su influencia en la tipología lexicográfica: el diccionario de aprendizaje especializado y consulta. *Romanica Olomucensia* 29(1): 1-16. https://doi:org/10.5507/ro.2017.001

Caruso, V. 2024. Between Collocation and Colligation: An Experiment in Collaborative Lexicography. *International Journal of Lexicography* 37(1): 75-94. https://doi.org/10.1093/ijl/ecad026

Caruso, Valeria, Alessandra Chervino and Giulia Daniele. 2022. Disseminating Dictionary Skills with e-lex Tools. Klosa-Kückelhaus, Annette, Stefan Engelberg, Christine Möhrs and Petra Storjohann (Eds.). 2022. *Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress, Mannheim, Germany*, *12–16 July 2022*: 138-139. Mannheim: IDS-Verlag.

Chaves-Yuste, B. and M. Márquez-Cruz. 2022. Lexicografía digital y adquisición de léxico en inglés: una experiencia b-learning en los tiempos de Covid-19. Puebla-Martínez, B., R. Vinader and N. Navarro (Coord.). 2022. *Covid-19, multiverso de disciplinas. Una mirada desde la salud, la educación y la comunicación*: 350-367. Madrid: Dykinson.

Chaves-Yuste, B. and M. Márquez-Cruz. 2024. A Cooperative Digital Dictionary to Improve English for General Purposes Reading Comprehension in Secondary Education. Bobkina, J. and E. Domínguez Romero (Eds.). 2024. *Transformative Trends in Language Education. Pioneering Autonomous Learning Technologies*: 171-188. Berlin: Peter Lang. https://doi.org/10.3726/b21998

Climent, J. 2005. *El uso del diccionario bilingüe castellano/valenciano como instrumento didáctico en los ciclos de enseñanza secundaria obligatoria y bachillerato.* Unpublished PhD Thesis. Alicante: Universidad de Alicante.

Costa, Hernani, Gloria Corpas, Miriam Seghiri and Ruslan Mitkov. 2015. iCorpora: Compiling, Managing and Exploring Multilingual Data. Corpas, G., M. Seghiri, R. Gutiérrez Florido and M. Urbano (Eds.). 2015. *Nuevos horizontes en los Estudios de Traducción e Interpretación. Libro de resúmenes de AIET17, Málaga, Spain, 29–31 January 2015*: 74-76. Geneva: Tradulex.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Flores, E. 2022. Compilación de un corpus de especialidad con BootCaT para la traducción jurídica italiano / español: el caso de un contrato de agencia. *Onomázein. Revista de lingüística, filología y traducción* 57: 167-214. https://doi.org/10.7764/onomazein.57.09

Heuberger, R. 2016. Learners' Dictionaries: History and Development; Current Issues. Durkin, P. (Ed.). 2016. *The Oxford Handbook of Lexicography*: 25-43. Oxford: Oxford University Press.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Horák, Aleš and Adam Rambousek. 2007. Dictionary Management System for the DEB Development Platform. Sharp, Bernadette and Michael Zock (Eds). 2007. *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science, NLPCS 2007, Funchal, Portugal, 12–13 June 2007*: 129-138. Funchal: Insticc Press.

Jia, W., L. Zhang, A. Pack, Y. Guan and B. Zou. 2024. Digital Game-based Learning's Effectiveness on EFL Learners' Receptive and Productive Vocabulary Knowledge. *Language Learning & Technology* 28(1): 1-21.
https://hdl.handle.net/10125/73554

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): 1-12.
http://doi.org/10.1177/21582440211008422

Kovalevskaitė, J., L. Boizou, A. Bielinskienė, L. Jancaitė and E. Rimkutė. 2020. The First Corpus-driven Lexical Database of Lithuanian as L2. *Frontiers in Artificial Intelligence and Applications* 328: *Human Language Technologies — The Baltic Perspective:* 245-252.
https://doi.org/10.3233/FAIA200630

Mangeot, Mathieu and Antoine Chalvin. 2006. Dictionary Building with the Jibiki Platform: The GDEF Case. Calzolari, N. et al. 2006. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, 22–28 May 2006*: 1666-1669. Genoa: European Language Resources Association.
http://www.lrec-conf.org/proceedings/lrec2006/pdf/105_pdf.pdf

Márquez Cruz, M. 2024. *Modelos lexicográficos digitales en la lexicografía didáctica.* Madrid: Guillermo Escolar Editor.

Mavrić, M., I. Srdanović and D. Špica. 2023. Analiza kulturno specifičnih termina na temu tradicionalnih japanskih festivala. Mikulaco, I. (Ed.). 2023. *Jezik, prijevod i međukulturna komunikacija*: 50-61. Pula: University of Pula.

Měchura, Michal Boleslav. 2017. Introducing Lexonomy: An Open-source Dictionary Writing and Publishing System. Kosem, Iztok, Carole Tiberius, Miloš Jakubíček, Jelena Kallas, Simon Krek and Vít Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 Conference, Leiden, Netherlands, 19–21 September 2017*: 662-679. Brno: Lexical Computing CZ s.r.o.

Molnár, György, Beáta Orosz and Katalin Nagy. 2022. Current Issues and Possible IT Solutions for Digital Competence Development. Turčáni, Milan, Zoltán Balogh, Michal Munk, Martin Magdin, Ľubomír Benko and Jan Francisti (Eds). 2022. *14th International Scientific Conference on Distance Learning in Applied Informatics, Conference Proceedings, Štúrovo, Slovakia, 2–4 May 2022:* 267-276. Alphen aan den Rijn: Wolters Kluwer.

Nguyen, L.T. and K. Tuamsuk. 2022. Digital Learning Ecosystem at Educational Institutions: A Content Analysis of Scholarly Discourse. *Cogent Education* 9(1): 1-17.
https://doi.org/10.1080/2331186X.2022.2111033

Redecker, C. and Y. Punie (Eds.). 2017. *European Framework for the Digital Competence of Educators: DigCompEdu*. Seville: Publications Office of the European Union.
https://doi.org/10.2760/159770

Roebianto, A., S.I., Savitri, I. Aulia, A.R.I.E. Suciyana and L. Mubarokah. 2023. Content Validity: Definition and Procedure of Content Validation in Psychological Research. *TPM — Testing, Psychometrics, Methodology in Applied Psychology* 30(1): 5-18.
https://doi.org/10.4473/TPM30.1.1

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2004 (Article)

Rojo, G. 2021. *Introducción a la lingüística de corpus en español.* New York: Routledge.

Rubio, R.Y., J.E. Bonilla and J.A. Bernal. 2021. Dictionary Writing Systems y otras herramientas informáticas para la elaboración, administración y publicación de diccionarios. *Lingüística y Literatura* 42(80): 340-360.
https://doi.org/10.17533/udea.lyl.n80a20

Shahjahan, M., K.R. Ahmed, A. Al Hadrami, R. Islam, S. Hossain and S. Khan. 2021. Factors Influencing Poor Academic Performance among Urban University Students in Bangladesh. *International Journal of Evaluation and Research in Education (IJERE)* 10(4): 1140-1148.
http://doi.org/10.11591/ijere.v10i4.21158

Shaxrizoda, Q. 2024. The History of Corpus Linguistics and Basics. *Best Journal of Innovation in Science, Research and Development*: 104-111.

Sinclair, J. 2005. Corpus and Text — Basic Principles. Wynne, M. (Ed.). 2005. *Developing Linguistic Corpora: A Guide to Good Practice*: 1-16. Oxford: Oxbow Books.

Snezhko, Z., D. Babaskin, E. Vanina, R. Rogulin and Z. Egorova. 2022. Motivation for Mobile Learning: Teacher Engagement and Built-in Mechanisms. *International Journal of Interactive Mobile Technologies (iJIM)* 16(01): 78-93.
https://doi.org/10.3991/ijim.v16i01.26321

Srdanović, I. 2020. From Specialized Web Corpora of Tourism to a Learner's Dictionary. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 46(2): 1059-1083.
https://doi.org/10.31724/rihjj.46.2.31

Strelan, P., A. Osborn and E. Palmer. 2020. The Flipped Classroom: A Meta-analysis of Effects on Student Performance across Disciplines and Education Levels. *Educational Research Review* 30: 100314.
https://doi.org/10.1016/j.edurev.2020.100314

Tarp, S. 2019. La ventana al futuro: despidiéndose de los diccionarios para abrazar la lexicografía. *RILEX. Revista sobre investigaciones léxicas* 2(2): 5-36.
https://doi.org/10.17561/rilex.v2.n2.1

Tarp, S. and R.H. Gouws. 2023. A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and Implications for the Future. *Lexikos* 33(1): 425-447.
http://doi.org/10.5788/33-1-1826

Tiberius, C. and K. Heylen. 2022. ELEXIS Pathfinder to Computational Lexicography for Developers and Computational Linguists. Version 1.0.0. DARIAH-Campus. [Training module].
https://campus.dariah.eu/resources/pathfinders/elexis-pathfinder-to-computional-lexicography-for-developers-and-computational-linguists

Toledo, M.C. and A. Camero. 2021. Compilación semiautomática con Sketch Engine de un corpus ad hoc comparable bilingüe (inglés–español) de contratos de distribución exclusiva (DISTRIBUCOR). Sartor, E. (Ed.). 2021. *Los corpus especializados en la lingüística aplicada: traducción y enseñanza*: 15-41. Mantua: Pliegos Hispánicos, Universitas Studiorum S.r.l. – Casa Editrice.

University of Cambridge ESOL Examinations. 2015. *Business Vantage. Business English Certificate 5.* Cambridge: Cambridge University Press.

Van Laar, E., A.J.A.M. van Deursen, J.A.G.M. van Dijk and J. de Haan. 2020. Determinants of 21st-Century Skills and 21st-Century Digital Skills for Workers: A Systematic Literature Review. *SAGE Open* 10(1): 1-14.
https://doi.org/10.1177/2158244019900176

Yong, H. and J. Peng. 2008. *Chinese Lexicography: A History from 1046 BC to AD 1911.* Oxford/New York: Oxford University Press.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

Krajšavar — An Algorithm for Automatic Recognition of Abbreviations in English Dictionary Entries Compiled in a Dictionary of Abbreviations

Mojca Kompara Lukančič, *Faculty of Criminal Justice and Security and Faculty of Tourism, University of Maribor, Slovenia (mojca.kompara@gmail.com) (mojca.kompara@um.si) (https://orcid.org/0000-0003-2368-4161)*

Abstract: This article describes the use of Krajšavar — an algorithm for the automatic recognition of abbreviations in compiling English dictionary entries for a dictionary of abbreviations (*Slovar krajšav*), published in 2025 and financed by the Slovenian Research and Innovation Agency (ARIS). Together with the Slovenian dictionary of abbreviations (*Slovenski slovar krajšav*) published in 2023, the mentioned dictionary adopts a pioneering approach to the compilation of dictionaries in Slovenia as these are the first contemporary dictionaries of abbreviations. The *Slovar krajšav* was compiled based on an analysis of the characteristics of English dictionary entries for abbreviations, and according to the characteristics of the compilation methods used for bilingual dictionaries. The dictionary of abbreviations includes entries in 22 languages, with the most frequent being English, Italian and French. In this article the focus is on compiling English dictionary entries and using the Krajšavar algorithm.

Keywords: ABBREVIATIONS, ENGLISH, DICTIONARY, ALGORITHM, DICTIONARY OF ABBREVIATIONS, LANGUAGES, KRAJŠAVAR

Opsomming: Krajšavar — 'n algoritme vir die outomatiese herkenning van afkortings in Engelse woordeboekinskrywings wat in 'n afkortingswoordeboek saamgestel is. Hierdie artikel beskryf die gebruik van Krajšavar — 'n algoritme vir die outomatiese herkenning van afkortings in Engelse woordeboekinskrywings wat in 'n afkortingswoordeboek (*Slovar krajšav*) saamgestel is. Dit is in 2025 gepubliseer en gefinansier deur die Sloweense Agentskap vir Navorsing en Innovasie (ARIS). Saam met die Sloweense afkortingswoordeboek (*Slovenski slovar krajšav*) wat in 2023 gepubliseer is, verrig die woordeboek onder bespreking baanbrekerswerk in die samestelling van woordeboeke in Slowenië, aangesien hierdie publikasies die eerste eietydse afkortingswoordeboeke is. *Slovar krajšav* se samestelling berus op 'n ontleding van die eienskappe van Engelse woordeboekinskrywings vir afkortings, en is volgens die eienskappe van die samestellingsmetodes wat vir tweetalige woordeboeke gebruik word, uitgevoer. Die afkor-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

tingswoordeboek sluit inskrywings in 22 tale in, met Engels, Italiaans en Frans wat die algemeenste is. In hierdie artikel val die klem op die samestelling van Engelse woordeboekinskrywings en die gebruik van die Krajšavar-algoritme.

Sleutelwoorde: AFKORTINGS, ENGELS, WOORDEBOEK, ALGORITME, AFKORTINGS-WOORDEBOEK, TALE, KRAJŠAVAR

1. Introduction

In 2011, Kompara Lukančič and Holozan (2011) examined the automatic compilation of simple and complex dictionary entries in a dictionary of abbreviations. The research was based partly on the development of an algorithm for the automatic recognition of abbreviations in electronic texts (Kompara Lukančič 2010, 2011b). As Kompara Lukančič (2009, 2010, 2017, 2018) mentions, abbreviations — grouped, joined or unified compositions of letters — emerge rapidly and suddenly in a language, which makes it difficult to keep track of them and include them in general, bilingual, terminological, orthographic dictionaries (Kompara Lukančič 2009). This might be because some abbreviations remain in a language for only a limited period, such as *COVID* or *SARS*, while others are no longer recognised as abbreviations, e.g. *radar*. This may make it is difficult to determine which abbreviations should be included in a dictionary. Establishing a strategy or methodology for including abbreviations in dictionaries is also subject to discussions since the "life expectancy" of an abbreviation cannot be anticipated.

Kompara Lukančič (2009) argues that abbreviations are a growing phenomenon in the Slovenian language, with her research in a diachronic frame showing that abbreviations were included in orthographic dictionaries of the Slovenian language (Kompara Lukančič 2009, 2018). Kompara Lukančič (2018) emphasises the need to compile a separate dictionary of abbreviations for the Slovenian language, akin the abbreviations dictionaries already available in other languages, such as English, Italian, German and French.

The need to compile a Slovenian dictionary of abbreviations is further justified by the existence of such dictionaries in other languages, particularly English as it has the biggest number of published dictionaries of abbreviations (Kompara Lukančič 2009). Kompara Lukančič (ibid.) advocates for a contemporary dictionary of abbreviations in Slovenian that includes both Slovenian and foreign abbreviations, and whose structure and content is comparable to similar dictionaries in other languages, including English (cf. Paxton 1983, De Sola 1986, Fergusson 2000), French (cf. Faudouas 1990, Murith and Bocabeille 1992), German (cf. Koblischke 1983, Steinhauer 2005), Italian (cf. Malossini 1999, Righini 2001) and Spanish (cf. Galende 1997, 2001).

The aim of the paper is to expose the need for the compilation of a dictionary of abbreviations for Slovenian, specifically a dictionary in which both Slovenian and foreign abbreviations and expansions are provided, and that has dic-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

tionary entries composed of language qualifiers, filed qualifiers and translations. As an example of good practice in the compilation process of such a dictionary, similar dictionaries compiled for English, Italian and French is considered. In this paper, the focus is on the process of compiling the English dictionary entries in the *Slovar krajšav*, as English is the dominant foreign language among the 22 foreign languages included in the dictionary.

2. Dictionaries of abbreviations in the Slovenian language

In the past 25 years, abbreviations in the Slovenian language have been extensively discussed by Kompara Lukančič (2009, 2017, 2018, 2023a), Logar (2005), Verovnik and Logar (2006), Verovnik (2018) and Tonin (2022). The first dictionary of abbreviations *Kratice* was published in 1948 (Župančič 1948), followed by *Rečnik jugoslovenskih skraćenica* (Zidar 1971), and two more recent online dictionary attempts: *Slovarček krajšav* (Kompara Lukančič 2006) and *Slovar krajšav* (Kompara Lukančič 2011a). *Slovarček krajšav* (Kompara Lukančič 2006) was compiled as part of student work, with basic lexicographical knowledge yet without an appropriate structure. It is a collection of over 6 000 manually collected Slovenian and foreign abbreviations with Slovenian translations of foreign abbreviations. Among the foreign languages, are English, Italian, German, Spanish, Latin and French abbreviations, to name a few.

Figure 1: The dictionary entry for C (*Slovarček krajšav* 2006)

C	1. lat.: centum: rimska številka 100 2. Carboneum: kemijski simbol za ogljik 3. angl.: see: glej 4. stopinja Celzija 5. it.: codice: zakonik 6. programski jezik C

As evident in Figure 1 above, the dictionary entry is relatively simple. It is composed of an abbreviation, e.g. *C* in bold, followed by numbered expansions, e.g. (1), the abbreviated language qualifier (it.) for *Italian* next to (5), and the foreign expansions, e.g. *centum* and *Carboneum*, and the Slovenian translation given at (1) *rimska številka (Roman number)*. In the dictionary entry, both Slovenian and foreign expansions are given. Unfortunately, the dictionary does not have a concise structure, expansions should be alphabetically ordered, translations should be verified, and cross-references, encyclopaedic data and field qualifiers should be unified.

Slovar krajšav (Kompara Lukančič 2011a) is a dictionary of abbreviations that was compiled entirely automatically. Abbreviations and expansions were automatically extracted from a corpus containing Slovenian texts from the newspaper *Delo*, with this corpus containing five years' worth of *Delo* editions and a total of 60 million words. This automatic extraction was performed by using the algorithm for the automatic recognition of abbreviations and expansions in electronic texts (Kompara Lukančič 2009, 2010, 2017, 2018). After extraction,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

abbreviation–expansion pairs were manually verified, expansions were lemmatised, and language qualifiers were added automatically (Kompara Lukančič and Holozan 2011). The collection of 2 571 dictionary entries is available for free on the *Termania* website (https://www.termania.net/).

At the initial stage despite filtering Slovenian texts, Slovenian and foreign abbreviations were found among the dictionary entries. As shown in Figure 2 below, foreign abbreviations, such as English abbreviations and their expansions, are included in the dictionary. The foreign expansions are followed by the abbreviated language qualifier in brackets (en) *English*, which was added automatically (Kompara Lukančič and Holozan 2011). As seen from the dictionary entry, no Slovenian translations are provided, however. This information is crucial for dictionary users and would make the dictionary more usable and oriented towards them. By providing a translation the dictionary would solve users' translation problems. A similar approach is visible in the Italian dictionary of abbreviations (Righini 2001) but is not present in the German (Steinhauer 2005).

Figure 2: The dictionary entry for BA (*Slovar krajšav* 2011a)

BA
British Airways (en)
Bank Austria (en)
Budapest Airport (en)

Figure 3 shows that Slovenian abbreviations were also included in the collection. Slovenian expansions were lemmatised, and abbreviated language qualifier in brackets, such as (sl) *Slovenian*, were added automatically. As seen in Figure 3 below, no additional information or encyclopaedic data is included in the dictionary entry. The inclusion of such data could however be appreciated by dictionary users as it would give additional information concerning an abbreviation and expansion. It would clarify the meaning to the users, and for that reason such data should be seen as relevant and an improvement of the dictionary. A similar approach of inclusion of additional or encyclopaedic data is visible in the Italian dictionary of abbreviations (Righini 2001).

Figure 3: The dictionary entry for DA (*Slovar krajšav* 2011a)

DA
Demokratska akcija (sl)
Demokratična alternativa (sl)
državni aparat (sl)

Slovarček krajšav (Kompara Lukančič 2006) and *Slovar krajšav* (Kompara Lukančič

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

2011a) are important to consider because they led to the preparation of the Slovenian dictionary of abbreviations *Slovenski slovar krajšav* (Kompara Lukančič 2023b). This new dictionary is a collection of 3 400 alphabetically ordered dictionary entries and over 4 000 expansions. The dictionary was compiled using the algorithm for the automatic recognition of abbreviations and expansions in Slovenian electronic texts as well as the algorithm for language detection and lemmatisation of Slovenian expansions. In the first and last phases of the compilation, abbreviations were collected manually.

The dictionary includes abbreviation–expansion pairs from the general language and several terminological fields. The dictionary was published in print in 2023 thanks to a grant given by the Slovenian Research and Innovation Agency (ARIS), and is also available online on the *Termania* website. As may be observed in Figure 4 below, the dictionary entry is relatively simple. The headword is followed by alphabetically ordered expansions and some additional information or explanations, such as *pevski glas* (singing voice), when needed. Filed qualifiers are not included. For expansions with two or more abbreviations, cross-references are included as a separate dictionary entry, for example *a*.

Figure 4: The dictionary entry for A (*Slovenski slovar krajšav* 2023b)

A
adenin
alt; pevski glas → a
amper
as; igralna karta
masno število

3. The algorithm for the automatic recognition of abbreviations and expansions in electronic texts

Over the course of a decade, there has been a shift from a dictionary compilation process where an automated process was used to some degree, to the current advance towards a fully automatised process (Rundell 2023). Rundell (2023) describes some semi-automated projects, the gradual progress towards new approaches in compiling dictionaries (as in post-editing lexicography), the role of lexicographers in post-editing (namely evaluating and refining the first dictionary draft which was automatically generated and imported into a dictionary writing and editing system), and finally the arrival of algorithms in dictionary compilation.

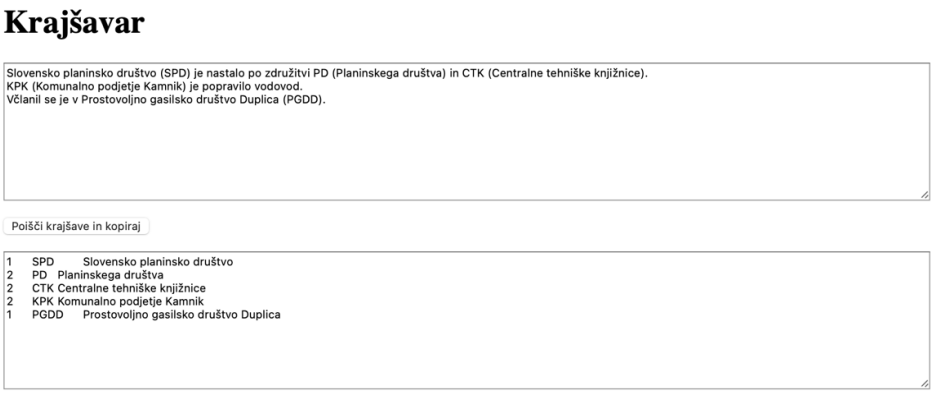
Algorithms for the automatic recognition of abbreviations and expansions in electronic texts have been researched for nearly three decades. The first such algorithm was developed in 1999 by Taghva and Gilbreth (1999). This pioneering approach was followed by Yeates (1999), Larkey et al. (2000), Park and Byrd (2001), Schwartz and Hearst (2003), Zahariev (2004), Xu and Huang (2005), Zhou, Torvik

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

and Smalheiser (2006), Šatev and Nikolov (2008), and Kompara Lukančič (2009, 2018). These algorithms focused on recognising abbreviations and expansions in English texts, mainly in the medical field (Schwartz and Hearst 2003).

Kompara Lukančič (2011b, 2018) developed the first algorithm for recognising abbreviations and expansions in Slovenian. This algorithm has since been modified, improved and adapted several times (Kompara Lukančič 2009, 2011b, 2018), also for filtering English texts. With the support of Peter Holozan, an IT specialist who developed the latest version of the algorithm's user interface shown in Figure 5 (below), the Krajšavar algorithm is currently capable of filtering and extracting abbreviations and expansions from both Slovenian and English texts.

Figure 5: Krajšavar — an algorithm for the automatic recognition of abbreviations and expansions in electronic texts



The final preparation of the algorithm entailed several steps. It started with establishing basic rules for recognising abbreviations that were based on abbreviations' characteristics in both Slovenian and English. This was followed by setting basic rules for recognising an expansion, and finally implementing all the concepts in a user-friendly digital form.

The Krajšavar algorithm was developed, initially for the author's purposes, in the interface shown above in Figure 5. The user interface is composed of two windows: in the first one, the text is included for filtration, while in the second one, the abbreviations and expansions appear. Behind this interface, a complicated algorithm operates that first cuts the text into separate words, treating punctuation marks as separate words. The algorithm then filters the words from the text in search of open brackets. Upon finding open brackets, the algorithm continues to search for the first closed bracket that follows the first open bracket. If between the brackets there is only one word that starts with an uppercase letter (or is entirely uppercase) and is made up of at least two letters, the algorithm assumes it is an abbreviation.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

To find expansions, the algorithm looks for words preceding the brackets and tries up to 10 words placed them. It then searches for the first word that matches the first letter of the abbreviation placed after open brackets (in uppercase). However, this procedure does not work well in languages like English in which words such as titles and proper names start with an uppercase letter. The algorithm tries to find abbreviations by following the same rule, namely if there are at least two words within brackets and the first letter of the first word matches the first word before the brackets, it assumes the abbreviation is before the brackets and the abbreviation's expansion is within the brackets.

The Krajšavar algorithm does not need an external server, is written in JavaScript, and included in the file krajšave.html. By clicking "Poišči krajšave in kopiraj" (find abbreviations and copy), the function "isciKrajsave" (find abbreviations) is recalled. In such a way, the text is read from the input file and sent to the function of tokenisation that cuts the text into words with the help of the regular expression. In the implementation in Python, an external library nltk.tokenize was used, in JavaScript the tokenisation was written manually to avoid any unnecessary external dependencies that might complicate the use of the webpage elsewhere. The candidates for abbreviations are given in the sequence "result" that at the end is written in the output file. At the very end, the sequence is copied onto the clipboard to enable further work and editing.

4. The compilation for the dictionary of abbreviations *Slovar krajšav*

Over the past 20 years, the need for a dictionary of abbreviations to be compiled has been discussed extensively (Kompara Lukančič 2009, 2010, 2011b, 2017, 2018; Kompara Lukančič and Holozan 2011). The compilation of the Slovenian dictionary of abbreviations, *Slovenski slovar krajšav* (Kompara Lukančič 2023b), led to the compilation of a more comprehensive work, namely a dictionary of abbreviations consisting of foreign abbreviations from 22 languages that have been compiled in a single volume. This dictionary provides the expansions of these abbreviations in the foreign languages along with Slovenian translations and, if they exist, Slovenian abbreviation equivalents. Preparation of the dictionary was made possible with the financial support of the ARIS, and the dictionary was published in 2025.

On a macrostructural level, the dictionary includes approximately 3 500 alphabetically ordered dictionary entries and over 4 200 expansions, noting that some abbreviations can have several expansions. The dictionary structure is based on the structure used in the Slovenian dictionary of abbreviations *Slovenski slovar krajšav* (Kompara Lukančič 2023b). Work on compiling the dictionary spanned over two decades. The dictionary was initially compiled using the algorithm for the automatic recognition of abbreviations and expansions in electronic texts (Kompara Lukančič 2009, 2010), and also the algorithm for language recognition and lemmatisation, with the final stage of the compilation process being completed manually. In later years, the algorithm was adapted

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

for filtering English texts, and the material obtained from these texts was added to the dictionary.

In the process of compiling English dictionary entries for the *Slovar krajšav*, the characteristics of dictionary entries in English dictionaries of abbreviations were examined (Kompara Lukančič 2023a). An extensive analysis of English dictionaries of abbreviations published between 1942 and 2019 was conducted in 2009 and updated in 2023 (Kompara Lukančič 2009, 2023a). The Krajšavar algorithm was introduced in the final phase of the compilation process. As discussed in Kompara Lukančič (2023a), while compiling the English entries for the *Slovar krajšav* based on the examination of dictionaries of abbreviations, the alphabetical order of expansions found in *Everyman's Dictionary of Abbreviations* (Paxton 1983) was retained, as well as language qualifier *angl. English*. The numerical order preceding an expansion was not preserved, however. Similar to *A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations* (Partridge 1942), some additional information or encyclopaedic data were added to a dictionary entry. As in the *Everyman's Dictionary of Abbreviations* (Paxton 1983), field qualifiers, abbreviated and provided in brackets were also used in a dictionary entry, for example (voj.) *army*. Cross-references and expansions with more than one abbreviation were also included as separate dictionary entries, which is similar to the approach in the *Everyman's Dictionary of Abbreviations* (Paxton 1983). Expansions of one abbreviation are written within a single dictionary entry, which differs from the *Dictionary of Abbreviations in Medical Sciences* (Heister 1989) where every expansion has a separate dictionary entry. English expansions are also followed by translations into Slovenian, like in *The Barnhart Abbreviations Dictionary* (Barnhart 1995), *The New Penguin Dictionary of Abbreviations* (Fergusson 2000), and the *Dictionary of Financial Abbreviations* (Paxton 2003), where foreign expansions are translated.

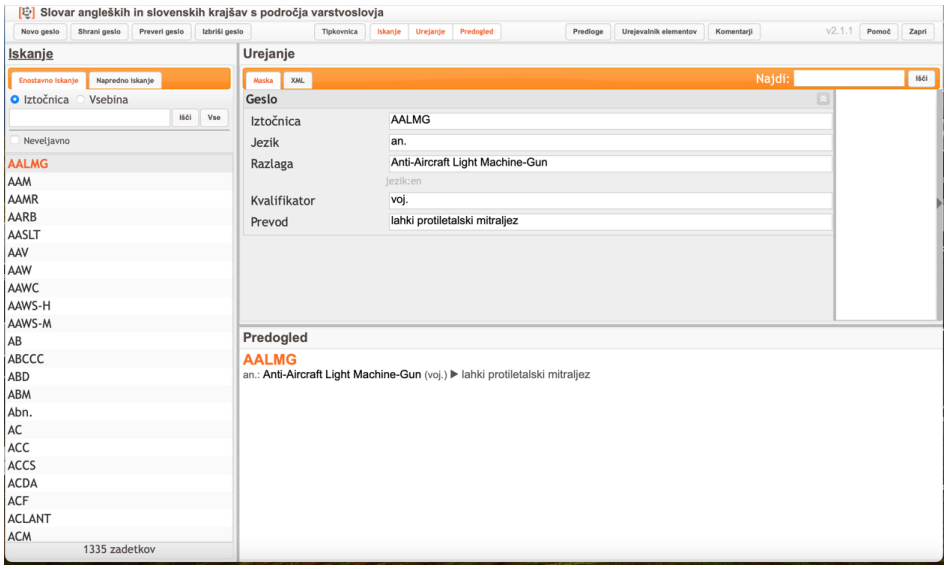
5. The use of the Krajšavar algorithm in compiling the *Slovar krajšav*

As mentioned, the Krajšavar algorithm was introduced in the final stage of the compilation process of the *Slovar krajšav*. The research was based on a collection of English texts following a text typology of texts from the field of criminal justice and security. These texts were collected using Google and Google Scholar as search engines and in line with a text typology for criminal justice and security (Kompara Lukančič, forthcoming). A total of 210 texts were filtered using the Krajšavar algorithm, resulting in 1 335 abbreviation–expansion pairs. The text filtration was followed by the automatic transfer of all abbreviation–expansion pairs into the *Termania* dictionary editing interface, as shown in Figure 6 below. As evident in Figure 6, the obtained abbreviations and expansions were included in the *Termania* dictionary mask, and language qualifiers, such as *an. English*, as well as field qualifiers, such as *(voj.) army*, were added together with the Slovenian translations. The database was named the Dictionary of English and Slovenian Abbreviations from the Field of Criminal Justice and Security

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

(*Slovar angleških in slovenskih krajšav s področja varstvoslovja*). Each dictionary entry in the preview section consists of a headword, namely the abbreviation (for example, *AALMG*), followed by a language qualifier that is abbreviated (such as *an.*), the foreign expansion (e.g., *Anti-Aircraft Light Machine-Gun*), a field qualifier, which is abbreviated in brackets, such as *(voj.)*, and finally the symbol ▶ denoting the Slovenian translation.

Figure 6: Example of a dictionary entry for AALMG



If an abbreviation has several expansions, all of them are included in the same dictionary entry, with the alphabetical order preserved. Symbols used in the process of compiling the *Slovar krajšav* are explained in the paragraph below. Abbreviation–expansion pairs obtained with Krajšavar were manually verified before being included in the *Slovar krajšav*. The structure of the dictionary entries in the *Slovar krajšav* is explained in the following section.

6. Examples of English dictionary entries in the *Slovar krajšav*

Figure 7 below presents a dictionary entry showing that the abbreviation *CC* has seven alphabetically ordered expansions in the dictionary entry. The dictionary entry consists of an entry word, followed by an abbreviated language qualifier (such as *angl.* – *English*), the English (or other languages) expansion, and an abbreviated field qualifier (such as *(ekon.) economics*) in brackets.

Some special symbols are used in the dictionary entry. Expansions are

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

followed by the symbol ►, denoting a translation, and the symbol ◊, indicating a descriptive sentence. The symbol ◊ is used for culturally specific expansions that do not have a translation equivalent, and thus a descriptive sentence is used instead. Examples of this are *član kanadskega viteškega reda* (Member of the Order of Knights of Canada) and *polje v e-poštnemu sporočilu za naslovnike, ki prejmejo kopijo pošte* (a field in the email message for other recipients who receive a copy of the email). The symbol ○ is used when a Slovenian equivalent abbreviation of the English one exists, for example *Kp, OZS* and *POVC*. In these cases, cross-references are also provided and included as separate dictionary entries. If an expansion has multiple abbreviations, such as *CC, c. c. and cc* for *carbon copy*, they are introduced in the dictionary entry with the word *tudi* (also) and included as separate dictionary entries for each abbreviation.

Figure 7: Example of the dictionary entry for CC

CC
angl.: carbon copy; tudi c. c., cc ◊ polje v e-poštnemu sporočilu za naslovnike, ki prejmejo kopijo pošte ○ Kp
angl.: Chamber of Craft (ekon.) ► Obrtna zbornica Slovenije ○ OZS
angl.: combat command (voj.) ► bojno poveljstvo
angl.: command center (voj.) ► poveljniški center ○ POVC
angl.: Companion of the Order of Canada ◊ član kanadskega viteškega reda
fr.: corps consulaire ► konzularni zbor
it.: Carabinieri (polic.) ► karabinjerji

As shown in Figure 8 below, the official translations, such as *srednjeevropski čas* (Central European Time), are mainly provided for English expansions that have previously been translated and can be found in dictionaries, databases and so forth. Where an official translation of an English expansion could not be found, a descriptive sentence containing additional information was used, for example *skupna carinska tarifa za uvoz iz držav nečlanic EU* (common external tariff).

Figure 8: Example of the dictionary entry for CET

CET
angl.: Central European Time (geo.) ► srednjeevropski čas
angl.: common external tariff (ekon.) ◊ skupna carinska tarifa za uvoz iz držav nečlanic EU

7. Discussion

As outlined in the introduction, the structure of the *Slovar krajšav* is based on the earlier compiled *Slovenski slovar krajšav* (Kompara Lukančič 2023b) and an

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

analysis of the characteristics of dictionaries of abbreviations in English (Kompara Lukančič 2023a). As seen in the above examples, in the dictionary entry the alphabetical order is preserved, like with the *Everyman's Dictionary of Abbreviations* (Paxton 1983). Additionally, the structure of the English dictionary entry is relatively simple and is composed of a headword, followed by an abbreviated language qualifier, an English expansion, a field qualifier in brackets (if the abbreviation is not generic, such as in the case of *carbon copy*), and a translation, equivalent or descriptive sentence in Slovenian, as well as a Slovenian abbreviation, if one exists, such as *Kp, OZS* and *POVC*.

All elements are introduced in the dictionary entry with the use of special symbols, as explained in the previous section. In the printed version of the dictionary, the use of special symbols as well as the dictionary's macro and microstructure are explained in the introduction. There dictionary users will find examples of dictionary entries and an explanation of the structure of elements of an entry.

In some entries, additional or encyclopaedic data are included in line with a specific terminological dictionary of abbreviations, namely *A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations* (Partridge 1942). Abbreviated field qualifiers in brackets, cross-references, and expansions with multiple abbreviations are included in the same manner as in *Everyman's Dictionary of Abbreviations* (Paxton 1983), while Slovenian translations of the expansions are included in the dictionary entries, similar to the approach in *The Barnhart Abbreviations Dictionary* (Barnhart 1995). English abbreviations represent the most frequent foreign abbreviations collected in the *Slovar krajšav* (cf. Figures 7 and 8).

By including abbreviation–expansion pairs obtained with the Krajšavar algorithm (cf. Figure 6), a larger number of abbreviations were included in the dictionary, namely those from the field of criminal justice and security. Filtering texts from this field, gathered in line with the text typology classification (Kompara Lukančič, forthcoming), ensured a broader range of abbreviations from the field, as well as the inclusion of up-to-date abbreviations from the field since the filtered texts came from a more recent time frame, namely the last five years. This approach of including abbreviations by using the Krajšavar algorithm also lowered the possibility of forgetting to include relatively known and new abbreviations that might have been overlooked in a manual compilation. In a way, the algorithm's application ensures the outcome is up-to-date and not missing relevant abbreviations. Moreover, thanks to the text typology classification for criminal justice and security (Kompara Lukančič, forthcoming), a similar classification of texts can be applied to other language for special purposes fields, such as tourism, medicine and economics, enabling the inclusion of abbreviations from other fields as well.

8. Conclusion

The article describes the application of the Krajšavar algorithm in the process of

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

compiling the *Slovar krajšav* and shows the need for a dictionary of abbreviations to be compiled in the Slovenian language. Dictionaries of abbreviations for the Slovenian language are presented in a synchronic and diachronic framework (cf. Kompara Lukančič 2018), namely two outdated dictionaries *Kratice* (Župančič 1948) and *Rečnik jugoslovenskih skraćenica* (Zidar 1971), and three more recent online dictionary attempts, namely *Slovarček krajšav* (Kompara Lukančič 2006), *Slovar krajšav* (Kompara Lukančič 2011a), and the most recently published *Slovenski slovar krajšav* (Kompara Lukančič 2023b).

The publication of the *Slovenski slovar krajšav* (Kompara Lukančič 2023b) led to the compilation of the single-volume dictionary of abbreviations *Slovar krajšav*, which is a collection of 3 500 alphabetically ordered dictionary entries and over 4 200 expansions from 22 foreign languages. The article outlines the overall compilation of the *Slovar krajšav* and discusses examples of dictionary entries for English abbreviations. As shown by the presented examples, a dictionary entry is composed following the compilation process used in previously published dictionaries the *Slovarček krajšav* (Kompara Lukančič 2006), *Slovar krajšav* (Kompara Lukančič 2011a) and *Slovenski slovar krajšav* (Kompara Lukančič 2023b), coupled with the characteristics of a range of English dictionaries of abbreviations (Kompara Lukančič 2009, 2018).

The compilation process took almost two decades to complete and included the application of several algorithms for lemmatisation, language detection and the automatic recognition of abbreviations. In the final preparation steps, the dictionary was compiled manually and with the help of Krajšavar algorithm, which allowed for the inclusion of abbreviations from a specialised field, as well as relevant abbreviations obtained from a range of texts following the relevant text typology. The *Slovar krajšav,* together with the *Slovenski slovar krajšav* (Kompara Lukančič 2023b), therefore represents an important contribution to the linguistic framework of abbreviations for the Slovenian language.

References

Barnhart, K.R. (Ed.). 1995. *The Barnhart Abbreviations Dictionary.* New York: John Wiley & Sons.

De Sola, R. 1986. *Abbreviations Dictionary.* New York: Elsevier.

Fergusson, R. 2000. *The New Penguin Dictionary of Abbreviations.* London: Penguin.

Faudouas, J.-C. 1990. *Dictionnaire des abréviations courantes de la langue française*. Paris: La Maison du Dictionnaire.

Galende, J.C. 1997/2001. *Diccionario general de abreviaturas españolas*. Madrid: Editorial Verbum.

Heister, R. 1989. *Dictionary of Abbreviations in Medical Sciences.* Berlin/Heidelberg/New York: Springer.

Koblischke, H. 1983. *Großes Abkürzungsbuch*. Third edition. Leipzig: VEB Bibliographisches Institut.

Kompara Lukančič, M. 2006. *Slovarček krajšav*. Ljubljana: Inštitut za slovenski jezik Fran Ramovš ZRC SAZU.

Kompara Lukančič, M. 2009. Prepoznavanje krajšav v besedilih. Weiss, P. (Ed.). 2009. *Jezikoslovni zapiski* 15(1–2): 95-112. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.

Kompara Lukančič, M. 2010. Krajšavni slovarji. Weiss, P. (Ed.). 2010. *Jezikoslovni zapiski* 16(2): 111-129. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

Kompara Lukančič, M. 2011a. *Slovar krajšav*. Kamnik: Amebis, Termania.

Kompara Lukančič, M. 2011b. Razvoj algoritma za samodejno prepoznavanje krajšav in krajšavnih razvezav v elektronskih besedilih. Weiss, P. (Ed.). 2011. *Jezikoslovni zapiski: Zbornik Inštituta za slovenski jezik Frana Ramovša* 17(2): 107-122. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.

Kompara Lukančič, M. 2017. Zasnova novega slovarja krajšav. Weiss, P. (Ed.). 2017. *Jezikoslovni zapiski* 23(1): 77-92. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.

Kompara Lukančič, M. 2018. *Sinhrono-diahroni pregled krajšav v slovenskem prostoru in sestava slovarja krajšav*. Maribor: Univerzitetna založba Univerze.

Kompara Lukančič, M. 2023a. Compilation of English Entries in the *Contemporary Slovene Dictionary of Abbreviations*. *International Journal of Lexicography* 36(2): 195-210. https://doi.org/10.1093/ijl/ecac016

Kompara Lukančič, M. 2023b. *Slovenski slovar krajšav*. Maribor: Univerza v Mariboru, Univerzitetna založba.

Kompara Lukančič, M. 2025. *Slovar krajšav*. Maribor: Univerza v Mariboru, Univerzitetna založba. (in press). https://press.um.si/index.php/ump/catalog/book/948

Kompara Lukančič, M. Krajšavar — An Algorithm for Recognizing English Abbreviations in Texts Related to Criminal Justice and Security (forthcoming).

Kompara Lukančič, M. and P. Holozan. 2011. What is Needed for Automatic Production of Simple and Complex Dictionary Entries in the First Slovene Online Dictionary of Abbreviations Using Termania Website. Kosem, Iztok and Karmen Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011, Bled, Slovenia, 10–12 November 2011*: 140-146. Ljubljana: Trojina, Institute for Applied Slovene Studies.

Larkey, L.S., P. Ogilvie, M.A. Price and B. Tamilio. 2000. Acrophile: An Automated Acronym Extractor and Server. *Proceedings of the Fifth ACM Conference on Digital Libraries, San Antonio, Texas, USA, 2–7 June 2000:* 205-214. New York: Association for Computing Machinery.

Logar, N. 2005. Norma v slovarju sodobne slovenščine: zloženke in kratice. Družboslovne razprave (Ed.). 2005. *Družboslovne razprave* 21(48): 211-225. Ljubljana: Slovensko sociološko društvo: Fakulteta za družbene vede.

Malossini, A. (Ed.). 1999. *Dizionario delle sigle e degli acronimi*. Milan: A. Vallardi.

Murith, J. and J.-M. Bocabeille. 1992. *Dictionnaire des abréviations et acronymes*. Second edition. Paris: Technique et documentation — Lavoisier.

Park, Y. and R.J. Byrd. 2001. Hybrid Text Mining for Finding Abbreviations and Their Definitions. *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing, Pittsburgh, Pennsylvania, 3–4 June 2001*. https://aclanthology.org/W01-0516.pdf

Partridge, E. 1942. *A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations.* London: Routledge.

Paxton, J. 1983. *Everyman's Dictionary of Abbreviations.* London: Dent & Sons.

Paxton, J. 2003. *Dictionary of Financial Abbreviations.* New York: Routledge.

Righini, E. 2001. *Dizionario di sigle, abbreviazioni e simboli*. Bologna: Zanichelli.

Rundell, Michael. 2023. Automating the Creation of Dictionaries: Are We Nearly There? *ASIALEX 2023: Lexicography, Artificial Intelligence, and Dictionary Users (Asialex 2023), 22–24 June 2023, Seoul, Korea:* 9-17. Seoul: Yonsei University.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2005 (Project)

Schwartz, A.S. and M.A. Hearst. 2003. A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Texts. *Proceedings of the Pacific Symposium on Biocomputing* 8: 451-462. http://psb.stanford.edu/psb-online/proceedings/psb03/schwartz.pdf

Steinhauer, A. 2005. *Das Wörterbuch der Abkürzungen*. Mannheim: Duden Verlag.

Šatev, V. and N. Nikolov. 2008. Using the Web as a Corpus for Extracting Abbreviations in the Serbian Language. Erjavec, T. and J. Žganec Gros (Eds.). 2008. *Jezikovne tehnologije: Zbornik 11. mednarodne multikonference Informacijska družba — IS 2008*: 75-79. Ljubljana: Institut Jožef Stefan.

Taghva, K. and J. Gilbreth. 1999. Recognizing Acronyms and Their Definitions. *International Journal on Document Analysis and Recognition* 1(4): 191-198. https://doi.org/10.1007/s100320050018

Tonin, G. 2022. Kratični termini in priporočila za njihovo uporabo. *Jezik in Slovstvo* 67(1–2): 209-221. https://doi.org/10.4312/jis.67.1-2.209-221

Verovnik, T. 2018. Obravnava kratic v prenovljenih pravopisnih pravilih. *Jezikoslovni zapiski* 24(2): 43-54. https://doi.org/10.3986/jz.v24i2.7104

Verovnik, T. and N. Logar. 2006. O jeziku, stilu i utjecaju slovenskih tiskanih oglasa. Granić, J. (Ed.). 2006. *Jezik i mediji: jedan jezik: više svjetova:* 743-752. Zagreb, Split: Hrvatsko društvo za primijenjenu lingvistiku.

Xu, J. and Y. Huang. 2005. A Machine Learning Approach to Recognizing Acronyms and Their Expansions. *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 18–21 August 2005. Volume 4: 2313-2319. https://doi.org/10.1109/ICMLC.2005.1527330

Yeates, S. 1999. Automatic Extraction of Acronyms from Text. *Proceedings of the Third New Zealand Computer Science Research Students' Conference, Te Kohinga Marama Marae, Hamilton, New Zealand, 6–9 April, 1999. University of Waikato 1999:* 117-124.

Zahariev, M. 2004. *A (Acronyms).* Unpublished PhD Thesis. School of Computing Science, Simon Fraser University.

Zhou, W., V.I. Torvik and N.R. Smalheiser. 2006. ADAM: Another Database of Abbreviations in MEDLINE. *Bioinformatics* 22*:* 2813-2818. https://doi.org/10.1093/bioinformatics/btl480

Zidar, J. 1971. *Rečnik jugoslovenskih skraćenica.* Beograd: Međunarodna politika.

Župančič, J. 1948. *Kratice.* Ljubljana: Državna založba Slovenije.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

The Effectiveness of the "Real-Life-Based School Dictionary" as an Electronic Dictionary*

Ayça Çağlayan Özkan, *Department of Turkish, Guangdong University of Foreign Studies, P.R. China*
(aycacaglayanozkan@gmail.com)
(https://orcid.org/0000-0002-8870-818X)

Abstract: The Real-Life-Based School Dictionary (GYTOS) is designed to include the general vocabulary, terms, collocations, idioms, proverbs, synonyms and antonyms of Turkish for primary and secondary school students who are the target audience with monolingual, virtual and code generation and decoding functions. The corpus linguistics method was used in the creation phase of the dictionary and all lemmas were identified and structured on a real-life basis through the created corpus (https://okulsozlugum.com/).

The aim of the study was to determine the effect of GYTOS on the vocabulary development of primary and secondary school students. For this purpose, the students' level of experience in using the dictionary was first determined. Then, an achievement test was administered to students who used the dictionary and students who did not use the dictionary to statistically determine the effect of GYTOS on students' vocabulary development.

In the study, a 13-question questionnaire was used to determine students' experience of using GYTOS and it was found that they found GYTOS to be highly adequate. Furthermore, students stated that GYTOS was an easy-to-use and accessible resource. As a result of the achievement test, it was found that the scores of two different groups of students at the same level and with the same number of users of GYTOS from the same achievement test showed a statistically significant difference.

Keywords: REAL-LIFE-BASED SCHOOL DICTIONARY, EFFECTIVENESS, SCHOOL DICTIONARY, LANGUAGE SKILLS

Opsomming: Die effektiwiteit van die "Real-Life-Based School Dictionary" as 'n elektroniese woordeboek. Die Real-Life-Based School Dictionary (GYTOS) is ontwerp om die algemene woordeskat van Turks in te sluit, asook terminologie, kollokasies, idiome, spreekwoorde, sinonieme en antonieme met eentalige, virtuele en kodegenererings- en dekoderingsfunksies vir laer- en hoërskoolleerders, wat die doeltaalgehoor vorm. Die korpuslinguistiekmetode is in die samestellingsfase van die woordeboek gebruik en alle lemmas is aan die hand van die werklike lewe geïdentifiseer en gestruktureer deur die korpus wat saamgestel is (https://okulsozlugum.com/).

* This study is based on Tübitak-Sobag-1003 National Research Project titled "Real-Life-Based School Dictionary Applications"(Project No. 118K109).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

Die doel van die studie was om die effek van GYTOS op die woordeskatontwikkeling van laer- en hoërskoolleerders vas te stel. Daarom is die leerders se ervaringsvlak in die gebruik van die woordeboek eers vasgestel. Daarna is leerders wat die woordeboek gebruik het en dié wat dit nie gebruik het nie, aan 'n prestasietoets onderwerp om die effek van GYTOS op die leerders se woordeskatontwikkeling statisties vas te stel.

In die studie is 'n vraelys met 13 vrae gebruik om leerders se ervaring in die gebruik van GYTOS vas te stel. Daar is bevind dat hulle GYTOS hoogs toereikend vind. Daarbenewens het leerders verklaar dat GYTOS 'n bron is wat maklik is om te gebruik en toeganklik is. Uit die prestasietoets is afgelei dat die punte van die twee verskillende groepe leerders op dieselfde vlak en met dieselfde aantal gebruikers van GYTOS van dieselfde prestasietoets 'n statisties beduidende verskil toon.

Sleutelwoorde: REAL-LIFE-BASED SCHOOL DICTIONARY, EFFEKTIWITEIT, SKOOLWOORDEBOEK, TAALVAARDIGHEDE

1. Introduction

Dictionaries have always been used as important reference works in language teaching. In this sense, they are indispensable for both native speakers and learners of a language as a foreign language. Today, it is possible to find dictionaries displaying a variety of features, depending on the expectations of the users and the purpose of the dictionary. In this context, researchers have classified dictionaries according to their features. Considering these features, it is clear that dictionaries are classified according to the number of languages they contain, the medium in which they are published, their size, function, ordering of the articles, scope and user group (Bekdaş 2017). One of the types of dictionaries that have been proposed in terms of dictionary typology is school dictionaries. For these dictionaries prepared for school-age students, terms such as school dictionary, children's dictionary, high school dictionary and college dictionary are used in the literature (Hartmann and James 1998). The concept in question corresponds to "a dictionary designed for school-age children". Dictionaries corresponding to this concept generally contain a controlled vocabulary, are supported by visuals and have a simple design (Özkan 2022).

It can be said that the Real-Life-Based School Dictionary (GYTOS), whose effect on vocabulary teaching is emphasised in the study, is a monolingual electronic dictionary that includes the general vocabulary of Turkish, terms, collocations, idioms, proverbs, synonyms and antonyms for the target audience with its code generation and decoding function.

The literature emphasises the need for school dictionaries. It is stated that the need to use dictionaries increases with age and grade level, especially in secondary schools, and that the intensification of course content as grade level increases, and thus the expansion of the vocabulary of the texts used in class, leads to this need. In this sense, students stated that they use dictionaries more frequently each year in order to understand the words in the texts whose meanings they do not know (Melanlıoğlu 2013).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

Researchers also stated that there are some shortcomings in the existing school dictionaries. It was stated that a school dictionary should not only be a source that contains the meaning of words but also a source that shows the pronunciation, form and usage characteristics of words (Akcan 2012). On the other hand, Melanlıoğlu (2013) emphasised that the problems students see in dictionaries should be given in bold letters, examples of the connotative and figurative meanings of words should be given, and visual elements should be included. It was also stated that school dictionaries are not created on the axis of a criterion (Çotuksöken 1999).

In general, it can be said that school dictionaries, just like general-purpose dictionaries, have deficiencies in terms of preparation and presentation criteria. In this sense, an original and real-life-based school dictionary can be created as an educational/teaching material by using a special purpose corpus that includes texts from natural language environments such as textbooks, children's literature products, periodical children's publications, etc. that children may encounter from primary school to high school. In this sense, it can be said that the Real-Life-Based School Dictionary, which is based on information technologies, is easily accessible, and has qualities that are different and fully meet the needs when other school dictionaries are considered (Özkan 2022).

1.1 GYTOS as a school dictionary

A special domain corpus was used to construct the Real-Life-Based School Vocabulary. The special field corpus includes two main layers. The first layer of the corpus is the Turkish Children's Literature Corpus, which is a special domain corpus in which Turkish Children's Literature is the subject of qualitative and quantitative research. The "Turkish Children's Literature Special Area Corpus", which contains 8,639,522 million lexemes, has been built using the principles and methods of corpus linguistics. Textbooks approved by the Ministry of Education were included in the second layer of the corpus. In this sense, the second layer of the corpus includes textbooks and periodicals containing 2,047,135 lexemes.

GYTOS contains a total of 49,279 items, which were sourced from the aforementioned corpus of these lemmas. 13,801 are enriched with visuals/graphics. 13,801 of these items were enriched with visuals/graphics and the pronunciation of each item was added with a voice button in the word search interface of the dictionary.

In addition, this GYTOS contains 4,712 idioms and their explanations, 2,847 proverbs and their explanations, 1,090 abbreviations, 9,698 personal names and their meanings, 18,422 tagged synonyms, 3.919 tagged antonyms, 760 misspelled words, the most frequently encountered spelling rules and 6 different games such as meaning puzzles, idiom puzzles, proverb puzzles, close meaning puzzles, antonym puzzles and correct spelling puzzles (see https://okulsozlugum.com/).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

In this context, in order to determine the effect of GYTOS on vocabulary development, the level of students' experience was first determined and secondly, its effect on students' vocabulary development was attempted to be determined with an achievement test.

2. Aim of study

The aim of this study is to determine the extent to which GYTOS meets the expectations of the target group and whether GYTOS contributes to the development of the language skills of the target group, primary and secondary school students.

3. Method

The research used a 13-question evaluation questionnaire which was applied to determine the extent to which GYTOS met the expectations of the target audience. Two groups were formed from students studying in a public secondary school. There are equal numbers of students in both groups and both groups have students from the same grade level (5th and 7th grades). The students in the

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

first group were allowed to use and analyse GYTOS, while the students in the second group were not given a dictionary. Through this 13-question questionnaire, GYTOS was asked to be evaluated by the students using it. In the first stage, the results of this questionnaire are given.

In the second stage of the study, an achievement test was administered to determine the effect of GYTOS on language skills and the results were evaluated using descriptive statistics. The achievement test was administered to 5th graders (40 students with a dictionary and 40 students without a dictionary) and 7th graders (43 students with a dictionary and 43 students without a dictionary). T-test was applied on the achievement scores obtained from this test and the differences were tried to be evaluated statistically.

4. Findings and interpretation

4.1 GYTOS in terms of user (student) experiences

Another concept that appears in lexicographic studies is "user research". User research is carried out to find out what kind of users have what kind of expectations of the dictionary to be presented to the user, what the user finds useful in a dictionary and what the user does not find useful in a dictionary. Jackson (2002: 163) emphasises that the target group that will use the dictionary should be identified before the planning stage. This stage can be carried out in a number of ways, such as field research, surveys, observation or obtaining expert opinion (Atkins and Rundell 2008: 30). Who will be the target users of the dictionary, the age range, the level of education and the purpose of use will reveal a plane that will influence the basic structure of the dictionary. (Factors such as the target user group of the dictionary, age range, education level, and purpose of use reveal a ground that will affect the basic structure of the dictionary.) According to Bowern (2008), the readership, the format of the dictionary, the content of the dictionary, the way in which lexical items are conveyed (indicating the type of word, whether construction and derivation suffixes are shown, whether dependent morphemes are included, etc.), the stages of dictionary preparation and publication are among the most important elements that should be taken into account when creating a dictionary.

In relation to the evaluation of dictionaries by users, user research is also concerned with the attitudes of users towards the dictionary and whether the dictionary has an impact on vocabulary. In this sense, it was ensured that GYTOS was used and experienced by the students. Then, students were asked to evaluate GYTOS.

The survey questions and responses regarding the dictionary experiences of students who are stakeholders in GYTOS are presented below (see Table 1 and Table 2). A total of 83 students participated in the survey, 40 of whom were 5th grade students and 43 of whom were 7th grade students.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

Table 1: 5th Grade students' answers

Grade 5 Answers (40 Students)			
Questions	Yes	No.	Partially
1. The dictionary's screen designs are adequate.	%70.00	%7.50	%22.50
2. In a dictionary, explanations of the meaning of words are sufficient.	%70.00	%7.50	%22.50
3. In the dictionary, the pronunciation of the article titles is sufficient	%57.50	%22.50	%20.00
4. In the dictionary, witness sentences are in number to explain the meaning.	%85.00	%5.00	%10.00
5. In the dictionary, collocation/joining notations are sufficient.	%75.00	%5.00	%20.00
6. The explanatory visuals in the glossary are sufficient.	%50.00	%32.50	%17.50
7. Explanatory texts in the dictionary are sufficient.	%72.50	%5.00	%22.50
8. I was able to find other words in the dictionary.	%62.50	%10.00	%27.50
9. The colors on the dictionary screens are compatible.	%57.50	%20.00	%22.50
10. I could easily find the word I was looking for in the dictionary.	%70.00	%10.00	%20.00
11. The dictionary is easy to use.	%77.50	%10.00	%12.50
12. The dictionary can be accessed at any time.	%67.50	%7.50	%25.00
13. I had a hard time using the dictionary.	%7.50	%80.00	%12.50

Table 2: 7th Grade students' answers

Grade 7 Answers (43 Students)			
Questions	Yes	No.	Partially
1. The dictionary's screen designs are adequate.	%53.49	%20.93	%25.58
2. In a dictionary, explanations of the meaning of words are sufficient.	%65.12	%13.95	%20.93
3. In the dictionary, the pronunciation of the article titles is sufficient.	%60.47	%13.95	%25.58
4. In the dictionary, witness sentences are in number to explain the meaning.	%72.09	%2.33	%25.58
5. In the dictionary, collocation/joining notations are sufficient.	%65.12	%13.95	%20.93
6. The explanatory visuals in the glossary are sufficient.	%67.44	%13.95	%18.60

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

7. Explanatory texts in the dictionary are sufficient.	%60.47	%6.98	%32.56
8. I was able to find other words in the dictionary.	%58.14	%11.63	%30.23
9. The colors on the dictionary screens are com patible.	%60.47	%9.30	%30.23
10. I could easily find the word I was looking for in the dictionary.	%60.47	%18.60	%20.93
11. The dictionary is easy to use.	%74.42	%18.60	%25.58
12. The dictionary can be accessed at any time.	%67.44	%4.65	%27.91
13. I had a hard time using the dictionary.	%16.28	%74.42	%9.30

An analysis of the responses to the 13-question questionnaire designed to determine the students' experience of using the dictionary shows that a high percentage of students (when the sum of yes and partial responses is taken into account) found the screen designs, meaning explanations, pronunciation feature, witness sentences, collocation/joining feature, explanatory text and explanatory images to be sufficient. Again, students found GYTOS easy to use and stated that they had no difficulty using it. They also stated that they could access the dictionary at any time and that they could easily find the words they were looking for in the dictionary. On the other hand, the students also stated that the colours of the dictionary screens were in harmony with each other.

4.2 The effectiveness of GYTOS and its impact on vocabulary teaching

Another dimension of user research is the effectiveness of the dictionary. The extent to which a dictionary is effective in educational settings and whether it contributes to students' vocabulary can also be revealed through user research. In this framework, the effectiveness of a new, original and real-life-based school dictionary as a teaching/learning aid was attempted to be shown, based on corpus linguistic and lexicographic practices. This research on effectiveness was carried out on the basis of the difference between the scores of the two groups in the language proficiency test (see Appendix 1) given to students who used GYTOS and those who did not.

The achievement test was administered to 5th graders (40 students with and 40 students without the dictionary) and 7th graders (43 students with and 43 students without the dictionary). The mean, standard deviation and standard error values of the descriptive statistical calculations revealing the basic statistical properties of the data obtained in the study are given in the table below (Table 3).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

Table 3: Descriptive statistics

	Group		Statistics	Std. Error
Grade 5 Test Scores	Students who don't use the dictionary	Mean	49.4750	1.93516
		Std. Deviation	12.23905	
	Students who use the dictionary	Mean	71.4000	1.96599
		Std. Deviation	12.43403	
Grade 7 Test Scores	Students who don't use the dictionary	Mean	54.5349	2.06493
		Std. Deviation	13.54068	
	Students who use the dictionary	Mean	69.5814	2.42272
		Std. Deviation	15.88685	

According to the table above, the mean score of the fifth-grade students who answered the common test without using a dictionary was 49, the mean score of those who used a dictionary was 71, and the standard deviation of both groups was 12. The difference between the mean scores is quite large, 22 points. However, the standard deviation of both groups is almost the same. This shows that the deviation values of the scores of both groups from the mean are equal.

When we look at the descriptive statistics of seventh-grade students, the mean score of those who did not use a dictionary while answering the common test was 54, while the mean score of those who used a dictionary was approximately 70. There is a difference of 16 points between the two groups in the expected direction. In addition, the standard deviations of the two groups are also different. It is noteworthy that especially the seventh-grade students who used a dictionary had a higher deviation from the group mean than the others. This situation can be evaluated separately.

In order to decide whether the analyzed data are suitable for independent sample T-test, it is necessary to check the assumption of normal distribution of difference scores. The results of the normality tests and T-tests for the fifth and seventh-grade levels, respectively, are presented below (see Table 4).

Table 4: Normality test of 5th Grade

	Group	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Grade 5 Test Scores	Students who don't use the dictionary	.107	40	.200*	.980	40	.695
	Students who use the dictionary	.119	40	.157	.965	40	.256

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

The results of the Shapiro-Wilk test (.695 and .256, p>.05) and the Kolmogorov-Smirnov test (.200 and .157, p>.05) on the difference in the scores of the exam results of two groups of fifth grade students who used GYTOS and those who did not use GYTOS were both greater than the reliability level of .05. Therefore, it can be assumed that the research data are normally distributed. Therefore, the data are suitable for a T-test. As the aim was to show the difference between two independent groups on a single variable, it was decided to conduct an independent sample T-test (see Table 5).

Table 5: Group statistics of 5th Grade

	Group	N	Mean	Std. Deviation	Std. Error Mean
Grade 5 Test Scores	Students who don't use the dictionary	40	49.4750	12.23905	1.93516
	Students who use the dictionary	40	71.4000	12.43403	1.96599

It can be seen that the mean scores of the fifth-grade students who use a dictionary are significantly higher than the mean scores of those who do not use a dictionary. However, in order to make this difference statistically clearer, it is necessary to look at the T-test results (see Table 6).

Table 6: Independent sample T-test of 5th Grade

		Levene's Test for Equality of Variances		T-test for Equality of Means				
		F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference
Test Scores of 5th Graders	Equal variances assumed	.115	.735	-7.948	78	.000	-21.92500	2.75862
	Equal variances not assumed			-7.948	77.981	.000	-21.92500	2.75862

When the above findings are analyzed, it is seen that there is a significant difference between the test scores of fifth grade students who use dictionaries and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

those who do not use dictionaries [t(78)=.000, p<.05]. Based on this result, it can be argued that the number of correct answers given by the two different groups of students at the same level and with the same number of correct answers to the same test questions showed a statistically significant difference (see Table 7).

Table 7: Normality test of 7th Grade

	Group	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Test Scores of 7th Graders	Students who don't use the dictionary	.107	43	.200*	.961	43	.148
	Students who use the dictionary	.121	43	.118	.948	43	.052

Both the results of the Shapiro-Wilk test (.148 and .052, p>.05) and the results of the Kolmogorov-Smirnov test (.200 and .118, p>.05) on the difference in the test scores of the two groups of seventh-grade students who used GYTOS and those who did not were greater than the .05 level of reliability. Since the research data showed normal distribution and met the normality assumption, it was accepted that the data were suitable for the application of T-test. As the aim was to show the difference between two independent groups on a single variable, it was decided to conduct an independent sample T-test (see Table 8).

Table 8: Group statistics of 7th Grade

	Group	N	Mean	Std. Deviation	Std. Error Mean
Test Scores of 7th Graders	Students who don't use the dictionary	43	54.5349	13.54068	2.06493
	Students who use the dictionary	43	69.5814	15.88685	2.42272

It is seen that the mean scores of the seventh-grade students who use a dictionary are higher than the mean scores of those who do not use a dictionary (see Table 9).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

Table 9: Independent sample T-test of 7th Grade

		Levene's Test for Equality of Variances		T-test for Equality of Means				
		F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**
Test Scores of 7th Graders	Equal variances assumed	.105	.746	-4.727	84	.000	-15.04651	3.18332
	Equal variances not assumed			-4.727	81.943	.000	-15.04651	3.18332

When the results of the independent sample T-test are analyzed, it is clear that there is a significant difference between the exam scores of seventh-grade students who use dictionaries and those who do not use dictionaries [t(84)=.000, p<.05]. Therefore, it can be said that the use of dictionary created a statistically significant difference in the number of correct answers given by two different groups with the same level and equal number of students to the same exam questions.

The dictionary was also tested in classroom settings and it was concluded that it has the potential to meet user needs to a high degree. This result was obtained by recruiting students. In the study conducted with 5th and 7th grade students, it was statistically determined that there were significant differences in the achievement tests administered between those who used GYTOS and those who did not.

The study found that there was a significant difference between the test scores of fifth grade students who used a dictionary and those who did not [t(78)=.000, p<.05].

Again, it was observed that the exam scores of seventh-grade students who used a dictionary and those who did not use a dictionary showed a significant difference [t(84)=.000, p<.05]. Based on this result, it can be said that the use of GYTOS shows a statistically significant difference in the scores of two different groups of students at the same level and with the same number of students on the same achievement test.

5. Conclusion

Scientific methods should be used at every stage of a dictionary, from its creation to its delivery to the end user. It would not be wrong to say that the dictionaries created so far contain word lists and configurations at an intuitive

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

level. In this sense, it can be said that GYTOS is a dictionary created with modern lexicographic methods. On the other hand, whether the dictionary is effective in educational environments and whether it can contribute to the users' vocabulary is a situation that needs to be determined separately.

It can be said that GYTOS, which was tested in course environments, has the potential to meet user needs to a high degree. It was found that there was a significant difference between the exam scores of the fifth grade students who used GYTOS and those who did not use a dictionary [t(78)=.000, p<.05]. Again, a significant difference was found between the test scores of seventh grade students who used a dictionary and those who did not [t(84)=.000, p<.05]. Based on this result, it can be said that the use of GYTOS shows a statistically significant difference in the scores of two different groups of students at the same level and with the same number of students on the same achievement test.

6. References

Akcan, İ.P. 2012. Türkçe Öğretiminde Sözlük Kullanımı: Türkçe Çalışma Kitaplarında Sözlüğe Gönderen Alıştırmalar. Aksan, M. and Y. Aksan (Eds.). 2012. *Türkçe Öğretiminde Güncel Çalışmalar*: 93-102. Ankar: Şafak Yayınları.

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography.* Oxford/New York: Oxford University Press.

Bekdaş, M. 2017. Sözlük Türlerinin Tasnifi (Classification of Dictionary Types). *Route Educational and Social Science Journal* 4(4): 52-59.

Bowern, C. 2008. *Linguistic Fieldwork: A Practical Guide.* New York: Palgrave Macmillan.

Çotuksöken, Y. 1999. Okul Sözlükleri Üzerine. *Kebikeç* 7–8: 61-65.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography.* London/New York: Routledge.

http://okulsozlugum.com/. Date of access: 2-4-2024.

Jackson, H. 2002. *Lexicography: An Introduction.* London/New York: Routledge.

Melanlıoğlu, D. 2013. Ortaokul Öğrencilerinin Sözlük Kullanma Alışkanlıkları: Nitel Bir Araştırma. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi* 2(2): 266-284.

Özkan, B. 2022. Tübitak-Sobag-1003 National Research Project titled "Real-Life-Based School Dictionary Applications" (Project No. 118K109).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

Appendix 1: Achievement test

	2. Grup (Sözlüksüz)
	Sınıfı:
	Okulu:
1.	*Mustafa Kemal'in dünyaya geldiği günlerde Osmanlı Devleti; siyasi, sosyal ve ekonomik alanda sıkıntılı bir **dönem**den geçiyordu.* Cümlede geçen "**dönem**" kelimesinin eş/yakın anlamlılarını yazınız. ..... "**dönem**" kelimesi başka kelimelerle birlikte kullanılabilir. Örneğin: "**alışma dönemi**", "**çocukluk dönemi**" gibi. "**dönem**" kelimesi başka hangi kelimelerle birlikte kullanılabilir. Yazınız. .....
2.	*Dünyadaki büyük buluşlar, zaferler, kahramanlıklar hayal gücümüzün **eseridir**. "İnsan dünyada hayal ettiği sürece yaşar." demiş şair. Demek ki insanoğlu için hayal kurmak başarıya giden yolu açabilecek bir güce sahiptir.* Metinde geçen "**eser**" kelimesinin eş/yakın anlamlılarını yazınız. ..... "**eser**" kelimesi başka kelimelerle birlikte kullanılabilir. Örneğin: "**sanat eseri**", "**eser vermek**" gibi. "**eser**" kelimesi başka hangi kelimelerle birlikte kullanılabilir. Yazınız. .....
3.	**"Göz açıp kapayıncaya kadar."** deyiminin anlamı aşağıdakilerden hangisidir? **A) Uzunca bir zaman.** **B) Kısa bir süre içinde.** **"Göz açıp kapayıncaya kadar."** deyimini cümle içerisinde kullanınız. .....
4.	**"Dile kolay."** deyiminin anlamı aşağıdakilerden hangisidir? **A) Söylemesi kolay ancak yapması çok güç.** **B) Isırdıktan sonra yutması zor olmayan.** **"Dile kolay."** deyimini cümle içerisinde kullanınız. .....

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2006 (Lexiconote)

5.	**"bandaj"** nedir? …… **"bandaj"** hakkında bilgi veriniz. …… …… …… ……
6.	**"cirit"** nedir? …… **"cirit"** hakkında bilgi veriniz. …… …… …… ……
7.	**"ansızın"** kelimesinin *eş/yakın anlamlıları* nelerdir? Yazınız. ……
8.	**"zar zor"** kelimesinin *karşıt/zıt anlamlıları* nelerdir? Yazınız. ……
9.	Bir deyim yazıp anlamını açıklayınız. **Deyim:**…… **Anlamı:**……
11.	Bir *yazım/imla kuralı* yazarak örnek veriniz. **Kural:** …… **Örnek:**……
12.	Yazımı/imlası karıştırılan kelimelere örnekler veriniz. ……
13.	Bir atasözü yazıp anlamını açıklayınız. **Atasözü:**…… **Anlamı:**……
14.	Adınızın anlamı nedir? Yazınız. ……

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

A Comparative Study on the Effectiveness of AI Chatbots and Dictionary Apps for Lexical Tasks and Retention

Yuzhen Chen, *College of Foreign Languages, Putian University, Fujian, P.R.C. (287323222@qq.com) (https://orcid.org/0009-0004-6431-7768)*

Abstract: This study compared an AI chatbot (Kimi) and a bilingual dictionary app (NCD) in supporting vocabulary tasks among Chinese junior English majors. Sixty-six participants used either Kimi or NCD to complete both receptive and productive lexical tasks. Questionnaires gathered user feedback on tool use, and a surprise retention test assessed long-term vocabulary retention one week later. Results showed that Kimi significantly outperformed NCD in vocabulary comprehension, collocation production, and productive knowledge retention. Additionally, Kimi demonstrated more consistent performance than NCD across all test items, highlighting its reliability. The study underscores the potential of AI chatbots to address language-related queries and enhance vocabulary acquisition. It also advocates for aligning technological advancements with pedagogical goals to optimize language learning tools and create a sustainable learning environment.

Keywords: AI CHATBOT, DICTIONARY APP, VOCABULARY RECEPTION, COLLOCATION PRODUCTION, RETENTION

Opsomming: 'n Vergelykende studie van die effektiwiteit van KI-kletsbotte en woordeboektoepasssings in leksikale take en retensie. In hierdie studie word 'n KI-kletsbot (Kimi) en 'n tweetalige woordeboektoepassing (NCD) vergelyk ten opsigte van die steun wat hulle aan Chinese junior studente met Engels as hoofvak in woordeskatopdragte bied. Ses-en-sestig deelnemers het óf Kimi óf NCD gebruik om beide reseptiewe en produktiewe leksikale opdragte te voltooi. Gebruikersterugvoer oor hulpmiddelgebruik is met behulp van vraelyste ingesamel, en 'n week later is die langtermynwoordeskatretensie in 'n onverwagse retensietoets bepaal. Die resultate het getoon dat Kimi beduidend beter as NCD in woordeskatbegrip, kollokasieproduksie en produktiewe kennisretensie gevaar het. Daarbenewens het Kimi in al die toetsitems konsekwenter vertoon as NCD, wat die betroubaarheid van Kimi beklemtoon. Hierdie studie benadruk die potensiaal van KI-kletsbotte om taalverwante vrae te hanteer en om woordeskatverwerwing te verbeter. Dit bepleit ook die belyning van tegnologiese vooruitgang met pedagogiese doelwitte om hulpmiddels vir die aanleer van taal te optimaliseer en 'n volhoubare leeromgewing te skep.

Sleutelwoorde: KI-KLETSBOT, WOORDEBOEKTOEPASSING, WOORDESKATRESEPSIE, KOLLOKASIEPRODUKSIE, RETENSIE

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

1. Introduction

The field of Artificial Intelligence (AI), especially Natural Language Processing (NLP), has achieved significant breakthroughs since the emergence of transformer-based architectures, exemplified by generative Large Language Models (LLMs) such as OpenAI's GPT series, Google's PaLM, Meta's LLaMA and Baidu's Ernie. These models possess unprecedented application potential and have brought disruptive impacts to scientific research, technological innovation, and daily life (Yuan 2023: 8). In the field of lexicography, a new age of the successful application of generative AI has dawned (De Schryver 2023: 1). An increasing number of scholars are investigating how LLMs can reshape lexicographical practices and the roles of dictionaries and lexicographers in the AI era. However, a critical area remains under-researched: the effectiveness of AI chatbots in resolving language-related uncertainties and enhancing vocabulary acquisition.

Recent studies have presented mixed results regarding AI's impact on vocabulary comprehension and production. For instance, Rees and Lew (2024) found no significant difference between AI-generated and human-authored definitions in reading comprehension tasks. In contrast, Ptasznik et al. (2024) and Lew et al. (2024) revealed the advantages of AI chatbots over monolingual online dictionaries in lexical reception and production. Despite these findings, the long-term impact of AI chatbots on vocabulary retention remains unexplored. This gap points to the need for further research to evaluate the role of intelligent tools in meeting learners' lexical needs through comparative studies from a user perspective.

2. Literature review

2.1 Generative AI and lexicography

Current research on generative AI in lexicography primarily explores three dimensions. The most prominent strand examines the capabilities and limitations of AI models in producing lexicographical content, e.g. De Schryver and Joffe (2023) who successfully integrated ChatGPT into the TLex dictionary system, Phoodai and Rikk (2023) who revealed ChatGPT's superiority in structuring entries for high-frequency words, capturing nuanced elements like phonetic transcription and morphological details, and McKean and Fitzgerald (2024) who found that AI models do not meet human editorial standards, necessitating significant human oversight (see also De Schryver 2023, Jakubíček and Rundell 2023, Lew 2023 and Rundell 2023).

The second dimension explores the opportunities, challenges, and future trajectory of AI in lexicography, investigating the dynamics of human-AI collaboration and AI's capabilities to streamline workflows and increase productivity. Tarp and Nomdedeu-Rull (2024) and Huete-García and Tarp (2024) researched

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

the development of a Spanish AI writing assistant project, underscoring the complexity of human-machine interaction and the crucial role of human lexicographers in maintaining quality, while Zhao (2023) indicated that LLMs can cluster concordance lines and streamline semantic categorization, thereby significantly reducing lexicographers' workload (see also Fuertes-Olivera 2024 and Lew 2024).

The last and relatively understudied but arguably most relevant dimension for this article, gauges the efficacy of AI chatbots in addressing real-world vocabulary inquiries from a user perspective. Rees and Lew (2024) divided 43 students into three groups for a reading comprehension task. One group used definitions from the *Macmillan English Dictionary* (MED), another used AI-generated definitions, and a third had no dictionary access. The results indicated that the MED group surpassed the non-dictionary group in performance. However, there were no significant differences between the MED and AI groups, or between the AI and non-dictionary groups. Furthermore, no notable variations were found in the time taken to complete the task between the two groups exposed to definitions. Nonetheless, the authors maintained that AI's benefits for language learners, educators, translators, and lexicographers remain attainable (Rees and Lew 2024: 65).

Ptasznik et al. (2024) investigated the utility of AI in facilitating vocabulary reception and production compared with the online *Longman Dictionary of Contemporary English* (LDOCE). The research involved 223 university students divided into two groups, with one group using ChatGPT and the other resorting to LDOCE. These participants performed two tasks: Translating 20 low-frequency English words from contextualized sentences into Polish, and converting twenty Polish sentences, with the English verbs highlighted, into English. The results indicated ChatGPT's superior performance in both receptive and productive tasks. Additionally, ChatGPT was faster in the production task than LDOCE. The study positions AI chatbot as a formidable competitor to traditional dictionaries and advocates for hybrid strategies that preserve learner autonomy.

Lew et al. (2024) assessed the effectiveness of ChatGPT against a monolingual dictionary (LDOCE) and a popular bilingual dictionary (Diki.pl) in production and reception tasks mirroring Ptasznik et al.'s (2024) experimental design. The findings showcased ChatGPT's enhanced effectiveness over both dictionaries in the production task, while in the reception task, it surpassed the monolingual but not the bilingual dictionary. The study confirmed that a general-purpose chatbot such as ChatGPT can be a viable alternative to traditional dictionaries in both production and reception tasks (Lew et al. 2024: 8).

The research conducted by Ptasznik et al. (2024) and Lew et al. (2024) not only emphasizes the prospects of AI chatbots in language learning but also raises important questions about the future role of traditional dictionaries. As intelligent tools become more prominent, further research is needed to probe how these new technologies can be integrated into language learning and how they compare with traditional dictionaries in addressing users' lexical reference needs.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

In particular, efforts should be made to explore how AI chatbots affect learners' long-term vocabulary retention, an area that remains uncharted so far.

2.2 Dictionary use in lexical tasks

While AI tools offer new possibilities, traditional dictionaries remain a cornerstone in lexical learning, as evidenced by numerous studies on their effectiveness in vocabulary tasks.

Researchers have long examined the utility of dictionaries in vocabulary acquisition by comparing different conditions under which lexical tasks are fulfilled. In a study conducted by Chen (2012), participants were divided into three groups to complete a reading comprehension task that focused on ten target lexical items. Each group used either a paper dictionary, an electronic version of the same dictionary, or no dictionary at all. The results indicated that both groups using diction significantly outperformed the group that did not use a dictionary. This suggests that dictionaries are more effective for vocabulary comprehension than relying solely on contextual cues.

Several other studies have confirmed the beneficial contribution of dictionaries in lexical learning. Li and Xu (2015) observed substantial improvements in task performance after consulting an online dictionary for meaning determination. Alzi'abi (2017) noted increased appropriate responses when electronic dictionaries were used in a verb-adverb collocation task. Chen (2017, 2020) reported that the use of CALL (Computer-Assisted Language Learning) and online dictionaries significantly improved learners' accuracy in producing target collocations.

The impact of dictionary medium on lexical learning has also been a focal area. Chen (2010) compared pocket electronic dictionaries and paper dictionaries in receptive and productive tasks. Despite the faster processing speed of electronic dictionaries, there was no significant difference in vocabulary comprehension, production, or retention between the two types. Dziemianko (2010, 2017) evaluated the efficacy of electronic and paper dictionaries in decoding, encoding, and retaining target vocabulary. Results were mixed, with one study reporting advantages for electronic dictionaries, while the other found no substantial differences. These discrepancies emphasize the complexity of understanding the effects of different dictionary forms on vocabulary learning.

Beyond the medium, researchers have also investigated how specific design features in electronic dictionaries affect vocabulary learning outcomes. Lew and Doroszewska (2009) revealed that the availability of L1 equivalents, either alone or with L2 definitions, strongly predicted word retention. Dziemianko (2015) highlighted the efficacy of color-coded functional labels in accelerating search efficiency and enhancing information recall. Further advancing this line of inquiry, Dziemianko's (2022) comparative analysis identified line drawings as the most pedagogically impactful visual format, demonstrating their dual capacity to streamline cognitive processing and reinforce long-term memory consolidation.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

From the literature review, it is apparent that the investigation of dictionaries' contribution to vocabulary acquisition predominantly focuses on electronic dictionaries. However, with the rise of generative AI, there is growing ambiguity regarding user preferences, the efficacy of AI models in meeting learners' reference needs, and the competitiveness of electronic dictionaries compared with AI chatbots. As we transition into the AI era, it becomes crucial to undertake comparative studies that examine the relative advantages and synergies between electronic dictionaries and intelligent tools within language learning.

3. Study design

Inspired by Ptasznik et al. (2024), this study aims to compare an AI chatbot and a dictionary app in vocabulary tasks. It employs a Chinese AI model and a bilingual dictionary app, differing from ChatGPT and LDOCE, the tools used in the original study. Notably, the study involves an investigation into vocabulary retention.

3.1 Research questions

The research seeks to address the following questions.

RQ 1: Which tool more effectively enhances the comprehension of target words in a receptive lexical task, the Kimi intelligent assistant (Kimi henceforth) or the mobile app of *New Century English–Chinese Chinese–English Dictionary* (NCD henceforth)?
RQ 2: Which tool provides more effective support for the production of target items in a productive lexical task, Kimi or NCD?
RQ 3: Which tool leads to superior retention of target items, Kimi or NCD?

3.2 Participants

The study involved 75 junior English majors from a Chinese university, with 67 females and 8 males aged 21–22. All participants had eight to nine years of EFL learning experience and had passed the Test for English Majors (TEM, Band Four), a national test to evaluate learners' English proficiency. None had used the NCD app before, and only a few had used Kimi, though they were familiar with other AI chatbots or dictionary apps.

3.3 Instruments

The study adopted a pre-test, main test and post-test design. Each phase incorporated two paper-based lexical tasks — one receptive and the other productive — which remained uniform throughout all three stages (see Appendix A).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

The receptive task

In this task, participants were asked to provide meanings of the target words highlighted within sentences, either in Chinese or English. For example, "*She was led away in a state of* ***stupor***". These target words were chosen from the 10,000-word level of the Vocabulary Levels Test (Schmitt et al. 2001: 87-88) based on several criteria. Firstly, they were deemed unfamiliar to the participants, necessitating the use of tools for task completion. Secondly, they spanned various parts of speech, encompassing content words of significant pedagogical relevance. Lastly, they were not morphologically complex, not obsolete, technical or specialized terms, thereby reducing retention difficulties. After consulting with three teachers from the participants' classes, ten words were chosen: *dabble*, *scrawl*, *vie*, *stint*, *wily*, *squirm*, *torrid*, *banter*, *swagger*, and *translucent*.

Example sentences for the task were sourced directly from either online *Oxford Advanced Learner's Dictionary* (OALD, https://www.oxfordlearnersdictionaries.com/) or LDOCE (https://www.ldoceonline.com), both renowned globally for their exceptional quality. It was ensured that the sentences in the task did not overlap with those provided by NCD, confirmed through trial consultations, or those produced by Kimi in response to prompts like "Please give an example sentence for (the target word)" or "how to use (the target word) in a sentence".

The productive task

In the productive task, participants had to complete missing verbs in verb–noun collocations within sentences, with meanings provided in Chinese, e.g., "*I could see he was trying to ___ a fight with me (寻衅)*". The selection of these collocations was guided by several criteria: (1) mutual information score >3 from the British National Corpus, indicating strong collocation strength; (2) verbs and nouns are among the top 2000 most common words but rarely combined; (3) covered in NCD and retrievable via Kimi; (4) semantically transparent, with meanings inferable from components; (5) some collocations and their translation are congruent while others are not; (6) excluding specialized, formal, or informal collocations to focus on general use. These criteria ensured that the collocations were representative, unfamiliar to most participants, and held significant experimental relevance. Ultimately, ten target collocations were chosen: *write a cheque, lift a ban, blow a kiss, cut a tooth, jump the queue, wear perfume, say a prayer, pick a fight, break a habit, take a joke*. Sentences for the task were similarly sourced from the online versions of OALD and LDOCE, avoiding overlap with NCD or Kimi-generated examples.

It is worth noting that the current study featured only 10 target items per task — half the amount in Ptasznik et al.'s (2024) research. This reduction was necessary, as participants were required to retain the target items, and memory capacity is inherently limited.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

The questionnaire survey

Two concise semi-structured questionnaires, each with eight questions, were designed to gather participants' feedback on Kimi or NCD (see Appendix B). They focused on: (1) perceived quality and quantity of retrieved information, (2) tool usability, (3) challenges encountered during tool usage, and (4) comparative advantages and disadvantages relative to alternative resources.

Despite their brevity, the surveys demonstrated adequate reliability and validity. Both employed structured, closed-ended questions with clearly defined response options, such as Likert scales and multiple-choice formats, which enhanced their internal consistency. Moreover, the inclusion of open-ended questions enriched the construct validity by allowing for the capture of more nuanced user perspectives.

Tools for Task Assistance

For the study, two tools were selected: Kimi and NCD, based on accessibility and participants' dictionary usage habits.

Given the unavailability of ChatGPT at the author's institution, Kimi, a domestic AI chatbot in China, was chosen as an alternative out of several considerations. To start with, it is a free AI chatbot with high accessibility, making it a practical choice for academic and research purposes. Secondly, it has earned wide recognition for its reliability and robust capabilities. Furthermore, Kimi excels in text interpretation and processing, particularly in Chinese, which aligns well with the needs of Chinese users and the context of this study.

Surveys point to the fact that dictionary apps are very popular among Chinese EFL learners (Liu et al. 2019, Ma 2019, Liu et al. 2021). Using these apps in the study reflects participants' real-life dictionary habits. The chosen app, NCD, is the first of its kind in China that integrates a prestigious L2–L1 dictionary (the *New Century English–Chinese Dictionary,* 2016) with an L1–L2 one (the *New Century Chinese–English Dictionary,* 2nd edition, 2016). It offers an innovative "two-in-one" functionality with a convenient "jump" feature for bidirectional searches.

The choice of Kimi and NCD was also informed by participants' dictionary habits. Extensive research has confirmed that most EFL learners in China prefer bilingual dictionaries over monolingual ones. NCD and Kimi both offer bilingual functionality, fitting well with participants' habits and preferences.

3.4 Procedure

The study lasted three weeks. Participants were randomly assigned to the Kimi or NCD group. In Week One, they attended 15-minute preparatory sessions: One group was granted free access to download NCD on smartphones and received training on its features, functions, and consultation methods, while the other group learned to use Kimi, focusing on composing prompts for lexical inquiries.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

In Week Two, participants took a pretest to assess their prior knowledge of target lexical items, performing receptive and productive tasks without reference tools. After the pretest, they completed a main test with the same tasks but in a randomized order to reduce carry-over effects. One group used NCD, while the other interacted with Kimi. Participants had a 20-minute time limit and were not allowed to use other reference sources. Afterward, they filled out a 5-minute questionnaire. In Week Three, a surprise delayed retention test was conducted without reference tools.

3.5 Data processing

Each participant received six scores: three for receptive tasks and three for productive tasks across the pre-test, main test, and retention test. The maximum score for each task was ten points, with one point for each correct response. Minor errors like spelling or verb inflections were ignored. SPSS 27 was employed for data analysis.

The pre-test results affirmed that most participants were unfamiliar with the test items. Only four participants knew the meaning of two or more target words, and five had prior productive knowledge of target collocations. To ensure uniform familiarity, these nine students were excluded from further analysis, leaving a final sample of 66 participants (33 in each class). The extremely low pre-test scores were not included in the analysis. Questionnaire responses were analyzed both quantitatively and qualitatively.

4. Results

4.1 The receptive task

The main test results, presented in Table 1, indicate that students using Kimi achieved perfect scores on the receptive task (M = 10.000, SD = 0.000), while the performance of the NCD group exhibited slight variability (M = 9.697, SD = 0.529). An independent-samples *t*-test revealed a statistically significant difference between the two groups (t = 3.288, p = 0.002, df = 64, two-tailed; Cohen's d = 0.81) (Table 2), suggesting that Kimi surpassed NCD in the receptive main test. As manifested in Figure 1, Kimi achieved a 100% success rate, whereas NCD experienced a few noticeable dips.

Table 1: Group Statistics (Max = 10) (receptive task, main test)

	Class	N	Mean	Std. Deviation	Std. Error Mean
Score	Kimi	33	10.000	0.000	0.000
	NCD	33	9.697	0.529	0.092

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Table 2: Independent Samples Test (receptive task, main test)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Score	Equal variances assumed	80.188	0.000	3.288	64	0.002	0.303	0.092	0.119	0.487
	Equal variances not assumed			3.288	32.000	0.002	0.303	0.092	0.115	0.491

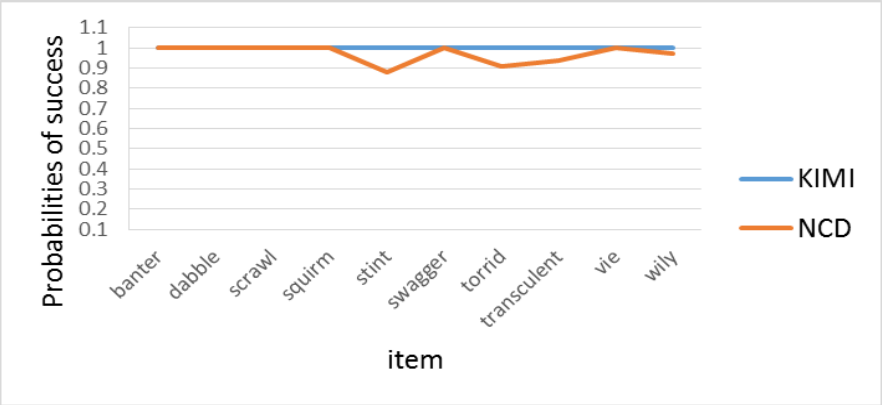


Figure 1: Probabilities of success for each test item in the receptive main test

In the retention test (Table 3), both groups obtained a comparable retention rate of target words. The Kimi group scored 7.667 out of 10 (76.7%), while the NCD group garnered 7.576 out of 9.697 (78.1%). An independent-samples *t*-test found no significant difference between the groups (t = 0.174, p = 0.863, df = 62.197, two-tailed) (Table 4). As depicted in Figure 2, while overall performance was similar, Kimi demonstrated more consistent retention patterns, whereas NCD showed greater variation across test items.

Table 3: Group Statistics (Max = 10) (receptive task, retention test)

	Class	N	Mean	Std. Deviation	Std. Error Mean
Score	Kimi	33	7.667	2.300	0.400
	NCD	33	7.576	1.937	0.372

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Table 4: Independent Samples Test (receptive task, retention test)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Score	Equal variances assumed	0.850	0.360	0.174	64	0.863	0.091	0.524	-0.954	1.136
	Equal variances not assumed			0.174	62.197	0.863	0.091	0.524	-0.955	1.137

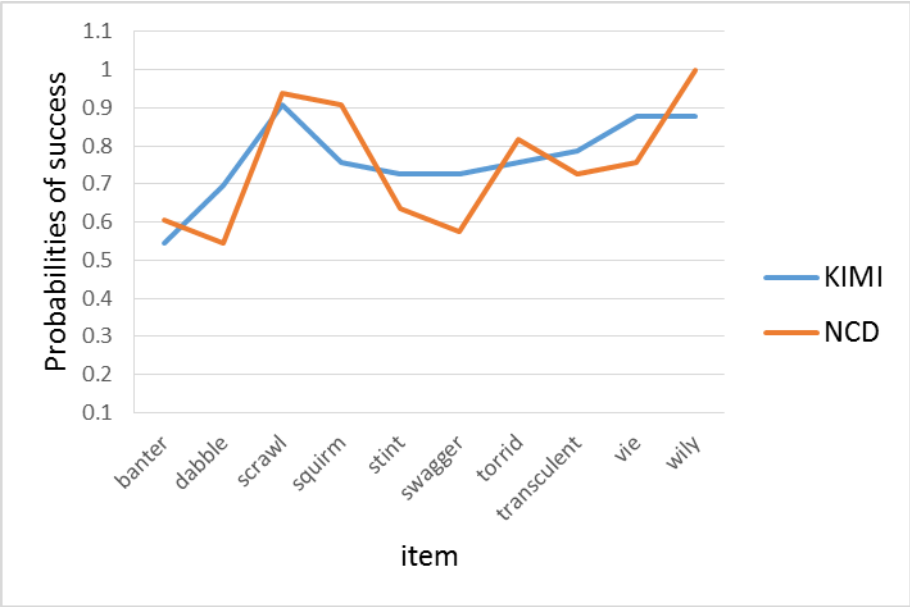


Figure 2: Probabilities of success for each item in the receptive retention test

Table 5 displays the mean score differences for each test item in the main and retention tests as well as detailed performance differences between the groups.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Table 5: Mean difference in each test item in main and retention tests (receptive task)

Test item	Tool	Mean score (main test)	Mean difference (main test)	Mean score (retention test)	Mean difference (retention test)
dabble	Kimi	1.00	0.00	0.70	0.15
	NCD	1.00		0.55	
scrawl	Kimi	1.00	0.00	0.91	0.03
	NCD	1.00		0.94	
vie	Kimi	1.00	0.00	0.88	0.12
	NCD	1.00		0.76	
stint	Kimi	1.00	0.12	0.73	0.09
	NCD	0.88		0.64	
wily	Kimi	1.00	0.03	0.88	0.07
	NCD	0.97		0.81	
squirm	Kimi	1.00	0.00	0.76	-0.15
	NCD	1.00		0.91	
torrid	Kimi	1.00	0.09	0.76	-0.06
	NCD	0.91		0.82	
banter	Kimi	1.00	0.00	0.55	-0.06
	NCD	1.00		0.61	
swagger	Kimi	1.00	0.00	0.73	0.15
	NCD	1.00		0.58	
translucent	Kimi	1.00	0.06	0.79	0.06
	NCD	0.94		0.73	

In the main test, Kimi users achieved full scores across all target words, while NCD users obtained marginally lower on four items, with differences ranging from 0.03 to 0.12 points. Analysis identified two main reasons for these errors: incorrect sense selection and partial comprehension of definitions. Some students chose the wrong sense for polysemous words. For instance, for "torrid," three students chose the wrong sense (sense 3, 热情似火的) in the sentence "They face a torrid time in tonight's game," despite the correct sense (sense 4, 难熬的，艰难的) being available, probably due to a mismatch with the contextual clues. Another example is "stint," where four students mistranslated it as "工作 (work)","分配 (allotment)", or "disturbed task" instead of understanding it as a "period of work" as a result of incomplete comprehension of the definition. These findings imply that the effectiveness of vocabulary tools depends on users' proficiency in utilizing them.

In the retention test, both groups had a balanced performance, with each outperforming the other on certain items. The highest retention scores were observed for "scrawl" (Kimi: 0.91, NCD: 0.94), while the lowest were for "banter" (Kimi: 0.55, NCD: 0.61). Despite being morphologically simpler, "banter" was less remembered than "scrawl". This may be due to the sentence contexts: "He scrawled his name at the bottom" provides clearer context than "He enjoyed exchanging banter with the customers", which allows for varied interpretations due to ambiguous context. Incorrect responses included terms like "交易 (transaction)", "货物" (goods)", "讨价还价 (bargain)", "友善" (kindness)", "零钱" (changes)", "柜台 (counter)", and "小道消息 (gossip)". It seems the context of a word can significantly influence its retention, irrespective of the tools employed.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

4.2 The productive task

In the main test (Table 6), the Kimi group scored higher (M = 8.697, SD = 0.951) than the NCD group (M = 7.939, SD = 1.144) in the productive task. An independent-samples *t*-test indicated a significant difference between the groups (t = 2.925, p = 0.005, df = 64, two-tailed; Cohen's d = 0.72) (Table 7). As evident from Figure 3, Kimi maintained consistent success rates across items, while NCD displayed more fluctuation in performance.

Table 6: Group Statistics (Max = 10) (productive task, main test)

	Class	N	Mean	Std. Deviation	Std. Error Mean
Score	Kimi	33	8.697	0.951	0.166
	NCD	33	7.939	1.144	0.199

Table 7: Independent Samples Test (productive task, main test)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Score	Equal variances assumed	0.015	0.902	2.925	64	0.005	0.758	0.259	0.240	1.275
	Equal variances not assumed			2.925	61.944	0.005	0.758	0.259	0.239	1.275

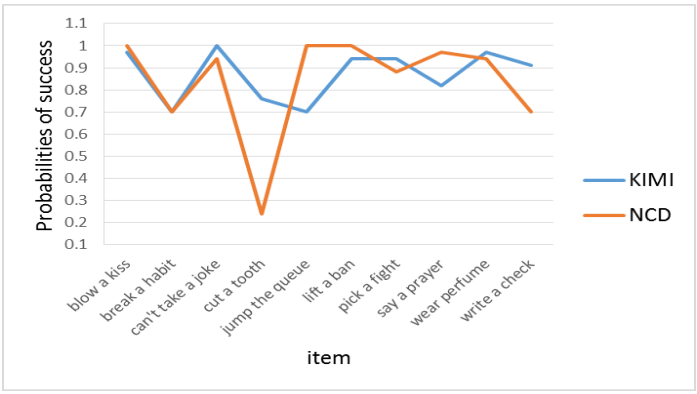


Figure 3: Probabilities of success for each item in the productive main test

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

In the retention test (Table 8), the Kimi group retained significantly more productive collocational knowledge (6.152/8.697, 70.7%) compared with the NCD group (3.606/7.939, 45.4%), with a 25.3% difference in retention rates. An independent-samples *t*-test confirmed a large, statistically significant advantage for Kimi (t = 4.085, df = 64, p < 0.001, two-tailed; Cohen's d = 1.006) (Table 9). Figure 4 further depicts Kimi's consistent retention versus NCD's item-specific fluctuations.

Table 8: Group Statistics (Max = 10) (productive task, retention test)

	Class	N	Mean	Std. Deviation	Std. Error Mean
Score	Kimi	33	6.152	2.949	0.513
	NCD	33	3.606	2.030	0.353

Table 9: Independent Samples Test (productive task, retention test)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Score	Equal variances assumed	7.042	0.010	4.085	64	0.000	2.545	0.623	1.300	3.790
	Equal variances not assumed			4.085	56.770	0.000	2.545	0.623	1.297	3.793

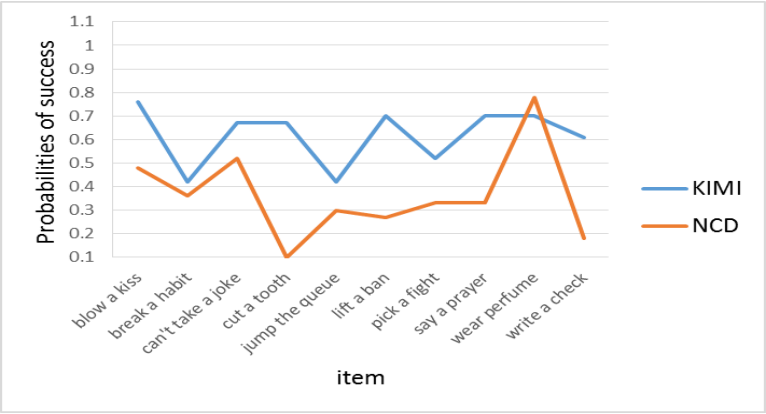


Figure 4: Probabilities of success for each item in the productive retention test

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Compared with the receptive task, the productive task was more challenging for participants due to the higher cognitive demand required for production. Table 10 illustrates the differential performance of the Kimi and NCD groups on each target collocation in both the main and retention tests. In the main test, Kimi users scored higher on some collocations, whereas NCD users excelled in others. However, in the retention test, the Kimi group consistently outperformed the NCD group across all target collocations, with score differences between 0.04 and 0.43 points.

Table 10: Mean difference in each test item in main and retention tests (productive task)

Test item	Tool	Mean score (main test)	Mean difference (main test)	Mean score (retention test)	Mean difference (retention test)
插队 jump the queue	Kimi	0.70	- 0.30	0.42	0.12
	NCD	1.00		0.30	
开支票 write a cheque	Kimi	0.91	0.21	0.61	0.43
	NCD	0.70		0.18	
寻衅 pick a fight	Kimi	0.94	0.06	0.52	0.19
	NCD	0.88		0.33	
祷告 say a prayer	Kimi	0.82	- 0.15	0.70	0.37
	NCD	0.97		0.33	
戒掉习惯 break a habit	Kimi	0.67	-0.03	0.42	0.06
	NCD	0.70		0.36	
解除禁令 lift a ban	Kimi	0.94	0.06	0.70	0.42
	NCD	1.00		0.28	
长牙齿 cut a tooth	Kimi	0.88	0.43	0.48	0.09
	NCD	0.45		0.39	
送飞吻 blow a kiss	Kimi	0.97	- 0.03	0.76	0.18
	NCD	1.00		0.48	
抹香水 wear perfume	Kimi	0.97	0.03	0.70	0.04
	NCD	0.94		0.66	
开不起玩笑 can't take a joke	Kimi	1.00	0.04	0.67	0.15
	NCD	0.94		0.52	

In the main test, the Kimi group scored lowest on "戒掉习惯" (0.67 points) and "插队" (0.70 points). For "戒掉习惯," 11 students wrote "quit a habit" instead of the correct "break/kick a habit" generated by Kimi. Since "quit" is a direct translation of "戒掉," these students might have instinctively paired "quit" with "habit," considering it a natural collocation. This suggests that they did not seek Kimi's assistance.

For "插队", nine students in the Kimi group chose "cut" instead of "jump," resulting in the incorrect collocation "cut the queue." Actually, Kimi provided two translations: "cut in line" and "jump the queue", along with detailed explanations and sentence examples. However, these students were not sufficiently attentive when reading the information.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

The NCD group scored very low on "长牙齿" (0.45 points), with only 19 out of 33 students providing correct responses. Eight students used "teethe," creating the erroneous phrase "teethe another tooth." According to the questionnaire, some students reported difficulty locating the collocation in NCD, as it was placed in a less accessible Sentence Bank Module which requires additional navigation. This user-unfriendly arrangement clearly affected the students' performance.

Another low-scoring collocation was "开支票" (0.7 points), with eight students using "invoice a cheque" instead of "write a cheque." These students probably knew that "invoice" means "开支票" and assumed "invoice a cheque" is an acceptable combination. Had they consulted the noun "cheque" in NCD, they would have found the correct collocation.

In the retention test, the NCD group had the lowest scores on "开支票" (0.18 points) and "解除禁令" (0.28 points). For "开支票," ten students left it unanswered, and others wrote incorrect phrases like "open a cheque" or "invoice a cheque." For "解除禁令", four left it blank, and others used wrong verbs like "abandon the ban", "stop the ban", "forbid the ban," or "loose the ban". These low retention scores may be attributed to the incongruent nature of these two collocations between English and Chinese.

4.3 Responses to the questionnaires

Questionnaire responses regarding Kimi

Among the 33 respondents, 63.6% (n = 21) had no prior experience with Kimi, though some were aware of it. Only a few used it regularly (15.2%, n = 5) or occasionally (18.9%, n = 6) for EFL learning, with one participant using it solely for non-linguistic purposes. In contrast, 60.6% (n = 20) regularly used other AI chatbots for EFL learning, primarily for tasks including vocabulary acquisition, grammar clarification, translation, writing support, speech drafting, text analysis, and error correction. A smaller subset (33.3%, n = 11) used AI tools sporadically for language learning, while two respondents reported non-linguistic applications.

Regarding lexical information provision for tasks in the main test, 69.7% (n = 23) affirmed that Kimi generated all required data, while 30.3% (n = 10) noted near-complete coverage. Satisfaction levels with the quality of information were notably high: 33.3% (n = 11) expressed strong satisfaction, 63.6% (n = 21) reported satisfaction, and only one respondent remained neutral.

All participants found Kimi easy to access, with 63.6% (n = 21) describing it as "convenient" and 36.4% (n = 12) as "highly convenient." Nevertheless, while 63.6% (n = 21) experienced smooth interactions, 33.3% (n = 11) reported uncertainty in formulating effective prompts, and one noted insufficient content depth. Compared with traditional dictionary apps, 75.8% (n = 25) of respondents rated Kimi as "far superior", 18.2% (n = 6) as "slightly better", and two considered it "comparable".

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

When it comes to the strengths and weaknesses of AI chatbots versus traditional dictionary apps, a clear consensus emerged among the respondents. They generally agreed that AI chatbots have three main advantages. To begin with, AI tools can provide prompt and instantaneous responses to queries, which is highly convenient. In addition, they can engage users in interactive learning through conversations, making the learning process more engaging and tailored to individual needs. Lastly, AI tools can offer a broader range of information than any single dictionary app.

Conversely, AI tools also have three main disadvantages compared with dictionary apps. First, the quality of the generated content depends on the prompts provided by users. Inadequate prompts may result in simplistic or superficial information. Secondly, while dictionaries present all relevant lexicographical information in a single entry, chatbot users must formulate prompts to elicit specific information. Thirdly, AI chatbots typically provide only textual information for lexical reference, whereas dictionary apps often offer visual aids such as images and audio pronunciations, which can enhance the learning experience.

Questionnaire responses regarding NCD

According to the questionnaire on NCD, all the 33 respondents reported no prior experience with this tool. In terms of dictionary information accuracy, the majority rated it positively: 33.3% (n = 11) described it as "very good," and 57% (n = 19) deemed it "good," while only three participants (9.1%) considered it "neutral." Similarly, assessments of the richness of dictionary content confirmed broad satisfaction, with 21.2% (n = 7) expressing "high satisfaction" and 57.6% (n = 19) indicating "satisfaction." The remaining 21.2% (n = 7) found it "average".

Regarding usability, 30.3% (n = 10) of respondents were "very satisfied" with NCD's interface and functionality, while 48.5% (n = 16) found it "satisfactory." A minority (21.2%, n = 7) felt the usability was merely "neutral". When asked about the effectiveness of NCD for vocabulary tasks, 27.3% (n = 9) reported being "very satisfied," and 66.7% (n = 22) expressed general satisfaction. Only two participants (6%) remained neutral.

Despite these positive trends, 66.7% of respondents (n = 22) faced challenges while using NCD dictionary use. Common issues included the absence of full-sentence translation capabilities, difficulties locating English equivalents for specific Chinese collocations, and lack of English definitions alongside Chinese ones. However, a strong majority (78.8%, n = 26) still considered NCD superior to mainstream alternatives like Youdao, Eduict, and Baidu.

Suggestions for improvement were provided by 81.8% (n = 27) of participants. Key recommendations included expanding vocabulary coverage, integrating sentence-level translation features, enhancing access to collocations and phrases, redesigning the interface for greater intuitiveness, adding English explanations for Chinese entries, and incorporating tools to support word retention.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

A comparative analysis

From the questionnaire responses, several similarities were identified between Kimi and NCD users. Despite having limited or no prior experience with the respective tools, both groups generally provided positive evaluations regarding the quality and quantity of retrieved information, as well as the usability and usefulness of the tools for task fulfilment. They found the tools more useful for task fulfilment than other dictionary apps they usually use. Moreover, most users in both groups had balanced perceptions of the strengths and limitations of the tools. These findings may partially explain the better results of the current study relative to previous research, as will be discussed in the next section.

Significant differences also emerged between Kimi and NCD users. Kimi users reported higher levels of satisfaction with the quality of output and the usability of the tool compared with NCD users. For example, 100% of Kimi users rated the tool as convenient or highly convenient, while seven NCD users rated it as neutral. Each group encountered different obstacles in using the tools. In addition, NCD users found dictionary use more challenging than Kimi users. These disparities likely contributed to the better performance of Kimi users in vocabulary comprehension, production and productive retention.

5. Discussion

5.1 Tool effectiveness for vocabulary reception

To address the first research question, the study offers convincing proof that Kimi significantly outperforms NCD in facilitating vocabulary comprehension. On one hand, the advantages of Kimi may chiefly stem from its powerful capabilities in generating interactive and contextually relevant content in response to users' inquiries. These features enable users to understand word meanings more accurately and efficiently than traditional dictionaries, leading to better comprehension outcomes. On the other hand, NCD users exhibited inadequate dictionary usage skills, which hindered their ability to leverage the dictionary for lexical reception. As detailed in Section 4.1, some students were unable to discern the appropriate meaning of a polysemous word based on contextual cues and some failed to grasp word definitions accurately. These inadequacies obviously impacted on the dictionary's effectiveness for vocabulary reception.

The current study achieved higher success rates for both tools in the receptive task (Kimi: 100%, NCD: 97.0%) compared with Ptasznik et al.'s (2024) research, where the dictionary group had a 73% mean success rate and ChatGPT attained 87%. This discrepancy primarily stems from task difficulty differences. Ptasznik et al.'s task was more challenging, featuring more complex target words (e.g., "sycophant," "circumlocution", "equanimity," "boondoggle") and twice as many items (twenty vs. ten in the current study). Additionally, the sentences employed exhibited greater syntactic complexity and length.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

5.2 Tool effectiveness for vocabulary production

In answering the second research question, the study illustrates that Kimi notably surpasses NCD in supporting vocabulary production, maintaining a more consistent performance across ten target collocations. These results may be attributed to Kimi's advanced capabilities to generate context-specific suggestions through interactive dialogue. Users can receive immediately relevant feedback, free from error-prone and time-consuming dictionary searches (Ptasznik et al. 2024: 334). In contrast, NCD offers a broader range of options without the same contextual guidance, resulting in more variability in student performance.

The study identifies several reasons for the relatively low scores on certain test items in the productive task. Participants often viewed collocations as arbitrary word pairings rather than conventionalized patterns. This misconception led some students to overestimate their collocation knowledge and underestimate the need for reference tools. In addition, a lack of attentiveness when interpreting accessed information resulted in avoidable mistakes. Moreover, the unintuitive layout and presentation of some entry information in NCD made it difficult for students to quickly and accurately locate what they needed.

The current study demonstrated higher success rates in the productive task relative to Ptasznik et al. (2024), with Kimi achieving 87% and NCD 79% efficacy, surpassing their reported results of 53% for the dictionary group and 81% for ChatGPT. Similar to the receptive task, the improved outcomes in production may be attributed to differences in task design. The current study focused on ten collocations comprising high-frequency verbs and nouns, whereas Ptasznik et al.'s research involved twenty highly polysemous verbs in complex verb complementation patterns, such as "see something out," "carry something off," "play up to somebody," and "fix somebody up with somebody". Evidently, the simpler nature of the current productive task reduced cognitive demand, thereby enhancing performance.

5.3 Tool effectiveness for vocabulary retention

Regarding the third research question, the study offers interesting findings on the tools' effectiveness for lexical retention. For receptive retention, no significant difference was found between the Kimi and NCD groups, indicating that both tools are equally effective in helping users remember word meanings. Specifically, NCD users achieved a slightly higher retention rate (78.13%) than Kimi users (76.67%), probably as a result of inter-learner strategy variation. Nevertheless, the latter exhibited more consistent performance across ten target collocations.

In the productive task, Kimi achieved a retention rate nearly 25% higher than NCD (70.7% vs. 45.4%), affirming its superior ability to help users recall collocations. The Kimi group also exhibited more consistent performance across different items. These differences may be ascribed to Kimi's strong interactivity, which

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

can reduce users' cognitive load and enhance their engagement with the tool, thereby improving memory retention.

Two factors were identified as influencing word retention. One is the context in which a word appears. It seems more specific context cues improve retention by helping students associate words with relevant contexts. The other is L1–L2 congruency. Incongruent collocations between English and Chinese, which lack direct equivalents, tend to have lower retention rates due to the difficulty in establishing meaningful connections.

Compared with the author's previous study (Chen 2017), which used an electronic dictionary and reported a retention rate of 36.1%, the current study reaped much better results (Kimi: 70.7%; NCD: 45.4%). This improvement is due to the advanced technological tools employed in the current study, which offer more sophisticated features than the traditional electronic dictionary used previously.

6. Conclusion

The study provides compelling evidence for the superior effectiveness of Kimi over NCD in facilitating vocabulary comprehension and production, echoing the findings of Ptasznik et al. (2024) and Lew (2024) on the advantages of AI chatbots over traditional dictionaries. Notably, Kimi excels in retaining productive collocation knowledge. These results highlight the transformative potential of generative AI tools as valuable complements to conventional lexicographical resources in language-learning contexts. Beyond vocabulary acquisition, AI chatbots like Kimi can play a pivotal role in diverse linguistic activities, including translation, writing, and reading, by offering real-time, context-aware assistance that adapts to learners' needs.

However, the effectiveness of these tools hinges on users' ability to craft precise prompts, a skill that requires explicit training. To maximize the utility of AI chatbots, it is crucial to integrate AI literacy modules into language curricula. These modules should cover prompt engineering, critical evaluation of AI-generated content, and ethical considerations. Meanwhile, developers should refine AI algorithms, enhance computing power, and improve inference abilities to better understand user inputs and meet their needs. Furthermore, incorporating multimodal features like audio-visual aids could significantly enhance the usability of AI models.

While AI chatbots hold great potential for language learning, learners should caution against overreliance. These tools can sometimes generate inaccurate, biased or even hallucinated content. Traditional dictionaries, on the other hand, still remain indispensable because they can guarantee linguistic accuracy and preserve nuanced semantic distinctions. Future research should focus on combining AI's interactive features with the reliable editing standards of dictionaries. It is also important to gauge the long-term effects of using AI tools on learners' ability to think independently and critically. Such efforts will ensure that tech-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

nological advancements align with pedagogical goals, creating a balanced and sustainable environment for language learning.

Meanwhile, in the era of AI, to boost the competitiveness of dictionary apps against AI-powered resources, interdisciplinary collaboration is crucial. Experts from lexicography, education, information technology, and particularly the AI industry should work together to develop AI-enhanced dictionaries. The form, content, access structure, and functionalities of dictionaries need substantial upgrades to withstand the competitive onslaught of intelligent tools and ensure their continued relevance and utility. The integration of AI technologies can facilitate the provision of more contextually relevant examples, interactive learning features, and personalized recommendations, thereby significantly improving the overall learning experience for users.

7. Limitations and suggestions for future research

The study is not without its limitations. It focused on outcomes rather than the processes of tool use. Recording and analysing students' prompts could provide deeper insights into how they construct queries and use AI-generated responses. Future research could explore in depth the interaction process between AI models and users. This could employ qualitative research methodologies, such as case studies, interviews, observational studies, and think-aloud protocols to capture the varied experiences of individuals as they interact with these tools.

Another limitation of the study is that, each task included only ten target items, which may limit the generalizability of the findings. Considering the limited memory capacity and attention span of learners within a relatively short period of time, it is reasonable to involve twenty lexical items in a twenty-minute task session. However, the limited number of target items might influence the representativeness of results. Future research could explore the effects of varying the number of target items within certain time frames to determine the optimal balance between quantity and retention efficiency.

Furthermore, the study employed a between-group design rather than a within-group design. While this approach allowed for a comparison of different tools across distinct groups, it may have overlooked the nuanced differences in how the same learners interact with and evaluate multiple tools. Future research could adopt a within-group design to compare the same group's usage patterns and evaluations of different tools within the same set of tasks. This would provide a deeper grasp of each tool's strengths and weaknesses, as well as how learners adapt their strategies when switching between tools.

To extend the research scope of the present study, further investigations could involve a diverse array of language learning activities, such as reading, writing, and translation. By employing a variety of chatbots and dictionaries with differing features and capabilities, researchers could gain more insights into the efficacy of these tools across various linguistic tasks and learner profiles.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Given the rapid advancement of AI technologies, it is also imperative to explore the evolving role of dictionaries in language learning. As AI continues to permeate and transform the landscape of language resources, the traditional functions and usage patterns of dictionaries are likely to undergo significant shifts. Research should therefore focus on tracking and analysing the changes in users' dictionary use habits and preferences over time, as well as the potential impact of these shifts on language learning outcomes.

Acknowledgements

The author extends deep gratitude to the anonymous peer reviewers and André du Plessis, editor of *Lexikos*, for their invaluable comments and suggestions on the previous drafts of this article. Particular appreciation is directed to the participants of the study for their indispensable contributions. Special thanks are also due to my colleagues, Dr. Lin Lin and Ms. Wang Rui for their assistance in collecting the experimental data.

Funding disclosure statement

This work was supported by the Fujian Social Science Foundation under Grant FJ2023B031.

GenAI use disclosure statement

During the manuscript preparation, the author utilized Kimi (a free LLM accessible at https://kimi.moonshot.cn/) solely for enhancing readability and language polishing. The author carefully reviewed and edited all AI-generated content and assumes full responsibility for the final text.

References

Alzi'abi, S.E. 2017. Guessing Verb–Adverb Collocations: Arab EFL Learners' Use of Electronic Dictionaries. *Lexikos* 27: 50-77.

Chen, Yuzhen. 2010. Dictionary Use and EFL Learning: A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.

Chen, Yuzhen. 2012. Dictionary Use and Vocabulary Learning in the Context of Reading. *International Journal of Lexicography* 25(2): 216-247.

Chen, Yuzhen. 2017. Dictionary Use for Collocation Production and Retention: A CALL-based Study. *International Journal of Lexicography* 30(2): 225-251.

Chen, Yuzhen. 2020. The Effectiveness of Dictionary Use for Collocation Production and Retention: A Case Study of Smartphone Online Dictionaries. *Lexicographical Studies* 2: 20-31.

De Schryver, G.-M. 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

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/.

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. 2015. Colors in Online Dictionaries: A Case of Functional Labels. *International Journal of Lexicography* 28(1): 27-61.

Dziemianko, A. 2017. Dictionary Form in Decoding, Encoding and Retention: Further Insights. *ReCALL* 29(3): 335-356.

Dziemianko, A. 2022. The Usefulness of Graphic Illustrations in Online Dictionaries. *ReCall* 34(2): 218-234.

Fuertes-Olivera, P.A. 2024. Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos* 34(1):123-140.

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. *Lexikos* 34(1): 21-40.

Jakubíček, M. and M. Rundell. 2023. The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-editing Lexicography? Medveď, M. et.al. (Eds.). 2023. *Electronic Lexicography in the 21st century (eLex2023): Invisible Lexicography. Proceedings of the eLex 2023 Conference, Brno, 27–29 June 2023:* 518-533. Brno: Lexical Computing CZ s.r.o.

Lew, R. 2023. ChatGPT as a COBUILD Lexicographer. *Humanities and Social Sciences Communications* 10(704):1-10.

Lew, R. 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11(1): 1-8.

Lew, R. and J. Doroszewska. 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.

Lew, R., B. Ptasznik and S. Wolfer. 2024. The Effectiveness of ChatGPT as a Lexical Tool for English, Compared with a Bilingual Dictionary and a Monolingual Learner's Dictionary. *Humanities and Social Sciences Communications* 11(1): 1-10.

Li, Lingling and Hai Xu. 2015. Using an Online Dictionary for Identifying the Meanings of Verb Phrases by Chinese EFL Learners. *Lexikos* 25:191-209.

Liu, Dilin, Yaochen Deng and Shiyan Yang. 2021. Evaluating Popular Online English–Chinese Dictionaries in China by Applying Lew and Szarowska's (2017) Evaluation Framework. *International Journal of Lexicography* 34(2): 157-182.

Liu, Xiqin, Dongping Zheng and Yushuai Chen. 2019. Latent Classes of Smartphone Dictionary Users among Chinese EFL Learners: A Mixed-method Inquiry into Motivation for Mobile Assisted Language Learning. *International Journal of Lexicography* 32(1): 68-91.

Longman Dictionary of Contemporary English: https://www.ldoceonline.com

Ma, Qing. 2019. University L2 Learners' Voices and Experience in Making Use of Dictionary Apps in Mobile Assisted Language Learning (MALL). *International Journal of Computer-Assisted Language Learning and Teaching* 9(4): 18-36.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

McKean, E. and W. Fitzgerald. 2024. The ROI of AI in Lexicography. *Lexicography* 11(1): 7-27.

Oxford Advanced Learner's Dictionary: https://www.oxfordlearnersdictionaries.com

Ptasznik, B., S. Wolfer and R. Lew. 2024. A Learners' Dictionary Versus ChatGPT in Receptive and Productive Lexical Tasks. *International Journal of Lexicography* 37(3): 322-336.

Phoodai, C. and R. Rikk. 2023. Exploring the Capabilities of ChatGPT for Lexicographical Purposes: A Comparison with *Oxford Advanced Learners' Dictionary* within the Microstructural Framework. Medveď, M. et al. (Eds.). 2023. *Electronic Lexicography in the 21st Century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 Conference, Brno, 27–29 June 2023:* 345-375. Brno: Lexical Computing CZ s.r.o.

Rees, G.P. and R. Lew. 2024. The Effectiveness of OpenAI GPT-generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Orientated Reading Task. *International Journal of Lexicography* 37(1): 50-74.

Rundell, M. 2023. Automating the Creation of Dictionaries: Are We Nearly There? *Proceedings of the 16th International Conference of the Asian Association for Lexicography (Asialex 2023 Proceedings), 22–24 June 2023, Seoul, Korea: Lexicography, Artificial Intelligence, and Dictionary Users:* 9-17. Seoul: Yonsei University.

Schmitt, N., D. Schmitt and C. Clapham. 2001. Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test. *Language Testing* 18(1): 55-88.

Tarp, S. and A. Nomdedeu-Rull. 2024. Who has the Last Word? Lessons from Using ChatGPT to Develop an AI-based Spanish Writing Assistant. *Círculo de lingüística aplicada a la comunicación* 97: 309-321.

Yuan, Yulin. 2023. Theoretical Reflections on Linguistic Studies against the Background of AI Great Leap Forward. *Chinese Journal of Language Policy and Planning* 46(4): 7-18.

Zhao, Chong. 2023. *Research on Contextual Embedding for Sense Identification in Learner's Dictionaries.* Beijing: Beijing Foreign Studies University.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Appendix A

The receptive lexical task

Directions: Please provide the meaning of the target word that is highlighted in the sentence. You may explain it in either Chinese or English.

1. He **dabbles** in local politics.
   dabble : _____
2. He **scrawled** his name at the bottom.
   scrawl: _____
3. There are at least twenty restaurants **vying** with each other for custom.
   vie: _____
4. He hated his two-year **stint** in the Navy.
   stint: _____
5. The boss is a **wily** old fox.
   wily: _____
6. Christine **squirmed** uncomfortably in her chair.
   squirm: _____
7. They face a **torrid** time in tonight's game.
   torrid:_____
8. He enjoyed exchanging **banter** with the customers.
   banter: _____
9. He **swaggered** over towards me.
   swagger: _____
10. His skin was **translucent** with age.
    translucent: _____

The productive lexical task

Directions: Please complete each sentence by filling in the blank with a suitable verb in its correct form to complete an accurate verb–noun collocation. The intended meaning of the collocation is provided at the end of each sentence.

1. An argument developed when she tried to ___ **the queue**. (插队)
2. I had to ____ **a cheque** for £360 yesterday. （开发票）
3. I could see he was trying to___ **a fight** with me. (寻衅)
4. The children ____ their **prayers** and got into bed. (祷告)
5. I've smoked for years, but I really want to ___ **the habit**. (戒掉习惯)
6. There are no plans to ___ **the ban** on the sale of fireworks to children. (解除禁令)
7. Poor little Patrick was _____ **another tooth** and we had hardly had any sleep. （长牙齿）
8. As the train drew away he ____ her **a kiss**. （飞吻）
9. She was ____ too much **perfume**. （抹香水）
10. The trouble with her is she **can't** _____ **a joke**. （开不起玩笑）

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Appendix B

KIMI AI Assistant Usage Survey

Directions: This survey is designed to gather your feedback on your interaction with Kimi for the lexical tasks you completed just now. Please select your response (✔) or provide comments based on your recent experience. Thank you for your participation.

1. How often do you use KIMI?
   A. Never heard of it B. Heard of it but never used before
   C. Occasionally for non-language learning purposes D. Occasionally for English learning
   E. Frequently for English learning F. Other situations _____
2. How often do you use other AI models (such as ChatGPT, Baidu's Ernie, etc.)?
   A. Never
   B. Occasionally for non-language learning purposes
   C. Occasionally for English learning
   D. Frequently for English learning, such as _____
   E. Other situations _____
3. Did Kimi generate sufficient content needed for the lexical tasks just now?
   A. Yes, it provided all the content needed for the tasks.
   B. Yes, it generated most, but not all the needed information.
   C. No, it barely provided anything useful.
   D. No, it generated some irrelevant content.
4. How satisfied are you with the quality of the content generated by KIMI?
   A. Highly satisfied B. Satisfied C. Neutral D. Dissatisfied E. Extremely dissatisfied
5. How would you evaluate Kimi in terms of ease of use?
   A. Very convenient B. Convenient C. Neutral D. Difficult to access E. Slow in response
6. What obstacles did you encounter during your interaction with Kimi just now? (Multiple choices allowed)
   A. No obstacles, highly smooth interaction B. uncertainty about prompt formulation
   C. Slow response D. Inaccurate content
   E. Incomplete content F. Overloaded content
   G. Irrelevant content H. Complex interface
   I. Other issues _____
7. Compared with traditional dictionary apps, how would you rate KIMI's overall performance?
   A. Far Superior B. Slightly Better C. Comparable D. Slightly Worse E. Far Inferior
8. In your opinion, what are the key advantages and limitations of AI assistants (like Kimi, ChatGPT) compared with traditional dictionaries?
   Advantages: _____
   Limitations: _____

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2027 (Article)

Dictionary Application Usability Survey

Directions: This survey is designed to gather your feedback on the dictionary application you used to complete your tasks just now. Please select your response (✔) or provide comments below based on your recent experience. Thank you for your participation.

1. Have you used this dictionary application before?
   A. Yes B. No
2. How satisfied are you with the accuracy of the dictionary content? (e.g., definitions, examples, pronunciations)
   A. Extremely Satisfied B. Satisfied C. Neutral D. Dissatisfied E. Extremely Dissatisfied
3. How satisfied are you with the comprehensiveness of the information? (e.g., coverage of words, synonyms, usage notes)
   A. Extremely Satisfied B. Satisfied C. Neutral D. Dissatisfied E. Extremely Dissatisfied
4. How satisfied are you with the ease of searching/retrieving information? (e.g., search bar functionality, filters)
   A. Extremely Satisfied B. Satisfied C. Neutral D. Dissatisfied E. Extremely Dissatisfied
5. How satisfied are you with the overall effectiveness in assisting vocabulary tasks? (e.g., interface design, time efficiency, relevance of results)
   A. Extremely Satisfied B. Satisfied C. Neutral D. Dissatisfied E. Extremely Dissatisfied
6. Compared with other dictionary applications you regularly use, how would you rate this app's overall performance?
   A. Far Superior B. Slightly Better C. Comparable D. Slightly Worse E. Far Inferior
7. Did you encounter any difficulties while using this application?
   A. Yes (Please describe: _____)
   B. No
8. Do you have any suggestions for improving this application?
   A. Yes (Please specify: _____)
   B. No

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

Avoiding Information Overload — Ways of Exemplifying in Bilingual e-Dictionaries Depending on the Systematicity of the Linguistic Data

Alfonso Rascón Caballero, *Institute of Foreign Languages, Faculty of Philology, University of Vilnius, Vilnius, Lithuania (alfonso.rascon@flf.vu.lt) (https://orcid.org/0000-0003-4766-865X)*

Abstract: Although monolingual dictionaries have received more academic attention, bilingual dictionaries are essential for learners at all levels. Today's electronic dictionaries for language learning have no space limitations, so the information they offer can be exhaustive. This brings with it the risk of information overload — users may feel overwhelmed and find it difficult to find the information they are looking for. E-dictionary editors must therefore balance comprehensiveness with usability and decide which linguistic data to present. Since examples are the entry component that takes up the most space, several types of examples are considered, and the systematicity of the information to be presented is the criterion for adopting specific lexicographic solutions. Research shows that shorter examples, such as infinitive phrases, support generalisable patterns, while full sentences and supplementary examples highlight more complex, unpredictable uses. This article proposes an innovative type of example, the translated cotext item, to help users infer usage rules. Ultimately, the study emphasises that dictionaries should help learners to intuitively distinguish predictable structures from language-specific expressions, thereby enhancing their ability to use the target language idiomatically.

Keywords: INFORMATION OVERLOAD, USABILITY, ELECTRONIC BILINGUAL LEXICOGRAPHY, ENTRY COMPONENTS, DICTIONARY EXAMPLE, LANGUAGE SYSTEMATICITY

Opsomming: Die vermyding van inligtingoorlading — Metodes om toeligting in tweetalige e-woordeboeke volgens die sistematisiteit van die taalkundige data weer te gee. Alhoewel eentalige woordeboeke meer akademiese aandag geniet het, is tweetalige woordeboeke op alle vlakke noodsaaklik vir leerders. Eietydse elektroniese woordeboeke vir taalleer het geen spasiebeperkings nie, en die inligting wat hulle aanbied, kan omvattend wees. Dit bring die risiko van inligtingoorlading mee — gebruikers kan oorweldig voel en dit moeilik vind om die inligting waarna hulle soek, te identifiseer. E-woordeboekredakteurs moet dus omvattendheid met bruikbaarheid balanseer en besluit watter taalkundige data aangebied moet word. Aangesien voorbeeldmateriaal die data-inskrywing is wat die meeste ruimte in beslag neem, word verskeie tipes voorbeelde oorweeg, en die sistematisiteit van die inligting wat aangebied

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

moet word, is die maatstaf vir die goedkeuring van spesifieke leksikografiese oplossings. Navorsing toon dat korter voorbeelde, soos infinitiewe frases, veralgemenende patrone ondersteun, terwyl volsinne en aanvullende voorbeelde meer komplekse, onvoorspelbare gebruike uitlig. Hierdie artikel stel 'n innoverende tipe voorbeeld, die vertaalde koteksitem, voor om gebruikers te help om gebruiksreëls daarvan af te lei. Uiteindelik beklemtoon die studie dat woordeboeke leerders moet help om voorspelbare strukture intuïtief van taalspesifieke uitdrukkings te onderskei, om sodoende hul vermoë om die doeltaal idiomaties te gebruik, te verbeter.

Sleutelwoorde: INLIGTINGOORLADING, BRUIKBAARHEID, ELEKTRONIESE TWEETALIGE LEKSIKOGRAFIE, DATA-INSKRYWINGS, WOORDEBOEKVOORBEELD, TAALSISTEMATISITEIT

1. Introduction

In this article the language dictionary is considered as a reference work used as a tool for learning a foreign language (Dubois and Dubois 1971: 11; Tarp 2011: 223) and for improving language proficiency. Foreign language learners acquire second language (L2) skills in language courses, with language handbooks, learner's grammars, apps, and by exposure to the language. The (monolingual or bilingual) dictionary is one learning tool among others, but it is a qualified tool, because it is designed by specialists in the language or languages being described to show the user the meaning and usage of the lexical items of one or two paired languages. The dictionary is also an essential aid for translators.

Monolingual learner's dictionaries enjoy a well-deserved prestige, which has led to bilingual dictionaries being regarded as second-class works (Svensén: 2009: 19; Nomdedeu-Rull and Tarp 2018: 10). Yet bilingual dictionaries have always been the preferred aid in learning a foreign language and not only in the first stages (Béjoint 2004: 144-145; Corpas Pastor et al. 2001: 239-273; Bogaards 2003: 27; Adamska-Sałaciak 2006: 25; Frankenberg-Garcia 2011: 97; Lettner 2020: 20). While it is true that research on monolingual dictionaries is far superior to the attention paid to bilingual dictionaries (Béjoint 2004: 65-74; Lettner 2020: 12), it cannot be denied that quality bilingual dictionaries are now published (Oxford and Collins, for example). These dictionaries not only give the equivalents of headwords but also provide a large number of source language (SL) examples whose target language (TL) translation includes alternative equivalent to the one initially given, and even translations whose syntactic structure does not correspond to that of the SL example (see *supplementary example* below). Thus, in the example in the following entry the translation of the example does not include the equivalent, and the nature of the predicate is different:

> **tie** NOUN (obligation, constraint) atadura
> **the children are a tie** los niños atan mucho (literally 'children tie parents down a lot')
> (*Oxford Spanish*)

One of the aims of bilingual dictionaries is the contrastive representation of two linguistic systems — mainly lexically, but also grammatically and pragmati-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

cally — so that users learn one of the two languages with the help of a language they master (L1). Although they are considered as a first aid to learning a foreign language (Svensén 2009: 19), they should be compiled to the same high standards as their monolingual counterparts.

There is no need to emphasise the importance of electronic dictionaries today (Dziemianko 2018). The e-dictionary allows lexicographical data to be presented more explicitly than print dictionaries — e.g., without abbreviations. Additionally, since the given data is typically presented more vertically than horizontally on a screen, a sense of dispersion is created, as opposed to the concentration found in print dictionaries. It offers much faster access to information, which can be more comprehensive. However, the screen does not cover as much information as a page in a print dictionary — the latter contains many articles, while the screen does not always fit the entire entry, and the user loses the overview of the article and must scroll down (Torres del Rey 2009: 49). In principle, e-dictionaries have no space restrictions, because the capacity of the databases that support them is unlimited. This format gives free rein to exemplification, as the example is the component that takes up the most space and is the least condensed. This, however, is where the danger of information overload lies: excessive exemplification can clutter the page to the point that the user feels overwhelmed.

The expression *information overload* was popularised by Toffler (1970) in his book *Future Shock*.[1] He drew attention to the inability of human intelligence to process all the information it receives: "By classifying information, by abstracting and «coding» it in various ways, we manage to stretch these limits, yet ample evidence demonstrates that our capabilities are finite" (1970: 350).

Drucker (1968, 1989, 1993) developed the concept of *knowledge society*, initially referring to the field of business management, but it was soon to be extended to all areas of society with the introduction of ICT. It is precisely in this environment that information overload is most prevalent. Referring to electronic lexicography, Fuertes-Olivera and Tarp (2014: 63) state: "In recent years […] it has developed into an international plague, especially when browsing and searching for information on the Internet, where the user is increasingly stressed with too many data and frequently placed in a situation similar to the search for the proverbial needle in a haystack" (cf. Gouws and Tarp 2017: 394).

Lexicographers became enthusiastic about the possibilities offered by the electronic format, which allows much more information to be provided than in the print dictionary. However, just as compiling a paper dictionary requiring data selection and condensation, "devising an e-dictionary calls for smart information management and sensitive design on the part of the editors and the software engineers" (Atkins and Rundell 2008: 240, 139; cf. Gouws and Tarp 2017: 412). In today's lexicographic research, *user-friendliness* is a recurring concept, as it is a necessary criterion to avoid information overload, but it always clashes with the aim of providing exhaustive information on each lexical item.

The Collins dictionary is regarded as one of the best bilingual dictionaries, at least as far as English–Spanish pair is concerned, and can be used for free on

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

the Internet.[2] It offers exhaustive information, with many well-chosen and adapted examples. Nevertheless, sometimes the way of presenting that information is fairly crowded, as per the entry below (which presents the text that can be viewed on the screen, before scrolling down):

> **love**
>
> [lʌv]
> **NOUN**
> 1. (= *affection*)
> [*of person*] **amor** *m*
> I no longer feel any love for *or* towards him **ya no siento amor** *or* **cariño por él**
> it was love at first sight **fue amor a primera vista ♦ fue un flechazo**
> her love for *or* of her children **su amor** *m* **por sus hijos**
> *her children's love for her* **el amor de sus hijos por ella**
> *don't give me any money, I'm doing it for love* **no me des dinero, lo hago por amor al arte** (*hum*)
> *to marry for love* **casarse por amor**
> *for love of her son, out of love for her son* **por amor a su hijo ♦ por el amor que le tiene/tenía a su hijo**
> *for the love of God or Mike!* **¡por el amor de Dios!**
> to be/fall in love (with sb) **estar enamorado/enamorarse (de algn)**
> *they are in love (with each other)* **están enamorados (el uno del otro)** *(Collins Spanish)*

Information on the usage of the headword *love* is comprehensive — there are usual constructions, and they are quite well exemplified. However, constructions and examples are presented in a long sequence, some in italics and some in roman type (it is not clear to which typeface each kind of information each typeface correspond),[3] and can produce a sense of clutter. For other examples of overload in dictionaries, see Gouws and Tarp (2017: 398ff.).

One of the advantages of the electronic format, is that it allows the elements of the entry to be displayed in such a way that users can access the information contained in it as they need it. Thus, an electronic dictionary can offer a menu of the different meanings of the headword and users can click on the one for which they are looking, and it is displayed with all the information (Lew and Tokarek 2010: 193). Otherwise, after providing one or two basic examples, the entry offers the possibility to view more examples on a box or line marked "more examples" (as in Spanish Oxford online dictionaries, when they were available, or in the electronic Lithuanian–German now under construction, *LiVoLi*).

All these technical possibilities depend on IT specialists and designers, but it is up to lexicographers to decide how to present lexicographic data according to linguistic criteria: how to combine the different entry components (lemma, semantic gloss, cotext item, syntactic construction, examples, etc.) so that the linguistic data are described and presented in a comprehensive yet clear and user-friendly way.

In this study, the focus is on how to present examples considering the systematicity of the linguistic data on the headword being described, how to manage the entry components, especially examples, in order to display information

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

that ranges from more to less systematic or generalisable. Since dictionary users are usually not specialists in linguistics, lexicographers must help them perceive which linguistic facts can be inferred from the lexicographic data provided intuitively, and which facts should be highlighted as more idiosyncratic and therefore explicitly learned.

Here some conclusions from research will be presented that was carried out on a corpus of more than three thousand examples collected from Lithuanian–Spanish/French/Italian dictionaries (cf. Rascón Caballero 2021). One of its aims was to determine what kind of information is conveyed by different types of dictionary example. The practical guidelines drawn from the research, are being implemented in the *Lithuanian–Spanish online dictionary*.

To determine the ways of exemplifying the linguistic data in the bilingual e-dictionary, two coordinates are handled: firstly, types of examples and their relation to the components they illustrate, and secondly, the degree of systematicity of the linguistic phenomena for which the dictionary accounts. From the analysis of examples and the information they convey, a criterion can be drawn, and this criterion will be applied to several samples of dictionary entries to be presented later.

2. Bilingual dictionary entry components and examples

The dictionary example is a fundamental component that helps the user to understand the lemmatised lexical unit and also shows how to use it. It is an instance of the metalinguistic information provided by the other components in the entry and provides additional information (Rascón Caballero 2024: 92-93; *WLWF* 2010: 660 s.v. *Beispielangabe*). It is generally the least condensed component, closest to the real language, but also the most likely to overload the entry.

The example illustrates the information provided by the other entry components (Blanco Escoda 1996: 156; Drysdale 1987: 215; Rascón Caballero 2024: 147):

a. It shows a form of the *headword* and of the equivalent according to the word class: for instance, in the fragment below, sequence *judrios akys — ojos vivaces* ('sparkling eyes'), offers the feminine and plural form of the Lithuanian word *judrus*, a lemma marked with the *morphological indication "adjective"*. In this example, the adjective is acting as a modifier of a noun. In the translation, the plural form of the equivalent *vivaz* is also provided, *vivaces*. This form includes a particular spelling, which increases the informative value of the example:

> **judrus, -i** *adj* vivaz; **judrios akys** *ojos vivaces (Lith–Sp)*

b. It is used in a context that confirms the meaning explicitly expressed by a *semantic gloss* (Hausmann and Werner 1991: 2737; Lettner 2020: 315; Rascón Caballero 2024: 42) conveying the concrete sense of the word concerned;

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

for instance, the phrase *devėti žiedą* ('wear a ring') illustrates the sense indicated by the gloss in brackets in the entry for *žiedas*:

žiedas *(papuošalas* 'ornament'*)* bague; ***devėti žiedą*** porter una bague *(Lith–Fr)*

c. It illustrates the collocational information provided by the *cotext item* (Hausmann and Werner 1991: 2734-2737; Atkins and Rundell 2008: 11, 217, 470; Svensén 2009: 263; Jacinto García 2015: 29, 117; Lettner 2020: 317; Rascón Caballero 2024: 43),[4] which is a word that combines with the lemma sign. Thus, a sentence such as *we don't run a car — no tenemos coche* exemplifies the cotext item in square brackets "[car]" presented before equivalent *tener*:

run [car] tener
we don't run a car no tenemos coche *(Collins Spanish)*

d. It shows an instantiation of the syntactic pattern presented by the *constructional item* (Zgusta 1971: 138-139; Hausmann and Werner 1991: 2732; Atkins and Rundell 2008: 218-219; Svensén 2009: 141; Jacinto García 2015: 21, 117, 118; Lettner 2020: 356; Rascón Caballero 2024: 45). Here the example *siųsk man laišką — envíame una carta* ('send me a letter') illustrates the construction *siusti ką kam*:

siųsti enviar ('to send')
siųsti ką kam enviar algo a alguien ('to send something to someone')
siųsk man laišką envíame una carta *(Lith–Sp)*[5]

Bilingual dictionary examples can be classified in two ways: according to syntactic range and depending on whether the equivalent is included in the translation of the SL example. According to syntactic range or integrity and on the basis of the classification made by Rey-Debove (1971: 302-307) (cf. Cowie 2002: 77-80; Atkins and Rundell 2008: 225; Rascón Caballero 2024: 122-126), examples can be distinguished into:

a. (Condensed) non-verbal phrases:
laukinis kapitalizmas capitalisme sausvage *(Lith–Fr)*; ***labai gerai*** muy bien ('very well') *(Lith–Sp, s.v.* laukinis*)*

b. (Condensed) infinitive phrases:
išbandyti laimę probar suerte ('to take a chance') *(Lith–Sp, s.v.* laimė*)*

c. Sentences o free examples:
leisk vaikams žaisti déjales jugar a los niños ('let the children play') *(Lith–Sp, s.v.* leisti*)*

On the basis of Model's (2009: 131-134) classification, depending on whether or not the translation of the example includes the given equivalent, the example can be (cf. Rascón Caballero 2024: 134-136, 166-179):

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

a. Illustrative: when it includes the given equivalent:

offer
siūlyti(s); pa(si)siūlyti
to offer for sale siūlyti parduoti *(En–Lith)*

b. Supplementary:[6] when it provides another equivalent not given before (here *to direct*)

dirigir
(empresa) to manage, run
(debate) to lead, chair
dirigió la operación de rescate he led *or* directed the rescue operation *(Oxford Spanish)*

or the translation of the example has a different syntactic structure:

castigar to punish
lo castigaron sin postre as a punishment he was made to go without dessert (literally 'they punished him without desert') *(Oxford Spanish)*

3. The degree of systematicity of the language to be learned

Every language fluctuates between systematicity and regularity (learners can infer word combinations and constructions from others previously learned), and restrictions, irregularities, or unpredictable difficulties with which only a competent speaker is able to cope (cf. Atkins and Rundell 2008: 9). The dictionary entry is an attempt to reflect the language as a system at the level of the lexical item being described. Among all entry components, examples, within their formal variety, also fluctuate in an intermediate zone situated between *langue* and *parole* (Olimpio de Oliveira Silva 2009: 96; Jacinto García 2015: 19, 76; Rey-Debove 1971: 302-303).

Lexicographers therefore must arrange the entry in such a way that the components conveying more explicit metalinguistic information (grammatical, collocational, and pragmatic indications) are illustrated by the examples. That examples in turn complement this information, serving as a model for users to use the headword with some guarantee that their concrete linguistic realisations are idiomatic (Jacobsen et al. 1991: 2783). However, since language does not always function according to grammatical and lexical regularities, the examples must also show linguistic phenomena that are unpredictable for the user.

Thus, if learners of Spanish using a dictionary learn the lexical combination *cumplir la ley — to observe the law*, they can deduce that one can also say *cumplir una regla* or *una norma — to observe a rule, a norm*; or that just as one says *cumplir una promesa — to keep a promise* it is also possible to say *cumplir un juramento, la propia palabra — to keep an oath, one's own word*. However, a Lithuanian student of Spanish can hardly foresee that in Spanish one also says *cumplir años — he cumplido veinte años (I have turned twenty years old)* and that this in Lithuanian corresponds to *man sukako 20 metų*, where the syntactic structure is different ('for me twenty years was completed').

The dictionary must therefore account for the more or less systematic lin-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

guistic facts, but it must present them in such a way that users intuitively distinguish what is more predictable and what is more idiomatic, more restrictive and therefore unpredictable. This is where it is important to manage the different entry components in combination with different types of examples, and to arrange the entry so that the information is as explicit as possible, but at the same time user-friendly.

Rascón Caballero (2021: 746ff.) has shown that a first criterion for accounting for the more or less systematicity of the language at the lexical item level is this — the more abstract the linguistic data are, and therefore the more generalisable, the shorter, more standard and more condensed are the entry components that convey it, especially cotext items and phrase examples. On the other hand, more specific and more restricted information has to be shown with larger examples, especially sentences and explicit glosses as usage notes. A second criterion is the word class of the headword — an entry for a noun it is not the same as for an adjective, a verb, etc.

4. Samples of bilingual e-dictionary entries

The mentioned criteria can be illustrated with the following samples of bilingual e-dictionary entries in which different linguistic data are presented with different types of examples. The assumption is that both SL and TL linguistic data provided in the entry are intended for showing how the headword and equivalent are used and how the meaning expressed by the headword in syntagmatic relation with other words is verbalised in TL, i.e. information provided is valid for production in both ST and TL (Adamska-Sałaciak 2006: 495; Yong and Peng 2007: 33; Svensén 2009: 16).

In these samples the concern is the type of information, and the components and exemplification conveying it, not the page design (spacing between lines can be larger, components can be presented in different typefaces and colours, etc.). The following entries are from the planned Spanish–Lithuanian dictionary and the *Lithuanian–Spanish online dictionary* in progress. Examples can be provided in different ways, but this format of exemplifying is proposed according to the abovementioned criteria.

This is how the entry for the mentioned Spanish verb *cumplir* might be presented:[7]

cumplir /kumplír/ conjugación
VERBO TRANSITIVO
1 *(la ley, una regla +)* **laikytis** *(ko)* (įstatymo, taisyklės +)
2 *(una orden, una indicación, un objetivo +)* **(į)vykdyti** (įsakymą, nurodymą, tikslą +)
3 *(una promesa +)* **ištesėti** (pažadą +)
4 *(una misión, una función +)* **atlikti** (misiją, funkciją +)
5 *(un deseo, un sueño +)* **išpildyti, įgyvendinti** *(norą, svajonę +)*
6 *(una edad) asmuo, kuriam sukanka yra veiksnys, kiek laiko - papildinys* **sukakti**
he cumplido veinte años man sukako dvidešimt metų
mi hijo va a cumplir nueve años sūnui sukaks devyni metai

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

VERBO TRANSITIVO / INTRANSITIVO
7 ***cumplir (con) una obligación / un deber*** atlikti pareigą
+ ejemplos
8 ***cumplir (con) los requisitos / las condiciones*** atitikti reikalavimus / sąlygas
+ ejemplos
→ **ajustarse**
VERBO INTRANSITIVO
9 (comportarse según los estándares sociales / elgtis pagal socialines normas)
hemos comprado un regalito a Ana, así que ya hemos cumplido nupirkome dovanėlę Anai, pareigą atlikome *(padarėme tai, ko tikimasi tokiose situacijose)*
en el trabajo Tomás siempre cumple Tomas darbe visada atlieka savo užduotis

The verb *cumplir* has many equivalents in Lithuanian, which depend on the words with which it combines.[8] In bilingual dictionaries, in verb entries, the usual component for discriminating equivalents is the cotext item, along with the indication *transitive, intransitive*, etc., but here the cotext indications are translated after the equivalent, becoming a kind of condensed example. The example, in principle, has to include the headword and the equivalent in the translation, but the translated cotext is a clear solution that avoids repeating both the headword and the equivalent, as they are taken for granted, and provides more general and more systematic information (Rey-Debove 1971: 258; Model 2009: 53; Svensén 2009: 281; Lettner 2020: 171: Rascón Caballero 2021: 152-154; Rascón Caballero 2024: 153-156).[9] In sense 1, when Lithuanian users read *la ley, una regla* ('the rule, a norm') and the translation, they can understand that they are the normal objects of the verb and implicitly form the combinations *cumplir la ley — laikytis įstatymo* ('to observe the law'). It is a hint for users to make other phrases, such as *cumplir una norma*, un *código* ('to observe a norm, to comply with a code'). But the electronic dictionary even allows showing explicitly more collocates of the same lexical set — by clicking on the plus sign more words are displayed in both SL and TL:

cumplir /kumplír/
(...)
1 *(la ley, una regla, una norma, un código)* **laikytis** *(ko)* (įstatymo, taisyklės, normos, kodekso)

The same applies in the meaning 2 to the set *una orden, una indicación, un objetivo* ('to fulfil an order, to comply with an indication, to meet a goal') and their equivalents. Users can predict other words that combine with the headword, which can be completed with *un plan, un proyecto* ('to comply with a plan, to fulfil a project'). By clicking on the plus sign they can also access these collocates:

2 *(una orden, una indicación, un objetivo, un plan, un proyecto)* **(į)vykdyti** (įsakymą, nurodymą, tikslą, planą, projektą)

The same is true for the following senses: Meaning 3 *una promesa* ('to keep a promise') can be completed with *un juramento, la propia palabra* ('an oath, one's

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

word'), and so on.

Sense 6 is *cumplir una edad* ('to reach an age'). As mentioned, this is a rather idiosyncratic use of the word *cumplir* that a learner of Spanish cannot predict: in Lithuanian the person who turns twenty is in the dative case and the years in the nominative case, but in Spanish the person is the subject and the year the object. That information should be explained in more detail with a usage note (here in italics and it can be highlighted in another colour) and sentence examples to draw attention to the syntactic changes:

> 6 *(edad) asmuo, kuriam sukanka yra veiksnys, kiek laiko - papildinys* ('the birthday person is subject, the age is object') **sukakti**
> ***he cumplido veinte años*** ('I have turned twenty years old') man sukako dvidešimt metų ('for me twenty years was completed')
> ***mi hijo va a cumplir nueve años*** sūnui sukaks devyni metai ('my son is going to be nine years old')

In senses 7 and 8, complete examples are provided (*complete* in the sense that they include the headword, that is, they are not translated cotexts):

> 7 ***cumplir (con) una obligación / un deber*** atlikti pareigą
> + ejemplos
> 8 ***cumplir (con) los requisitos / las condiciones*** atitikti reikalavimus / sąlygas
> + ejemplos

Here the information is not as straightforward as in the previous meanings, because the verb can be used transitively or intransitively with the same meaning, so this is the best way to show the optional use of the phrases *cumplir (con) un deber, una obligación* ('to fulfil a duty, an obligation') and *cumplir (con) los requistos, las condiciones* ('to meet requirements, conditions') with or without a preposition. These examples are supplementary, because any equivalent has been provided before for these senses. It would be possible to give the equivalent *atlikti*, but the syntactical optionality and the lexical restriction make it advisable to present the equivalent embedded in the example. Here it is also possible to add optional sentence examples that can be viewed by clicking on + *ejemplos*.

More interesting is the intransitive use of *cumplir*, which means *to do one's duty*, and is even more idiosyncratic to Spanish:

> 9 (comportarse según los estándares sociales / elgtis pagal socialines normas) ('behave according to social standards')
> ***hemos comprado un regalito a Ana, así que ya hemos cumplido*** nupirkome dovanėlę Anai, pareigą atlikome *(padarėme tai, ko tikimasi tokiose situacijose)* ('we have bought a little present for Ana, so we have already done our duty', *we have done what is expected in this situation*)
> ***en el trabajo Tomás siempre cumple*** ('at work, Tomás always does his duty') Tomas darbe visada atlieka savo užduotis ('Tomas always makes his tasks')[10]

In this case, there is no clear Lithuanian equivalent for this meaning and the best way to show this usage is to give an explicit semantic gloss in italics and supplementary and preferably sentence examples. For the first example, a gloss

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

is added. It is up to lexicographers to consider whether or not the provided information is excessive.

This is the entry for the reflexive and intransitive verb *cumplirse*.

cumplirse /kumplírse/ conjugación
VERBO REFLEXIVO INTRANSITIVO
1. *[un deseo, un sueño +]* **išsipildyti** [noras, svajonė +]
mi deseo se cumplió mano noras išsipildė ('my wish came true')
→ **hacerse realidad**
2. *[un plazo]* **baigtis** [terminas]
el plazo de presentación de solicitudes se cumple el 21 de enero paraiškų pateikimo terminas baigiasi sausio 21d. ('the deadline for submitting applications is 21 january')
3. *[un periodo de tiempo]* **sukakti** *[laikotarpis]*
hoy se cumplen 100 años desde que murió don Tomás šiandien sukanka 100 metų, kai mirė ponas Tomas ('today marks 100 years since don Tomás died')

In intransitive verb entries the cotext item used to distinguish equivalents points to the subject of the verb, but it may not be so clear or sufficient to users. In this entry, the translation of the cotext indication is provided, but a short and simple sentence example is the best aid to show how the headword and equivalent are used. In senses 2 and 3, the sentence example is mandatory, as *cumplirse* in these meanings and their equivalents require at least two arguments to show their syntactic and collocational behaviour to users.

This is the entry for the noun *actitud*:

actitud /agtitúd/ plural **actitudes** /agtitúðes/
SUSTANTIVO FEMENINO
(*positiva*, *negativa* teigiamas, neigiamas) **požiūris**
actitud hacia požiūris į ('attitude towards')
cambia / ha cambiado la actitud hacia la educación keičiasi / pasikeitė požiūris į švietimą ('the attitude towards education changes / has changed')
adoptar / mostrar una actitud laikytis požiūrio / parodyti požiūrį ('to adopt/show a responsible attitude')
adoptó / mostró una actitud responsable jis laikėsi atsakingo požiūrio / jis parodė atsakingą požiūrį ('he adopted/showed a responsible attitude')
(ella) mostró una actitud negativa a mi proyecto ji neigiamai vertino mano projektą, mano projekto atžvilgiu ji buvo nusiteikusi neigiamai ('she negatively assessed my project')

In a noun entry, at the beginning, simple cotexts with adjectives or other modifiers, such as *actitud positiva, negativa*, can be inserted, maintaining the word order in both languages: in Lithuanian *teigiamas, neigiamas požiūris*.[11] There is a construction item *actitud hacia* and its translation, and then a sentence example that illustrates this syntactic information and is a natural combination of *actitud*, where this word is the subject. This usage can only be shown with the sentence example format.

The most interesting are the following examples, as they show the quite frequent collocations *adoptar, mostrar una actitud*, which are infinitive phrases. The

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

following is a sentence example where two collocations are together: *adoptó/mostró una actitud responsable — adopted/showed a responsible attitude*.

The last example, *mostró una actitud negativa a mi proyecto*, is very important, because it shows that very often it is not possible to translate word for word. Although the literal translation is possible (*parodė negiamą požiūrį į mano projektą*), it is not well-sounding in Lithuanian (this sentence was checked by Lithuanian natives and translators). One of the valid translations is this one: *ji neigiamai vertino mano projektą* 'she negatively assessed my project'. This is a supplementary example, in which the given equivalent is not included, even the syntactic structure of the translation is different from the Spanish example. Such supplementary examples are very significant and valuable for bilingual dictionaries because they show that each language is a different system in its own right. This kind of example does not solve the practical problem for the users, but when the users come across many supplementary examples, they start to understand that the foreign language must be learned also from the foreign language, that the L1 only gives a first help (Rascón Caballero 2023: 19).

This is the entry for the adjective *lygus* ('level, equal'):

lygus, lygi
(paviršius superficie, ***laukai*** tierra) **plano -a, llano -a** ('level surface, ground')
(plaukai pelo) **liso -a** ('straight hair')
(vienodas) (***kiekis*** cantidad) **igual** ('equal amount')
lygios dalys partes iguales ('equal parts')
lygūs prieš Dievą / įstatymą iguales ante Dios / la ley ('equal before God / the law')
eiti lygiu žingsniu andar al mismo paso ('to keep in step')[12]

The more general combinations of the headword are presented here in a more condensed way, with translated cotexts, but the following ones are more specific: the phrase *lygios dalys* ('equal parts') are usually mentioned in the plural and therefore the example should reflect this formal particularity and be complete (with headword and equivalent); the expressions *lygūs prieš Dievą/įstatymą* ('equal before God/the law') show more specific uses of the headword. The last example (*eiti lygiu žingsniu)* not only is an infinitive phrase which presents a wider context for the adjective *lygus* but is supplementary — it provides a specific equivalent not given before for this particular use of the headword (*mismo* 'same').

5. Conclusion

Nowadays, there is no lack of information when it comes to learning a foreign language. The problem is rather that there is an excess of information. Bilingual dictionaries are useful tools, among others, for accessing the linguistic data to be learned. Modern technology facilitates lexicographic work, but information has to be selected, processed and provided to learners by lexicographers in a comprehensive, but also clear and user-friendly way. In this article, the focus has been placed on how to present the dictionary entry components, and especially the examples, from the point of view of the systematicity of the language paired

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

with another language: how to describe the linguistic data so that users could infer what is generalisable, predictable and what is more specific and unpredictable. Research on bilingual dictionary examples has shown that short components (especially cotext items) and examples (non-verbal and infinitive phrases) provide more generalisable, systematic information. An unusual example format was proposed — the translated cotext that does not include the headword and the equivalent, as they are assumed. This type of example provides a clue for users to form phrases according to this model, although an e-dictionary can explicitly display more words combining with the headword and the equivalent. When the information is more difficult to generalise (e.g. certain flexional forms, figurative senses, how to use an intransitive verb, or to show a noun as the subject of a verb), more explicit examples should be given. Sentence examples should be provided to illustrate complex syntactic structures (especially when it is necessary to show the subject and when the predicate requires more than one argument). Supplementary examples (especially, when the syntactic structure of the translation is different from the one of the SL example) form a very valuable contribution to bilingual dictionaries, as they draw the users' attention to the real fact that the paired languages are not parallel systems, that the L1 is only a first aid to learn a foreign language.

Endnotes

1. Toffler acknowledges that the expression *information overload* had already been used in the context of mental health by Miller in 1963 (1970: 350 footnote).
2. The Oxford bidirectional dictionary for the pair Spanish/English, as comprehensive as Collins dictionary, was available online for free until 24 August 2022.
3. In this online dictionary, constructions are usually presented in normal typeface and examples in italics, but in this entry and others this distinction is not kept. The author of this paper made some suggestions to the editors regarding the unclear distinction between constructions and examples in this entry and received a very grateful response to these remarks.
4. A variant of this component is proposed below: the translated cotext.
5. https://www.webonary.org/lietuviukalba-espanol/g73277260-07e2-4d10-9a4c-fc8b063e1b5a/?lang=en
6. This reduces to *supplementary* the two types distinguished by Model — *supplementary* and *direct*, the former being the example whose translation includes an equivalent not given before in the entry and the latter being the example with its translation given without having previously provided any equivalent for the headword or for one of its senses being illustrated.
7. It should be noted that an e-dictionary allows the different components of an entry to be represented by different colours, as is the case in the *Lithuanian–Spanish online dictionary*. Examples are translated into English in the comments of each sense.
8. On the concept of collocation in bilingual lexicography, and how collocational information is conveyed in bilingual dictionaries, see Rascón Caballero (2022; 2024: 219-232)
9. This kind of example is used in two Piesarskas' dictionaries (1991 [1981], 1984) and the Lithuanian–Spanish dictionaries.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

10. These two examples were translated by translator and interpreter Nomeda Lukoševičienė.
11. Note that in these examples or translated cotexts, the SL word combined with the headword and the TS translation combined with the equivalent are together, not separately as in a verb entry.
12. Abridged from https://www.webonary.org/lietuviukalba-espanol/ga0b77b97-2aaa-4741-a9fb-65684d23d54f/?lang=en

References

Dictionaries

Collins Spanish. https://www.collinsdictionary.com/dictionary/english-spanish.

En–Lith: Anglonas. Anglų–lietuvių kalbų žodynas. Electronic version. Piesarskas, B. UAB Fotonija. 2006.

Learner's English–Lithuanian. Piesarskas, B. Kaunas: Šviesa. 1991 [1981].

Learner's Lithuanian–English. Piesarskas, B. Kaunas: Šviesa. 1984.

Lith–Fr (Lithuanian–French Dictionary): Dictionnaire lituanien–français. Lietuvių–prancūzų kalbų žodynas. Melnikienė, D. Paris: Hermann. 2020.

Lith–Sp (Lithuanian–Spanish Dictionary): Lietuvių–ispanų kalbų žodynas. Diccionario lituano–español. Rascon, A. Vilnius: Žodynas. 2002.

Lithuanian–Spanish Online Dictionary. Rascón, A. and N. Lukoševičienė. Since 2021. https://www.webonary.org/lietuviukalba-espanol/?lang=en

LiVoLi: Lietuvių–Vokiečių–Lietuvių kalbu žodynas. Plaušinaitytė, L., V. Zubaitienė and J. Gelumbeckaitė. Since 2020. http://www.livoli.flf.vu.lt/

Oxford Spanish Dictionary. Oxford/New York/Madrid. 1994, 2003.

WLWF: Wörterbuch zur Lexikographie und Wörterbuchforschung. Vol. 1, *A–C.* 2010. Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer and W. Wolski (Eds.). 2010. Berlin/ New York: De Gruyter.

Other literature

Adamska-Sałaciak, A. 2006. *Meaning and the Bilingual Dictionary: The Case of English and Polish.* Frankfurt am Main: Peter Lang.

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Béjoint, H. 2004. *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.

Blanco Escoda, X. 1996. L'exemple dans la lexicographie bilingue. Traitements métalinguistiques. *Le Français Moderne* LXIV(2): 156-168.

Bogaards, P. 2003. Uses and Users of Dictionaries. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography.* Terminology and Lexicography Research and Practice. Vol. 6: 26-33. Amsterdam/ Philadelphia: John Benjamins.

Corpas Pastor, G., J. Leiva Rojo and M.J. Varela Salinas. 2001. El papel del diccionario en la formación de traductores e intérpretes: análisis de necesidades y encuestas de uso. Ayala Castro, M.C. (Ed.). 2001. *Diccionarios y enseñanza*: 239-273. Alcalá de Henares: Universidad de Alcalá.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

Cowie, A.P. 2002. Examples and Collocations in the French 'Dictionnaire de langue'. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins:* 73-90. Göteborg: EURALEX.

Drucker, P. 1968. *The Age of Discontinuity: Guidelines to our Changing Society.* New York: Harper and Row.

Drucker, P. 1989. *The New Realities*. London: Routledge.

Drucker, P. 1993. *Post-Capitalist Society*. New York: HarperCollins.

Drysdale, P.D. 1987. The Role of Examples in a Learner's Dictionary. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*. Lexicographica. Series Maior 17: 213-223. Tubingen: Max Niemeyer.

Dubois, J. and C. Dubois. 1971. *Introduction à la lexicographie: le dictionnaire.* Paris: Larousse.

Dziemianko, A. 2018. Electronic Dictionaries. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 663-683. London/New York: Routledge.

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.

Fuertes-Olivera, P.A. and S. Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries*. Berlin/ Boston: De Gruyter.

Gouws, R.H. and S. Tarp. 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

Hausmann, F.J. and R.O. Werner. 1991. Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: eine Übersicht. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta. (Eds.). 1991. *Wörterbücher. Dictionaries. Dictionnaires.* Vol III: 2729-2769. Berlin/New York: De Gruyter.

Jacinto García, E.J. 2015. *Forma y función del diccionario. Hacia una teoría general del ejemplo lexicográfico.* Jaén: Universidad de Jaén.

Jacobsen, J.R., J. Manley and V.H. Pedersen. 1991. Examples in the Bilingual Dictionary. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta. (Eds.). 1991. *Wörterbücher. Dictionaries. Dictionnaires.* Vol III: 2782-2789. Berlin/New York: De Gruyter.

Lettner, K. 2020. *Zur Theorie des lexikographischen Beispiels: Die Beispielangaben in der ein- und zweisprachigen pädagogischen Lexikographie des Deutschen.* Berlin/Boston: De Gruyter.

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.

Model, B. 2009. El ejemplo en los diccionarios bilingües. Fuentes Morán, M.T. and B.A. Model (Eds.). 2009. *Investigaciones sobre lexicografía bilingüe*: 117-157. Granada: Tragacanto.

Nomdedeu-Rull, A. and S. Tarp. 2018. Hacia un modelo de diccionario en línea para aprendices de español como LE/L2. *Journal of Spanish Language Teaching* 5(1): 50-65.

Olimpio de Oliveira Silva, M.E. 2009. Discurso metalexicográfico sobre los diccionarios bilingües, la equivalencia y la equivalencia fraseológica. Fuentes Morán, M.T. and B.A. Model (Eds.). 2009. *Investigaciones sobre lexicografía bilingüe*: 81-116. Granada: Ediciones Tragacanto.

Rascón Caballero, A. 2021. *Teoría y práctica del ejemplo lexicográfico bilingüe. Análisis de los ejemplos diccionarios lituano–español/italiano/francés.* Unpublished PhD Dissertation. Barcelona: Universitat Autònoma de Barcelona.

Rascón Caballero, A. 2022. La indeterminación del concepto de colocación en la lexicografía bilingüe. *ELUA* 37: 93-116.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2028 (Article)

Rascón Caballero, A. 2023. Equivalente y ejemplo en los diccionarios bilingües: las traducciones de los ejemplos que no incluyen el equivalente. *ReDIILLet Revista digital internacional de lexicología, lexicografía y terminologia* 6: 1-26.

Rascón Caballero, A. 2024. *The Theory and Practice of Examples in Bilingual Dictionaries.* Berlin/Boston: De Gruyter.

Rey-Debove, J. 1971. *Étude linguistique et sémiotique des dictionnaires français contemporains.* The Hague: Mouton.

Svensén, B. 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Tarp, S. 2011. Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. *Lexikos* 21: 217-231.

Toffler, A. 1970. *Future Shock*. New York: Random House.

Torres del Rey, J. 2009. Diccionarios electrónicos bilingües: nuevas posibilidades de futuro. Fuentes Morán, M.T. and B.A. Model (Eds.). 2009. *Investigaciones sobre lexicografía bilingüe*: 29-79. Granada: Ediciones Tragacanto.

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: Mouton.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

Yongwei Gao. *A Dictionary of Contemporary English Neologisms and Trendy Words*. 2024, 406 pp. ISBN: 978-7-5213-5810-0 (Paperback). Beijing: Foreign Language Teaching and Research Press. Price $9.32.

1. Introduction

Dynamic evolution is one of the most captivating characteristics of a language, with the emergence and obsolescence of vocabulary serving as the most evident markers of this transformation. "New words enter the language every day, words acquire or lose meanings, and words cease to be used" (Minkova and Stockwell 2009: 5). In English, hundreds or even thousands of new words are coined each year, each reflecting distinct aspects of contemporary life and society. Published in October 2024, *A Dictionary of Contemporary English Neologisms and Trendy Words* (henceforth DCENTW) stands as a testament to this linguistic dynamism. The dictionary was compiled by Professor Gao Yongwei of Fudan University, a prominent figure in the field of English–Chinese lexicography. Professor Gao's scholarly contributions to the study and documentation of English neologisms are extensive. Prior to this publication, he played a key role in compiling five influential dictionaries dedicated to new English vocabulary: *A Supplement to the English–Chinese Dictionary* (《英汉大词典补编》, 1999), *A Dictionary of Contemporary English New Neologisms* (henceforth DCENN, 《当代英语新词语词典》, 2002), *A 21st Century English–Chinese Dictionary of Neologisms* (《21世纪英语新词语词典》, 2021), *A New Age Dictionary of English Neologisms* (《新时代英语新词语词典》, 2023b), and *An English–Chinese Dictionary of Neologisms in Present-day English* (《当代英语精选新词语词典》, 2023a). These works collectively showcase Professor Gao's commitment to capturing and analyzing the ever-evolving nature of the English lexicon, providing invaluable resources for linguistic scholars, translators, and language learners alike.

DCENTW includes approximately 17,000 newly emerged or frequently used terms in contemporary English, with the earliest words dating back to the 1970s. The dictionary includes nearly 200 idiomatic expressions and over 3,000 technical terms spanning a wide range of fields, including politics, economics, military affairs, sports, biology, and medicine. The selection of entries is the result of years of extensive research and compilation by Professor Gao, complemented by reference to usage frequency in news corpora. Each entry is carefully documented, featuring abundant examples, and a dedicated section on etymology and word-formation. This design not only illustrates the structural formation and historical context of the terms but also facilitates readers' comprehension and retention of the vocabulary.

Language manifests the changes in society, effectively documenting the complexities of social transformation and cultural shifts. In particular, "the new terminology of technological and scientific advance had an immediate impact on the language, adding tens of thousands of words to the English lexicon" (Crystal 2003: 78). DCENTW provides a multifaceted reflection of the dynamic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

evolution of global language and culture over the past few decades. Serving as both a practical linguistic resource and a comprehensive academic reference, the dictionary offers invaluable insights for language learners and researchers alike, enabling them to explore the development of present-day English and its cultural significance in greater depth. This review seeks to critically evaluate the distinctive features of the dictionary's compilation, including its strengths and innovative aspects, while also addressing its limitations and areas for improvement.

2. Compilation features

DCENTW includes most new entries found in the aforementioned dictionaries (excluding DCENN), as well as thousands of newly coined terms that have yet to be included in major English–Chinese dictionaries, such as *The English–Chinese Dictionary* (Unabridged, second edition) by Lu and *New Century English–Chinese Dictionary* by Hu. As Gao (2024: III) mentions in DCENTW's preface, "From a compilation standpoint, DCENTW also marks a notable breakthrough by combining the objective recording of new words found in the aforementioned new word dictionaries with an editorial approach that incorporates illustrative examples to aid selection". This paper primarily analyzes the compilation features of DCENTW through its word selection and entry translations.

2.1 Coverage of entries in DCENTW

Firstly, the new words included in DCENTW cover a wide range of fields, including politics, economics, science, and computing. It includes nearly 17,000 new and frequently used terms that have emerged in English since the 1970s, including AI (人工智能), big data (大数据), COVID (新型冠状病毒感染), goblin mode (摆烂模式), MOOC (慕课), permacrisis (长久危机), taikonaut (中国航天员), and vlog (视频博客). "Fundamental properties of language design indicate that a rich tradition is correct in regarding language as essentially an instrument of thought" (Chomsky 2015: 16). And significant social changes over time often lead to corresponding shifts in the forms of linguistic expression, particularly in vocabulary usage. This expansive inclusion of new words in DCENTW illustrates the dictionary's comprehensive approach to documenting contemporary language. By incorporating neologisms from diverse areas such as technology, global crises, education, and cultural trends, DCENTW offers a broad snapshot of how language evolves in response to societal changes. For instance, terms like "permacrisis" depicts the global political and environmental instability, while "MOOC" mirrors the growing influence of digital education. The inclusion of specialized terms like "taikonaut" also indicates the dictionary's ability to capture emerging language in fields like space exploration. Overall, DCENTW serves as

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

a valuable resource for understanding the intersection of language and the rapidly changing world.

Secondly, DCENTW includes over 3,000 terms from various fields, categorized under 66 subject labels, covering areas such as computer science, medicine, sociology, law, architecture, and oceanography. Specific examples can be found in Table 1. By assigning specific labels to each entry, such as "computer science", "medical science", or "astronomy", the dictionary not only enhances the organization and accessibility of specialized vocabulary but also indicates the growing interdisciplinary nature of contemporary language. For example, terms like "abandonware" (弃件) and "cloud mining" (云采矿) are clearly marked under computer science, allowing users to quickly identify technical jargon related to this field. Similarly, terms such as "keratomileusis" (角膜磨镶术) and "long COVID" (新冠长期症状) are grouped under medical science, helping users distinguish between terms related to healthcare and those from other domains. The inclusion of subject labels thus facilitates the efficient retrieval of relevant information, especially for those seeking expertise in a particular discipline. Moreover, this practice underscores the specialized nature of modern terminologies, ensuring that users understand the context and precise meaning of the terms within their respective fields. For instance, terms from fields like military science, such as "asymmetric warfare" (不对称战争) and "bunker buster" (掩体炸弹), are categorized separately, ensuring that users interested in defense studies can easily locate and comprehend these terms. In addition, the use of subject labels enhances the dictionary's capacity to reflect the rapid development and expansion of knowledge in various areas, from the emerging field of "cryptolect" (隐蔽语言) in linguistics to the complex concepts of "astroparticle" (天体粒子) in astronomy.

Table 1: The entries recorded in DCENTW

Subject label	Entry
计算机 (computer science)	abandonware (弃件), abend (异常终止), cloud mining (云采矿, 云挖矿), IVR (interactive voice response, 交互式语音应答)
医学 (medical science)	keratomileusis (角膜磨镶术), liposuction (脂肪抽吸), locked-in syndrome (闭锁综合征), long COVID (新冠长期症状)
化学 (chemistry)	curcumin (姜黄素), dicamba (麦草畏, 除草剂), fluoroelastomer (氟橡胶), gel permeation chromatography (凝胶渗透色谱法)
军事 (military)	asymmetric warfare (不对称战争), bunker buster (掩体炸弹), kill vehicle (击杀载具), soft bomb (石墨炸弹)
心理学 (psychology)	acrotomophilia (慕残), alien hand syndrome (异手症，异己手综合征), dark triad (黑暗三联征), enclothed cognition (具衣认知)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

语言学 (linguistics)	cryptolect (隐蔽语言), semantic bleaching (语义漂白), translanguaging (超语), uptalk (升调结尾)
天体学 (uranology)	astroparticle (天体粒子), baby universe (婴儿宇宙), brown dwarf (棕矮星), cannibal galaxy (食人者星系)

Thirdly, the inclusion of nearly 200 idioms in DCENTW, labeled with "idiom", underscores the dictionary's comprehensive approach to documenting not only technical and formal vocabulary but also the rich, figurative expressions that shape everyday language. By incorporating idioms like "Netflix and chill (边看电影边亲热)", "open a can of whoop ass (给予致命一击;使遭受惨败)", and "take a bullet (替某人挡子弹, 挨枪子)", the dictionary accentuates the playful, culturally specific ways language evolves in popular discourse. These expressions are deeply embedded in contemporary social and cultural contexts, showcasing attitudes toward relationships, gender, and humor, as seen in idioms such as "play the gender card (打性别牌)". Their inclusion broadens the dictionary's scope, making it a valuable resource for understanding both the literal and figurative meanings of modern English. Idiomatic expressions are particularly useful in conveying the nuances of natural conversation. Native speakers often use idiomatic expressions naturally, without consciously considering their figurative meanings. Non-native speakers often struggle to understand idioms, as "they (idiomatic and fixed expressions) form collocations with other items in the text as single units and enter into lexical sets which are different from those of their individual words" (Baker 2018: 73). By offering clear translations and contextual examples, DCENTW facilitates a deeper understanding of these expressions, helping learners navigate the complexities of everyday language. Moreover, the inclusion of idioms illustrates the ever-evolving characteristic of language, as these phrases often arise in response to shifting cultural, social, or technological trends, exposing the ongoing interaction between language and society.

Fourthly, the inclusion of combining forms in DCENTW reflects the dictionary's proactive approach to documenting emerging word formation patterns, thereby providing valuable insights into contemporary morphological innovation. By including highly productive morphemes in Table 2 like "-flation" "-fluencer" "-geddon" and "-preneur", the dictionary embodies the inventive and malleable qualities of present-day English, where new terms are frequently coined by attaching these combining forms to existing words. These combining forms are particularly significant because they represent linguistic trends that are not only prevalent in popular discourse but also play a pivotal role in shaping the evolving lexicon. For example, terms such as "Biden-flation" "mom-fluencer" and "foodpreneur" display contemporary social, political, and technological phenomena, and demonstrate the ability of language to adapt rapidly to changing cultural contexts. Through the documentation of these forms, DCENTW provides a comprehensive resource for understanding the processes by which new words are formed, emphasizing the role of productivity and abbreviation in the expansion of the lexicon. This inclusion not only enhances the dictionary's cov-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

erage of current language use but also ensures that it remains relevant to users seeking to understand both the meaning and formation of newly emerging expressions. Moreover, the entry of such combining forms into the dictionary marks a shift toward recognizing the importance of morphological creativity in language development, a trend that portrays broader shifts in how language interacts with technological advancements, media culture, and societal changes.

Table 2: The entries recorded in DCENTW

Combining form	Entry
anarcha-	表示 "无政府主义"（亦作anarcho-）：anarcha-feminists无政府女性主义者
cyber-	表示 "网络的"：cyber-operations网络行动 / cyber-surveillance网络监控
-flation	表示 "通货膨胀"：Biden-flation拜登通胀 ▲inflation的缩略
-fluencer	表示 "网红"：mom-fluencer网红妈妈 ▲influencer的缩略
-geddon	表示 "灾难事件"（亦作-mageddon）：eco-geddon 生态灾难 ▲Armageddon的缩略
-pocalypse	表示 "灾难性事件"：bee-pocalypse 蜂灾 / snow-pocalypse 大雪灾 ▲apocalypse的缩略
-preneur	表示 "企业家"：foodpreneur食品企业家 ▲entrepreneur的缩略
-tacular	表示 "非常棒的" "精彩的"：pec-tacular胸肌发达的 ▲spectacular的缩略

2.2 Translation in CEDU in DCENTW

Translation in DCENTW focuses on accurately conveying headwords and example sentences. Headwords are expressed to display their precise subject-specific meanings, while example sentences illustrate their practical use in context.

The translation of headwords in DCENTW strikes a careful balance between linguistic accuracy and contextual adaptability, which shows the diversity and specificity of the terms included. The dictionary consistently provides precise translations that align with the semantic nuances of each headword, ensuring clarity for users across various domains. For instance, technical terms such as "atrial natriuretic factor (心房肽)" and "atomic force microscope (原子力显微镜)"[1] are rendered with their exact scientific equivalents in Chinese, maintaining their terminological integrity. Conversely, colloquial expressions like "atomic wedgie (扯内裤套头)", "Covidiot (新冠憨憨)" and "bucketload (许多，大量)" are translated in a manner that reveals their idiomatic flavor and cultural connotations, bridging linguistic and cultural gaps effectively. Furthermore, the dictionary addresses contemporary linguistic phenomena by incorporating transliterations or hybrid translations for novel constructs like "bug-out bag (应急包)", "centennial (Z世代)", "chia (奇雅，芡欧鼠尾草) and "at sign (@符号)", thereby exhibiting ongoing linguistic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

evolution while preserving the original meaning. This comprehensive and adaptable approach underscores DCENTW's commitment to fostering an inclusive and practical lexicon.

DCENTW includes over 1,300 example sentences, which plays a crucial role in enhancing the understanding of newly introduced terms by providing contextually rich and accurate translations. By offering clear, practical examples for each entry, the dictionary ensures that readers can grasp both the meaning and application of these terms in everyday language. This approach not only bridges linguistic gaps but also reflects the dynamic process of language as it adapts to new phenomena, such as the emergence of "Covidiot" or "Generation Z". The provision of these sentence examples allows for a deeper comprehension of how specific terms, often rooted in cultural or technological shifts, function in real-world discourse. Additionally, the use of such contextualized translations as shown in Table 3 aids non-native speakers in navigating the complexities of modern lexicon, making it more accessible and relevant to diverse audiences. The integration of sentence exemplifications, therefore, extends beyond mere vocabulary listing, offering a more practical understanding of contemporary English usage in both domestic and global contexts.

Table 3: The example sentences recorded in DCENTW

Example Sentence	Translation
Please don't be fooled by this spring-like London weather as it is 100% going to *chuck it down* tomorrow.	请不要被像春季般的伦敦天气所愚弄，因为明天百分之一百会下大雨。
As the *climate crisis* has worsened, experts have put forth numerous solutions to curb greenhouse gas emissions.	随着气候危机的恶化，专家们提出了许多控制温室气体排放的解决方案。
Gen Z are consumers born between the mid-1990s and the early 2010s; they number 260 million digital natives in China.	Z世代是1990年代中期到本世纪10年代初出生的消费者；中国的这批数字原住民数量达2.6亿。
New technology will arrive sooner, domestic firms will outcompete foreigners, and steady employment will ensure *a chicken in every pot*.	新技术将更快出现，国内公司在竞争中会胜出外国公司，稳定的就业将确保家家都有肉吃。

3. Room for improvement in DCENTW

Despite the editor's tireless efforts, this lexicographic work still leaves room for further development, particularly in terms of entry inclusion, translation accuracy, and example selection.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

3.1 Problems with the entry inclusion

In DCENTW, certain newly coined terms have been inadvertently omitted, such as aforenoted, African massage, Bafumbira, Calennig, content creator, dagginess, elf-shot, grief tourism, GPT, Generation Alpha and hash collision. Furthermore, several established English words with newly emerging meanings have not been included. For example, argument as a noun in the domain of music refers to "a unifying theme, idea, or melody within a musical work, particularly a symphony", or "the cohesiveness of a piece of music"; blow job in the military context denotes "a jet aeroplane"; and scrimmage is used to describe "a disorderly crowd of people or things". A dictionary will inevitably overlook new words or new meanings to some extent because language is constantly evolving, and dictionaries can only take a snapshot of the language at a given time. The process of word inclusion involves criteria such as frequency of use, established meaning, and broad acceptance, which means that newly coined terms or emerging meanings may not meet these thresholds immediately. Additionally, the lag between the coinage of a term and its official recognition or widespread usage further contributes to this gap.

3.2 Problems with the translation

In DCENTW, some of the translations of entries are inaccurate or could benefit from further clarification. For instance, adultescent does not simply refer to middle-aged individuals who dress young, but rather to middle-aged people whose clothing, interests, and activities generally align with the lifestyle habits and cultural norms of younger generations. As such, the content in parentheses should be revised to "衣着、兴趣和活动通常与年轻人的生活习俗一致的中年人", providing a clearer description. Similarly, advergaming refers specifically to "the use of computer games to advertise products", and so the appropriate Chinese translation should be "广告嵌入式游戏", rather than "广告游戏的制作", which could misleadingly imply the process of creating such games, rather than their use as a marketing tool. Additionally, some supplementary information provided within parentheses is redundant. For example, the translation of A&E as "(医院的）急诊部，急诊室" is sufficient, as both of these Chinese terms already imply departments within hospitals that are responsible for providing emergency medical care. Therefore, the explanatory content in parentheses should be omitted to avoid unnecessary repetition. On the other hand, some definitions are overly simplified, which may lead to misunderstandings for readers unfamiliar with the term. For example, dark triad refers to three distinct yet interrelated personality traits, all of which are considered antisocial personality traits within Western psychological frameworks. These traits are collectively known as the "black triad of personality". The traits that constitute the dark triad include Machiavellianism, narcissism, and psychopathy. To provide a more comprehensive explanation, the content in parentheses should be expanded to in-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

clude: "三联征是指马基雅维利主义、自恋和精神病态", ensuring that readers gain a fuller understanding of the term's meaning and its significance in psychological contexts.

3.3 Problems with example selection

There are over 1,300 example sentences in DCENTW, but the selection of which headwords are provided with example sentences seems somewhat inconsistent and random. For instance, example sentences are given for terms like "boy band" and "global warming", with sentences such as: "It has been nearly three decades since Backstreet Boys made its debut, but the legacy of the American 90s boy band strongly lives on" and "The world's biggest ice sheet is more vulnerable to global warming than scientists previously thought". However, both "boy band" and "global warming" are terms whose meanings are relatively intuitive and easily understood by most readers, making the inclusion of example sentences unnecessary in these cases. These words are commonly known, and their definitions are self-explanatory in context, meaning additional examples do not add much value to the reader's understanding. Conversely, headwords like "gouch out", which may be less familiar or whose meaning could be more ambiguous, would greatly benefit from the inclusion of example sentences. For terms like these, providing context through examples is essential to help readers understand their correct usage in various scenarios. Including examples for such words would not only clarify their meaning but also guide readers on how to incorporate them appropriately into their own speech or writing. More selective and thoughtful inclusion of examples, especially for less commonly understood or more complex terms, would significantly enhance the dictionary's practical value and make it a more effective resource for learners.

4. Conclusion

Compiling a dictionary of new words presents inherent challenges as it involves not only capturing the rapid evolution of language but also distinguishing between transient fads and lasting linguistic trends. The inclusion of a new word requires careful consideration of its widespread use, cultural relevance, and potential to endure over time. Moreover, the process of defining and contextualizing these words demands expertise in various fields, including linguistics, culture, and technology. Balancing the need for accuracy with the desire to remain current can be difficult, especially when words emerge from specialized or niche areas that may not yet have broad recognition. DCENTW exemplifies how technology has transformed the efficiency of modern dictionary compilation. Beginning on June 9, 2022, Prof. Gao completed crucial tasks — new word selection, headword translation, example sentence selection and translation, and etymological information gathering — in just six months. This rapid progress was

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

facilitated by using advanced dictionary compilation software, online technologies, and corpus-based methods. These tools enabled him to process large volumes of language data quickly, ensuring the dictionary remains up to date with current linguistic trends. By incorporating these technologies, the project not only achieved a high level of precision but also a remarkably fast completion time, showcasing how modern tools can streamline traditionally time-consuming tasks in lexicography.

Endnote

1. The Chinese translations of "atrial natriuretic factor" and "atomic force microscope" are consistent with those provided by Term Online: https://www.termonline.cn/search?searchText=atrial+natriuretic+factor and https://www.termonline.cn/search?searchText=atomic+force+microscope. Term Online has become a comprehensive knowledge service platform, offering the most extensive global Chinese terminology resources, the highest data quality, and the most systematic functionality.

References

A. Dictionaries

Gao, Y.W. 2021. *A 21st Century English–Chinese Dictionary of Neologisms.* Shanghai: Fudan University Press.

Gao, Y.W. 2023a. *An English–Chinese Dictionary of Neologisms in Present-day English.* Shanghai: Fudan University Audio-Visual Publishing House.

Gao, Y.W. 2023b. *A New Age Dictionary of English Neologisms.* Beijing: The Commercial Press.

Gao, Y.W. 2024. *A Dictionary of Contemporary English Neologisms and Trendy Words.* Beijing: Foreign Language Teaching and Research Press (DCENTW).

Hu, Z.L. 2016. *New Century English–Chinese Dictionary*. Second edition. Beijing: Foreign Language Teaching and Research Press (NCECD).

Lu, G.S. 2007. *The English–Chinese Dictionary.* Unabridged, second edition. Shanghai: Shanghai Translation Publishing House.

Lu, G.S. and Y.W. Gao. 1999. *A Supplement to the English–Chinese Dictionary.* Shanghai: Shanghai Translation Publishing House.

Lu, G.S. and Y.W. Gao. 2002. *A Dictionary of Contemporary English New Neologisms.* Beijing: Foreign Language Teaching and Research Press (DCENN).

B. Other literature

Baker, M. 2018. *In Other Words: A Coursebook on Translation*. Third edition. London: Routledge.

Chomsky, N. 2015. *What Kind of Creatures Are We?* New York, NY: Columbia University Press.

Crystal, D. 2003. *English as a Global Language*. Second edition. Cambridge: Cambridge University Press.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2029

Minkova, D. and R. Stockwell. 2009. *English Words: History and Structure*. Second edition. Cambridge: Cambridge University Press.

Aling Shi
School of Foreign Studies
Tongji University
Shanghai
China
(shialing2011@163.com)
(https://orcid.org/0009-0001-6053-0930)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

La Lexicographie Scolaire comme Fondement de la Lexicographie Moderne au Gabon

Edgard Maillard Ella, *Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques (GREDYLEX), Institut de Recherche en Sciences Humaines, CENAREST Libreville, Gabon (ellamaillard@yahoo.fr) (https://orcid.org/0009-0000-6716-3371)*

Résumé: Il y a un lien étroit entre le processus de standardisation des langues et les ressources lexicographiques dont on dispose des données fiables. La lexicographie scolaire joue ainsi à la base un rôle capital dans la standardisation du français. En référence au modèle français, nous montrons la nécessité de prioriser la lexicographie scolaire comme fondement de la lexicographie moderne des langues locales au Gabon. Le processus de la standardisation des langues locales en sera le corollaire. Les dictionnaires scolaires bilingues français et ces langues locales, constituent le type d'ouvrage à confectionner dans cet ordre prioritaire. Ce travail renvoie exclusivement à celui de Boulanger (2006) qui a fait la démonstration du rôle essentiel de la lexicographie scolaire dans le processus de standardisation de la langue française.

Mots clés: DICTIONNAIRES SCOLAIRES BILINGUES, LEXICOGRAPHIE ANCIENNE, LEXICOGRAPHIE MODERNE, LEXICOGRAPHIE SCOLAIRE, STANDARDISATION

Abstract: School Lexicography as the Foundation of Modern Lexicography in Gabon. There is a close connection between lexicographic resources and the standardization process of languages for which reliable data is available. School lexicography thus plays an essential role in the standardization of French. With reference to the French model, we show the need to prioritize school lexicography as a foundation of modern lexicography of the local languages in Gabon. The process of standardization of local languages will be the corollary. Bilingual school dictionaries of French and these local languages constitute the type of work to be compiled in this priority order. This article refers exclusively to Boulanger (2006) who demonstrated the essential role of school lexicography in the standardization process of French.

Keywords: BILINGUAL SCHOOL DICTIONARIES, FORMER LEXICOGRAPHY, MODERN LEXICOGRAPHY, SCHOOL LEXICOGRAPHY, STANDARDIZATION

Introduction

Le Gabon est un pays d'Afrique centrale, ce qui le situe dans l'espace où sont parlées les langues dites bantu ou bantoues c'est selon, qui ont la particularité

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

d'être des langues à tons. Á l'instar de nombreux pays africains, l'ordre social du Gabon découle de la colonisation avec corolaire l'existence d'une langue européenne comme langue officielle. Autour de cette dernière, gravitent des langues locales ou nationales officialisées ou non[1]. Dans le dernier cas, ces langues assurent des fonctions dans la communication officielle. Le Gabon, avec les langues locales telles que le *fang*, l'*ikota*, l'*inzébi*, le *(yi)punu*, le *lémbaama*, le *myènè* et le *téké* pour ne citer que celles-là, figure dans le deuxième cas avec le français comme seule langue officielle.

Cependant, ces pays, le Gabon y compris, aspirent naturellement à standardiser leurs langues locales afin de les officialiser et de leur conférer à terme, le même statut que les langues européennes qui sont parfois les seules officielles sur leur sol. Dans la politique linguistique du Gabon, selon l'article 2 de la constitution, le français est reconnu comme langue officielle et l'État œuvre à la promotion des langues locales. Dans le système éducatif, les états-majors de l'éducation nationale consignent souvent les langues locales comme activités d'immersion. Sous la houlette de la Fondation André Raponda Walker et sur la base du *Rapport Final de la Session de Concertation sur l'Orthographe des langues Gabonaises* en 1999, ces langues étaient inscrites dans le programme d'enseignement et ont été incluses dans l'examen du BEPC.

Il y a un lien étroit entre le processus de standardisation des langues et les ressources lexicographiques dont on dispose des données fiables. Boulanger (2006) a montré le rôle capital joué à la base par la lexicographie scolaire dans la standardisation du français. Étant un état dans le cheminement de l'apprentissage de la langue qui va du bas vers le haut, le dictionnaire scolaire vulgarise et aide à se familiariser avec le bon usage de la langue dès le plus bas âge.

Nous allons d'abord, d'une part, discuter du critère lexicographique permettant de délimiter dans le temps, une lexicographie que nous pouvons qualifier au Gabon d'ancienne et l'autre, de moderne. D'autre part, nous allons essayer de dresser un bilan de la lexicographie scolaire. Ensuite, nous allons démontrer comment en priorisant les dictionnaires scolaires bilingues français et langues locales, la lexicographie scolaire peut être fondamentale dans le processus de standardisation de ces langues locales sous le même angle que celui décrit dans le cas du français. Cette lexicographie se positionnera alors comme le fondement de la lexicographie moderne au Gabon. Ce travail, renvoie exclusivement à celui de Boulanger (2006).

1. La lexicographie ancienne et moderne, aspects de la lexicographie scolaire au Gabon

Dans les lignes qui suivent, il sera question d'une part, de discuter des aspects des critères qui peuvent permettre de délimiter du point de vue scientifique ainsi que dans le temps, une lexicographie que nous pouvons qualifier au Gabon d'ancienne et l'autre, de moderne. D'autre part, sans être exhaustif nous allons dresser un bilan de la lexicographie scolaire dans ce pays.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

1.1 La lexicographie ancienne vs lexicographie moderne au Gabon et aspects des critères scientifiques et temporels de délimitation

1.1.1 La lexicographie moderne

La recherche de la littérature sur les travaux lexicographiques au Gabon peut conduire à l'identification d'une période capitale. Il s'agit de celle à partir de laquelle, à côté des productions de dictionnaires, ces travaux se sont penchés sur les révisions des ouvrages existants comme contribution en vue de leur amélioration. On peut mentionner ceux de Mihindou (2001), Mavoungou (2011), Ondo-Mebiame (2005), Mabika Mbokou (2006b), Ndinga-Koumba-Binza (2006) et Ondo-Mebiame et Ella (2011) pour ne citer que ceux-là. Ces travaux ont été souvent réalisés en convoquant les critères de la Théorie Générale de la lexicographie moderne telle qu'énoncée par Wiegand (1984: 13-30), s'articulant autour de quatre Constituants Théoriques A, B, C et D[2]. Wiegand est considéré comme le père de la lexicographie moderne. Un aperçu sur ces travaux montre que ces derniers s'appuient de façon générale d'une part, sur le Constituant Théorique B[3]. Il s'agit premier aspect qui traite de tout ce qui est planification dictionnairique, notamment le plan de conceptualisation dictionnairique (macrostructure, microstructure, structure d'accès, médio structure, etc.). D'autre part, ces travaux renvoient également au Constituant Théorique C[4]. Il s'agit précisément de la théorie de la collecte des données lexicographiques. C'est la raison pour laquelle nous pensons que ces travaux peuvent être au Gabon, représentatifs de la lexicographie que nous pouvons qualifier de moderne. Les autres aspects de la théorie de la lexicographie ont été évidemment convoqués dans ces travaux, mais n'avons évoqué ici que les plus saillants qui apparaissent. Par ailleurs, ces travaux, sans pour autant avoir été exhaustif ici dans leur collecte, le constat selon lequel ils commencent de façon générale à partir des années 2000, s'impose.

Le travail d'Ondo-Mebiame, publié en 2005 et intitulé «De la révision du Dictionnaire de Samuel Galley» par exemple, est une illustration des caractéristiques de ces contributions qui viennent d'être mentionnées. En dehors de nombreux autres aspects, l'auteur fait état des données théoriques que sont la macrostructure et précisément des aspects liés à la lemmatisation, aux sources primaires et secondaires (Ondo-Mebiame 2005: 153-154). Dans cette même composante structurale, l'auteur évoque également les aspects liés aux utilisateurs potentiels du de cet ouvrage par rapport à leurs besoins et leurs aptitudes. Dans la même veine, il évoque aussi avec force et détails les éléments théoriques tels que la microstructure et la structure d'accès (Ondo-Mebiame 2005: 155-160 et 161-162). Á l'instar de la contribution de cet auteur, le schéma directeur de ces travaux se décline à partir d'une grille d'analyse théorique claire et cohérente sous le triptyque *avantages-inconvénients-propositions d'amélioration*. La Figure 1 ci-dessous, présente la couverture de Mavoungou (2010), le premier ouvrage complet rédigé par un auteur gabonais théorisant sur la lexicographie locale.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

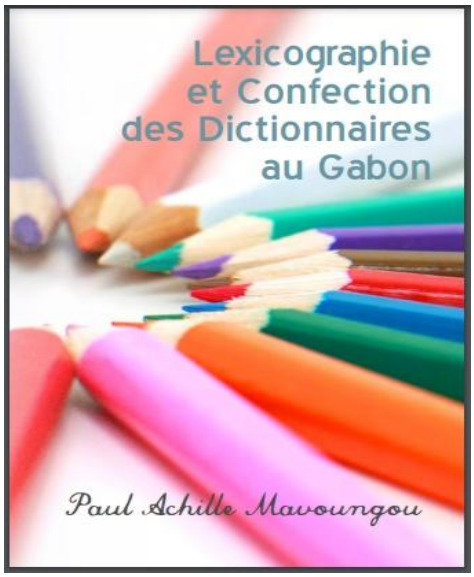


Figure 1: Couverture de Mavoungou (2010) théorisant sur la lexicographie gabonaise

Á côté de ces travaux qui ont porté sur la révision des dictionnaires existants comme contribution en vue de leur amélioration selon les critères de la Théorie Générale de la Lexicographie (telle qu'énoncée par Wiegand 1984: 13-30), il y en a également de nombreux qui, sur les mêmes bases, ont porté sur des aspects de la lexicographie gabonaise en général. Parmi ces travaux, on peut citer sans exhaustivité, Emejulu (2001 et 2002), Mavoungou (2001), cf. Figure 1, Nyangone Assam (2001), Tomba Moussavou (2002) pour ne citer que ceux-là. Un aperçu de ces travaux a conduit à l'identification de l'usage des révisions des dictionnaires existants comme contribution en vue de leur amélioration selon les critères métalexicographiques. C'est le cas du travail de Tomba Moussavou (2002), par exemple. Dans son travail, cette auteure passe au crible les aspects du public-cible (168-169), les composantes de la structure du dictionnaire, à savoir, les prétextes (169), la macrostructure (169-170), la microstructure (170) et les post-textes (171) pour une lexicographie monolingue réussie au Gabon.

Á l'instar du travail de Tomba Moussavou, tous ces travaux ont fait l'objet d'articles dans des ouvrages. Ce qui n'enlève rien à leur importance et qualité. Il y a un livre, intitulé *Lexicographie et Confection des Dictionnaires au Gabon* (Mavoungou 2010), cf. Figure 1, que nous pensons est emblématique et symbolise cette nouvelle démarche dans la lexicographie gabonaise. Autrement dit, celle qui consiste à réfléchir et à théoriser sur cette lexicographie gabonaise qui marque d'une pierre blanche la réalité selon laquelle à côté de la production des dictionnaires, il y a dorénavant une métalexicographie locale. Fort de tout ce qui vient d'être

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

dit, nous pouvons émettre les hypothèses qui suivent:

— La révision des dictionnaires existants comme contribution en vue de leur amélioration ainsi que la théorisation de la lexicographie gabonaise selon les critères de la Théorie Générale de la lexicographie (telle qu'énoncée par Wiegand 1984: 13-30), peuvent être ainsi vues comme critère scientifique qui permet de différencier au Gabon, la lexicographie moderne de celle qui sera considérée comme ancienne.
— L'année 2000, peut être envisagée comme la délimitation dans le temps de ces deux lexicographies.

1.1.2 La lexicographie ancienne

Au regard de ce qui a été dit dans les lignes précédentes, nous pouvons émettre l'hypothèse selon laquelle, la lexicographie ancienne va regrouper les travaux qui ont été produits avant l'année 2000. Symboliquement ou de façon factuelle, ces travaux n'ont pas porté sur les révisions des dictionnaires existants comme contribution en vue de leur amélioration selon les critères de la Théorie Générale de la lexicographie telle qu'énoncée par Wiegand (1984: 13-30, et 1989: 246-280). La recherche de la littérature sur ces travaux ainsi identifiés conduit essentiellement aux travaux des missionnaires religieux. C'est-à-dire que symboliquement ou de façon factuelle, la lexicographie ancienne au Gabon peut renvoyer à ces travaux. Á ce propos, nous pensons que le *dictionnaire français–fang et fang–français* de Galley (1964) communément appelé le «Galley» est emblématique et symbolique de cette lexicographie. La Figure 2 ci-dessous est présente la couverture intérieure du dictionnaire de Galley (1964).

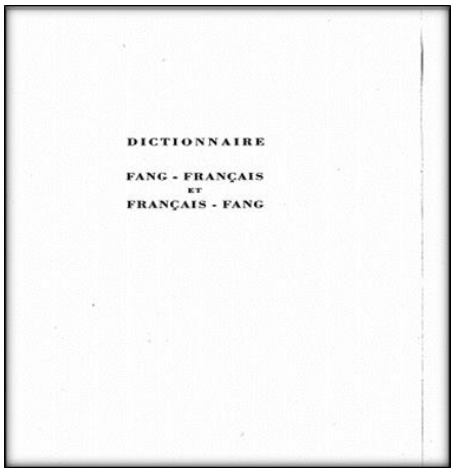


Figure 2: Couverture de Galley (1964): un classique de la lexicographie ancienne au Gabon

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Pour en avoir une idée assez claire et précise de ces travaux qui ont été les germes de la lexicographie au Gabon, nous présentons ci-dessous, une esquisse d'inventaire non exhaustif de ces travaux. Cet inventaire est établi par ordre chronologique et se présente ainsi que suit:

— 7 (sept) Dictionnaires monodirectionnels ou monoscopals

1. *Dictionnaire Fang–Français*: Marling, R.P. New York. 1872.
2. *Dictionnaire Francais–Fang ou Pahouin, précédé de quelques principes grammaticaux sur cette même langue*: Lejeune, L. Paris: A. Faivre et H. Teillard. 1872.
3. *Dictionnaire français–mpongoûé: Maisonneuve et Cie.* Missionnaires de la congrégation du St Esprit et du saint cœur de Marie. Paris. 354 pp. 1877.
4. *Dictionnaire français–vili:* Marichelle, C. Loango. 1900.
5. *Dictionnaire français–mpongwé, suivi d'éléments de grammaire*: Imprimerie La Libre Lorraine, XVII-640. (Réédité en 1961, Brazzaville, Imprimerie St. Paul). Metz. 722 pp. 1934.
6. *Dictionnaire français–mpongwé*: Raponda Walker. Classes Africaines. 2ème édition 1995. 742 pp. 1961.
7. *Dictionnaire ghetsogo–français*: Raponda Walker. Nd.

— 5 (cinq) Dictionnaires bidirectionnels ou biscopals

8. *Dictionnaire français–pongoué/pongoué–français:* Mgr Bessieux. Ed. Lenoël-Hérouart, en deux volumes 52 pp. 1847. Mission du Gabon. Maisonneuve et Cie.
9. *Dictionnaire français–ndumu et ndumu–français précédé d'éléments de grammaire***:** Biton, A. Nantes, Imp. Bourgeois. 1907.
10. *Dictionnaire français–fang et fang–français*: Galley, P.S, Neuchatel, Henri Messeiller. 588 pp. 1964.
11. *Dictionnaire français–yipounou/yipounou–français*: Eglise Evangélique du Sud Gabon. 1966.
12. *Dictionnaire ndumu–mbede–français et français–ndumu–mbede***:** Archevêché de Libreville. Petite flore de la région de Franceville (Gabon), Bar-le-Duc, Imp. St Paul. ADAM, J.J. 1969.

— 1 (une) Encyclopédie

13. *Encyclopédie Pahouine. Éléments de Grammaire et Dictionnaire Francais–Pahouin*: Largeau, V. Paris: E. Leroux. 1901.

— 1 (un) Dictionnaire spécialisé ou de langue de spécialité

14. *Le dictionnaire des noms propres du Gabon.* Raponda, Walker. (Éd.). 1993.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

1.2 Aspects de la lexicographie scolaire au Gabon

Pour analyser ces aspects, nous allons d'abord évoquer les généralités sur le dictionnaire scolaire qui renvoie à la lexicographie scolaire, avant de discuter d'une part, de l'état de traitement des données de cette lexicographie dans les dictionnaires existants. D'autre part, il sera question de se pencher sur l'état de confection de dictionnaires de lexicographie scolaire.

1.2.1 Généralités sur le dictionnaire scolaire

Nous allons seulement évoquer ici les généralités sur le dictionnaire scolaire, car les aspects sur cet ouvrage ont déjà été largement développés et sont connus avec en prime un travail, celui de Nyangone Assam (2002: 292-304), adapté au Gabon[5]. Nous allons donc donner un bref aperçu des caractéristiques de ce dictionnaire. Il s'agit d'une part, des caractéristiques typologiques de cet ouvrage et d'autre part, des aspects saillants de ses macro et microstructures par lesquels cet ouvrage se démarque. Après avoir discuté de la lexicographie ancienne et de la lexicographie moderne et des modalités de leur délimitation, l'autre nœud de ce travail et qui en est le plus déterminant est de démontrer l'articulation entre la lexicographie scolaire et le processus de standardisation des langues locales du Gabon.

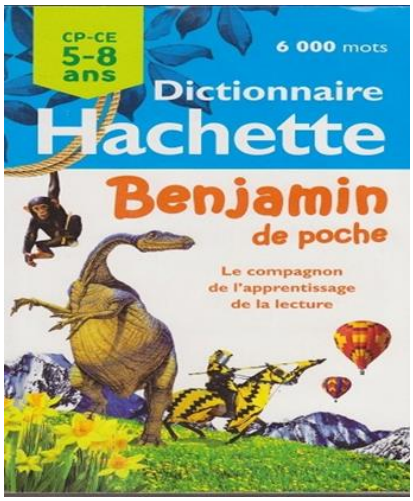


Figure 3: Couverture du *dictionnaire Hachette Benjamin* Junior de Mével (2011)

Le dictionnaire scolaire est un ouvrage général monolingue conçu pour son utilisation dans l'enseignement primaire et secondaire (Boulanger 2006: 3) et représente aussi une catégorie de dictionnaire spécialisé ou de langue de spécialité

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

ou encore restreint (Lombard 1990). Cependant, le dictionnaire scolaire peut avoir un sens plus élargi, car ainsi qu'il le précise, d'une part, on peut en étendre l'acception du côté de l'école maternelle et, d'autre part, du côté du début de la formation post-secondaire ou collégiale (Boulanger 2006: 3). On peut dire que ce type de répertoire couvre une fourchette d'âge qui s'étend de 5/6 ans jusqu'à 15/16 ans (Boulanger 2006: 3). L'image ci-dessus qui est celle d'un dictionnaire des 5–8 ans est une parfaite illustration des propos de cet auteur.

Le dictionnaire scolaire est une sous-catégorie du dictionnaire pédagogique et auquel il se confond[6]. Boulanger (2006: 3), décline les objectifs comme étant la sensibilisation des jeunes au vocabulaire et de les mettre à l'aise avec les mots. Il ajoute que le dictionnaire scolaire consigne le lexique vivant, dont une partie est déjà connue de la majorité des élèves, soit de manière active, parce que les jeunes locuteurs doivent s'exprimer, soit de manière passive, parce qu'ils doivent écouter et comprendre les autres. Cela est dit autrement par Gouws (2001: 113) qui définit le dictionnaire scolaire représente le vocabulaire de base que les écoliers emploient pour les conversations normales types et avec lequel ils travaillent à l'école.

En ce qui concerne les aspects saillants de ses macro et microstructure par lesquels cet ouvrage se démarque comme nous l'avons indiqué en supra, sa macrostructure est restreinte et affiche une faible densité des données avec un rapport quantitatif équilibré des entrées avec la microstructure (Gouws 2001: 113). Comme cela est illustré dans la Figure 4 ci-dessous, la macrostructure du dictionnaire scolaire est souvent renforcée par l'iconographie, car cet ouvrage a une fonction avant tout d'ordre descriptif (Boulanger 2006: 3)[7].

Figure 4: Renforcement de la macrostructure par l'iconographie dans la nomenclature du *dictionnaire Hachette Benjamin* Junior de Mével (2011)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Cette structure d'accès est conçue pour aider d'une manière fonctionnelle une catégorie spécifique d'utilisateurs que sont les jeunes élèves qui évoluent de l'école maternelle jusqu'au début de la formation post-secondaire ou collégiale. De ce fait, il important que la fonction pédagogique règne à tout moment dans le dictionnaire scolaire (Gouws 2001: 113). C'est la raison pour laquelle, on observe que la quantité d'illustrations est inversement proportionnelle à l'âge des utilisateurs cibles: plus l'âge croît, plus le nombre d'illustrations diminue.

1.2.2 État des lieux de la confection des dictionnaires de lexicographie scolaire

La recherche sur la littérature nous a conduit à identifier quelques dictionnaires scolaires existants. Sans être exhaustif, deux œuvres se démarquent de cette exploration et au premier regard cela pourrait sembler famélique. Ce qui n'est pas le cas. Il s'agit de *Le dictionnaire des Mwanas «avec Papito et Jolica»* (Unik Afrika 2013) et de *Mon Premier Dictionnaire Bilingue Français–Fang* (Collectif 2013). En effet, cela relève d'une fausse impression que ces dictionnaires sont faciles à confectionner et n'exigent qu'une simple approche de «couper-coller» en les extrayant des ouvrages classiques plus volumineux (Gouws 2001: 113).

— *Le dictionnaire des Mwanas*

Généralités

Á quelques légères différences près, *Mwanas* veut dire «enfants» dans toutes les langues locales du Gabon. Il était difficile de passer à côté de cet ouvrage qui était vendu sur les grandes surface de Libreville, la capitale du Gabon. Autorisé dans les programmes scolaires au dans ce pays, c'est une collection qui se décline sur 5 des principales langues nationales du Gabon. Il s'agit par ordre alphabétique du *fang*, de l'*inzébi*, du *lembaama*, de l'*omyènè* (sur l'image en Figure 5 ci-dessous) et du *yipunu*.

Dans les propos introductifs, il est dit que ce manuel permet aux petits enfants et débutants d'acquérir un vocabulaire de base dans ces différentes langues locales. Les atouts d'apprentissage de cette méthode étant: des angles arrondis; des illustrations simples, ludiques et en couleur pour une bonne compréhension; des personnages de race noire âgés de 4 à 5 ans pour l'appropriation et l'identification; un apprentissage pas à pas, organisé par thème et une contextualisation des mises en scène. Ces propos liminaires mentionnent également que la collection *Le dictionnaire des Mwanas* est le fruit d'un travail de deux longues années de préparation et de perfectionnement imaginé, rédigé, corrigé et traduit par des professionnels gabonais, afin de produire un enseignement de qualité aux enfants gabonais. Elle conclut par les propos selon lesquels, l'intégration de l'apprentissage des langues nationales dès le jeune âge est nécessaire pour la

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

pérennisation du patrimoine culturel du Gabon. Ce qui fait de *Le Dictionnaire des Mwanas*, un investissement important pour l'avenir des enfants et de la nation.

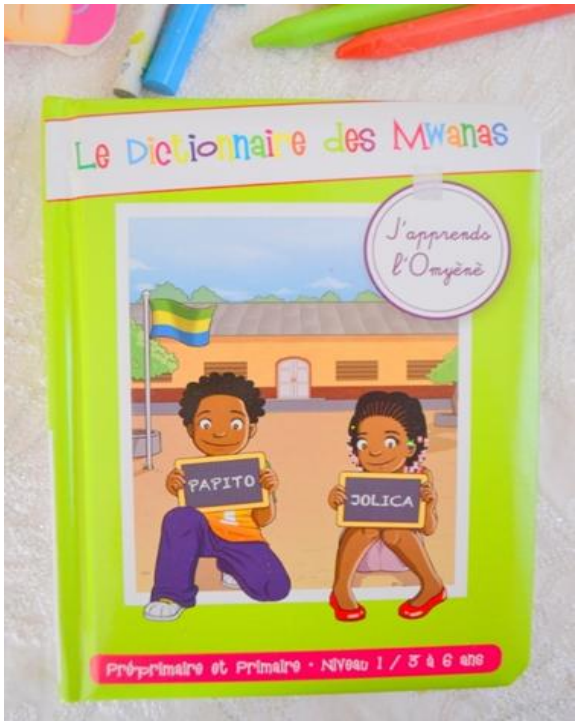


Figure 5: Couverture du *Dictionnaires des Mwanas «J'apprends l'Omyènè»*

Structure du dictionnaire

La Figure 6 ci-dessous est une page de la nomenclature de l'ouvrage qui peut être analysée comme un article de dictionnaire. C'est-à-dire, l'association d'un élément macrostructural autrement appelée entrée ou lemme, et son traitement microstructural (Gouws 2001: 122).

Dans l'article de dictionnaire de ces ouvrages, c'est l'image (la Figure 6 montre le fruit connu sous le nom d'*atanga* au Gabon et de *prune* ou *safou* dans d'autres pays d'Afrique centrale) qui fait office de l'entrée et non un élément lexical. Cette image est immédiatement précédée du traitement microstructural qui est le nom en *yipunu* de ce fruit, à savoir *dusafu* [dúsàfù]. L'ouvrage scolaire *Le Dictionnaire des Mwanas* obéit ainsi à l'architecture du dictionnaire scolaire, à savoir une structure à trois niveaux qui se répartie ainsi que suit: la macrostructure, la microstructure et l'iconostructure. La dernière, l'iconostructure, ayant autant

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

d'importance pédagogique que les deux autres structures comme aime à le rappeler (Boulanger 2006: 14).

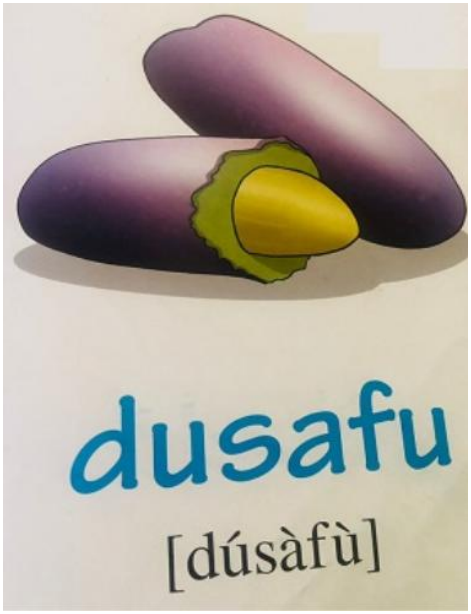


Figure 6: Article du *Dictionnaire des Mwanas «J'apprends le Yipunu»*, Unik Afrika (2013)

En faisant immédiatement précéder l'image du traitement microstructural, *dusafu*, la microstructure de l'ouvrage *Le Dictionnaire des Mwanas* est du type «intégré» et correspond à ce type de dictionnaire[8]. Il y a une relation d'adressage direct entre la paraphrase du sens, *dusafu,* et son entrée qui est l'image, facilitant ainsi l'accès au jeune utilisateur. C'est à ce propos que Gouws (2001: 113) dit que la structure d'accès d'un dictionnaire scolaire devrait être conçue pour aider, d'une manière fonctionnelle, la catégorie d'âge spécifique identifiée qui est constituée par des enfants en âge d'apprentissage.

Toujours dans la microstructure, un aspect passe difficilement inaperçu ici. Il s'agit de l'absence dans l'article du dictionnaire des équivalents en français des données présentées comme éléments du traitement microstructural. Ces derniers sont uniquement en langues locales. Un lexique ou une liste d'équivalence avec le français comme entrée est présenté à la fin de l'ouvrage et sur lequel nous allons revenir. C'est une méthode très avantageuse pour les jeunes gabonais qui vont directement associer l'orthographe du mot en reconnaissant l'image sans passer par le français. Cependant, cela peut être un inconvénient pour les non Gabonais qui ne reconnaitront pas dans quelques rares cas, ce que repré-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

sentent l'image. C'est notamment le cas de l'*atanga* ici, par exemple. Néanmoins, ce type de cas reste heureusement assez isolé, car les autres fruits tels que l'ananas, la banane, etc. sont reconnaissables de tous.

Du point de vue de la prononciation, l'image utilisée comme entrée du dictionnaire est un atout pour les jeunes enfants locuteurs de leurs langues endogènes. En dehors de l'orthographe du mot, ils vont aussi aisément associer la prononciation. C'est la raison pour laquelle Boulanger (2006: 14) indique à juste titre que dans les dictionnaires scolaires, l'iconostructure a autant d'importance pédagogique que la macro- et la microstructure. Toutefois, cela ne va pas constituer un obstacle pour les non locuteurs. Ces derniers pourront acquérir indirectement la prononciation par l'écoute en demandant ou en le percevant lors d'une conversation. En effet, l'acquisition des sons est chez l'homme est la capacité à reproduire les sons qu'il entend.

L'autre aspect saillant qui caractérise l'article du dictionnaire dans l'ouvrage *Le Dictionnaire des Mwanas* est la présence des aspects liés à la prononciation, à travers la transcription phonétique, cf. [dúsàfù] *dusafu*. Dans ce type de dictionnaire, les rubriques historiques (étymologie, datation) et la prononciation sont exclues dans le traitement microstructural par conformisme et convention, cf. Boulanger (2006: 16). En effet, c'est en début de collège et lycée que l'apprentissage de la phonétique est inscrit au programme. Cette transcription n'est leur est donc pas destinée. Par contre, cette dernière peut être utile pour les personnes qui ont la maîtrise des notions de base de la phonétique. Il peut s'agir des non gabonais qui veulent bien faire acquérir à leur progéniture le vocabulaire de base des langues autochtones de leur pays hôte. Il peut et surtout aussi s'agir des enseignants et des parents.

Ces derniers bénéficient des données sur les conventions orthographiques qui leur permet de faire le lien entre l'orthographe et la prononciation. Comme l'indique Boulanger (2006: 13):

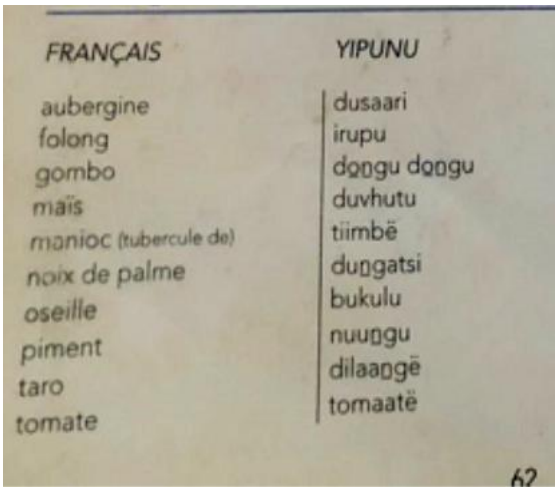
> «Le dictionnaire pédagogique est un ouvrage de référence sur la langue et sur le monde qui s'utilise d'abord à l'école de manière collective, orientée, puis à la maison de manière plus personnelle et plus spontanée, sans exclure une utilisation dirigée par les parents à l'occasion des devoirs, par exemple».

Ce qui vient d'être dit pour la transcription phonétique est également valable pour le lexique ou la liste d'équivalence avec le français comme entrée qui est présenté à la fin du dictionnaire, cf. Figure 7 sur la page suivante. Dans le dictionnaire scolaire, la liste centrale domine et un nombre limité de textes fonctionnent comme textes externes (Gouws 2001: 113). Cela devrait l'être encore plus ici avec *Le Dictionnaire des Mwanas* par rapport à la jeune tranche d'âge auquel il est destiné (3–6 ans). Il s'agit du niveau 1 de l'apprentissage. Ce lexique sera donc plus utile aux enseignants et aux parents, car il ne sera pas commode ou pratique pour les enfants pour qui ce renvoi d'article externe est assez complexe[9]. Par ailleurs, ce lexique aurait eu à gagner si les entrées correspondaient aux formes présentées dans la microstructure.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Plus précisément, l'élément microstructural, *dusaari* par exemple, dans le *Dictionnaire Français–Yipunu* (Unik Afrika 2013), ne renvoie pas l'utilisateur de cette position de référence à une adresse de référence qui aurait dû être *dusaari* suivi de son équivalent en français *aubergine*. En effet, si l'utilisateur ne sait pas à quoi renvoie *dusaari* malgré l'image par exemple, il devrait être contraint de regarder à l'intérieur du lexique et sans être guidé par l'ordre alphabétique qui y est de fait aléatoire. Dans ce lexique, la langue cible n'est pas la langue locale, mais le français, langue dans laquelle l'équivalent est cherché. La langue locale devrait donc se positionner ici comme la langue source avec ses items lexicaux comme entrées afin que ce lexique joue pleinement sa fonction.

La Figure 7 ci-dessous présente un extrait du lexique français/yipunu du le *Dictionnaire Français–Yipunu* (Unik Afrika 2013). On peut noter l'item lexical *dusaari* qui vient d'être mis en relief dans le paragraphe précédent.



FRANÇAIS	YIPUNU
aubergine	dusaari
folong	irupu
gombo	dɔŋgu dɔŋgu
maïs	duvhutu
manioc (tubercule de)	tiimbë
noix de palme	duŋgatsi
oseille	bukulu
piment	nuuŋgu
taro	dilaaŋgë
tomate	tomaatë

62

Figure 7: Extrait du *Dictionnaire Français–Yipunu* contenant l'item *dusaari*

— *Mon premier dictionnaire bilingue français–fang*

Généralités

Mon Premier Dictionnaire Bilingue Français–Fang (MPDBFF) du Collectif (2013), voir Figure 8 ci-dessous). *MPDBFF* est le deuxième ouvrage qui a été identifié dans notre recherche sur la littérature de la lexicographie gabonaise et choisi pour faire l'objet de l'analyse de ce travail. Il fait partie d'une collection d'imagiers ludiques faciles d'accès pour maîtriser dès le plus bas âge le français et des langues endogènes souvent rares et considérées comme en danger. Au premier regard, *MPDBFF* obéit aux particularités du dictionnaire scolaire avec une mise

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

en page attrayante, riche en illustrations et en couleurs. C'est un ouvrage bilingue monodirectionnel ou monoscopal, car il n'y a pas le pan fang–français. C'est-à-dire que toutes les entrées sont en français. Ce qui correspond toujours aux caractéristiques du dictionnaire scolaire qui est un ouvrage général monolingue et quelque fois bilingue conçu pour son utilisation dans l'enseignement primaire et secondaire (Boulanger 2006: 16).

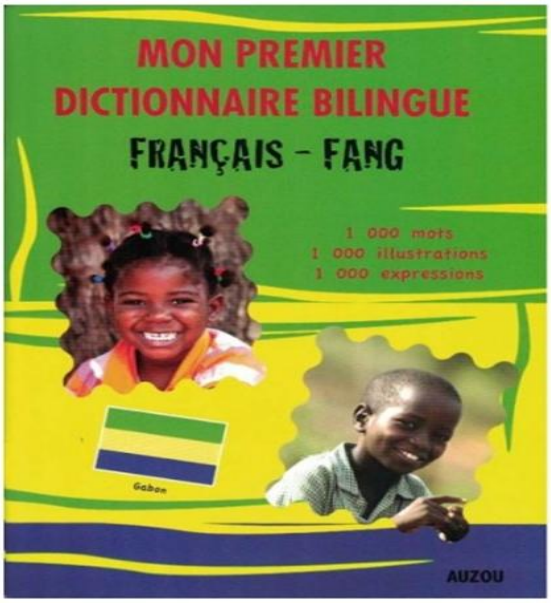


Figure 8: Couverture du *Mon Premier Dictionnaire Bilingue Français–Fang* du Collectif (2013)

Selon les éléments de description recueillis, *MPDBFF* présente les données suivantes: 1000 mots parmi les plus courants en français et en fang; 1000 phrases simples (sujet + verbe + compléments ou adjectif) traduites en français et en fang pour mettre les mots en situation; 1000 illustrations correspondant au nombre d'entrées; des planches thématiques (les animaux sauvages et aquatiques, les fruits, et un index) Pas étonnant qu'avec *Le Dictionnaire des Mwanas* (Unik Afrika 2013), ces deux ouvrages soient considérées selon notre avis, comme étant les têtes de gondole de la lexicographie scolaire au Gabon.

Structure du dictionnaire

La Figure 9 ci-dessous est une illustration du tiers d'une page de la nomenclature de *MPDBFF* (Unik Afrika 2013).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Figure 9: Illustration de la nomenclature de *MPDBFF* (Unik Afrika 2013)

La présentation d'une page entière donne une image complètement floue. Le tiers de la page l'est moins. Il y a dans chaque page, 10 images qui peuvent être analysées sans ambiguïté, comme des articles de dictionnaire comme cela a été le cas avec *Le Dictionnaire des Mwanas* (Unik Afrika 2013). Alors, chacun de ces articles de dictionnaire, présente l'entrée en français qui est immédiatement suivie de son équivalent en fang dans deux couleurs distinctes. C'est le cas avec *chemin* et son équivalent en fang *zen,* par exemple. En dessous de ces deux données, il y a une iconographie. Cette dernière dont la présence est plus sensible dans les ouvrages à petite nomenclature, comme c'est le cas ici, renforce la fonction du dictionnaire scolaire qui est avant tout d'ordre descriptif, la description étant le plus souvent renforcée par l'iconographie (Boulanger 2006: 3).

Cet auteur ajoute dans le même ordre d'idées que simultanément à sa fonction linguistique, le dictionnaire scolaire s'ouvre directement sur le monde, soit par la formule de la définition, soit par les exemples soit par l'iconostructure. Cette dernière, singularise l'entrée et la replace souvent dans un environnement réel, à savoir pluri-conceptuel. C'est le cas ici avec l'image dans laquelle il y a deux jeunes hommes empruntant un *chemin*. La troisième composante si on peut le dire de l'article du dictionnaire de *MPDBFF* est constitué par deux co-textes qui sont quasiment illisibles sur l'image[10]. Le premier *Le chemin est une petite route* rend compte de l'emploi en français comme étant représentatif de l'environnement syntaxique de l'élément lexical traité *chemin* dans lequel il se produit typiquement ou est généralement employé. Le deuxième co-texte rend compte de la traduction en fang *Zen ene ndzong ntoa*.

Au premier regard, *MPDBFF*, ouvrage monodirectionnel peut apparaître à tout égard comme étant destiné en premier lieu plus aux locuteurs français que

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

ceux du fang. Cela relève souvent d'une fausse impression à propos du dictionnaire monolingue que c'est la langue qui précède dans l'intitulé et dont les mots sont les entrées qui est la langue cible. C'est l'inverse. Avec les entrées en français, l'objectif est de satisfaire les besoins en fang de l'utilisateur qui est locuteur du français. Cette fonction est renforcée par le co-texte qui rend compte de l'emploi en français comme étant représentatif de l'environnement syntaxique de l'élément lexical traité et celui qui rend compte de la traduction en fang. Cette architecture de l'article du dictionnaire de *MPDBFF* permet à l'utilisateur de recevoir et de produire des textes en fang.

Par ailleurs, à l'instar des langues locales du Gabon, le fang n'est pas encore formellement standardisé. Le français qui l'est évidemment, demeure encore la seule ayant le statut officiel. Ajouté à cela, les jeunes qui constituent la majeure partie de la population ont le français comme langue maternelle (Mabika Mbokou 2006a: 4). De ce fait, qu'il soit locuteur ou pas du fang, l'utilisateur fang présente les caractéristiques d'un locuteur du français. Cela fait de *MPDBFF* un dictionnaire de fang quasiment au même titre qu'un dictionnaire monolingue. Il peut être alors intitulé *Mon premier dictionnaire fang* comme cela est illustré sur l'image ci-contre. Cela est possible dans le cadre d'une révision et d'une réédition de cet ouvrage. Les mentions *dictionnaire bilingue français–fang* vont apparaître comme des sous-textes descriptifs.

Si les entrées en fang de *MPDBFF* étaient en fang, l'objectif aurait été de satisfaire les besoins de connaissance en français de l'utilisateur locuteur du fang. Dans ce cas, l'ouvrage aurait été titré *Mon Premier Dictionnaire Bilingue de Français* avec les mentions *dictionnaire bilingue fang–français* comme des sous-textes descriptifs.

Remarques conclusives

Pour une meilleure analyse de deux ouvrages que nous avons répertoriés comme étant des dictionnaires scolaires, nous avons utilisé les outils de la théorie de la lexicographie en considérant les images qui constituent la nomenclature comme des articles de dictionnaire à part entière. Comme nous l'avons dit en l'entame de cette section, si deux ouvrages identifiés semble dérisoire au premier abord, cela reste une satisfaction à y regarder de près. En effet, les dictionnaires scolaires ne se confectionnent pas comme les ouvrages généraux d'une part, et ils ne résultent pas d'un simple «couper-coller» de ces derniers, d'autre part.

Ce sont des ouvrages qui ont une conception à part entière, qui sont beaucoup plus complexes qu'ils en ont l'air. En effet, ils exigent beaucoup plus d'attention et de subtilité qu'on ne le pense. Pour en donner une idée, nous reprenons les propos de Boulanger (2006: 3) ainsi que suit:

> Sur le plan des contenus, le souci pédagogique et son corollaire le souci de la protection morale de l'enfant prônent une sélection soigneuse et serrée des entrées et des contenus des articles. Les lexicographes qui explorent la constellation pédagogique ont tendance à écarter les mots rares ou littéraires, mais surtout ils

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

> éliminent les mots tabous ou à contenu tabou, ils font une croix sur les mots peu connus ou non filtrés par l'institution enseignante, ils délaissent le vocabulaire trop spécialisé, pour ne s'arrêter que sur les couleurs du lexique socialement assimilables. Ce programme éditorial aboutit à concevoir une géométrie dictionnairique qui repose sur un ordre conçu comme la régulation des usages. Cette régulation s'appuie souvent sur des décisions préalables à la description, c'est-à-dire sur ce qui précède la rédaction proprement dite — les mots écartés de la nomenclature —, plutôt que sur la prescription inscrite dans les articles.

Coïncidence ou fait de hasard, *Mon Premier Dictionnaire Bilingue Français–Fang* (Collectif 2013) paraît la même année que *Le Dictionnaire des Mwanas* (Unik Afrika 2013). Ces deux ouvrages pourraient être vues comme les symboles ou les productions emblématiques de la lexicographie scolaire au Gabon. Il a été jusqu'ici question, d'une part, de delimiter dans le temps et par la nature des activités dictionnairiques du point de vue pratique que théorique la lexicographie ancienne de la lexicographie moderne. D'autre part, il s'est agi d'identifier et d'analyser deux ouvrages majeurs de la lexicographie scolaire au Gabon. Dans les lignes qui suivent, nous allons discuter de l'articulation entre cette lexicographie et le processus de standardisation des langues locales.

2. La lexicographie scolaire dans le processus de standardisation des langues locales

Pour évoquer la question de la lexicographie scolaire dans le processus de standardisation des langues locales du Gabon, il est d'abord important de discuter du type de dictionnaire scolaire qui convient aux réalités objectives de ce pays: dictionnaire scolaire bilingue monodirectionnel ou monoscopal ou dictionnaire scolaire bilingue bidirectionnel appelé biscopal?

2.1 Le choix des dictionnaires bilingues monodirectionnels français et langues locales

En analysant *MPDBFF*, nous avons déjà insinué que notre choix portait sur le dictionnaire scolaire bilingue monodirectionnel. Nous avons évoqué des arguments relatifs aux commodités d'usage centrées sur l'utilisateur en termes de satisfaction des besoins communicationnels effectifs des enfants gabonais qui doivent être considérés comme locuteurs du français. Nous allons ajouter des arguments concernant le développement de la lexicographie scolaire au Gabon de façon générale.

Dans le cas du Gabon où les langues locales ne sont pas encore assez exposées et vulgarisées à l'écrit, la confection préalable des ouvrages de type bilingue monodirectionnel présente des avantages dans ce sens. Cela obéit à des principes lexicographiques définissant l'ordre prioritaire dans les situations dans lesquelles il y a une langue standardisée et des langues périphériques ne l'étant pas encore (Gallardo

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

1980). Par ailleurs, cela va permettre en amont, de constituer d'une part, une réserve importante de fonds linguistique et culturel. D'autre part, cela va permettre de constituer une documentation et un traitement métalinguistique et métalexicographique sur les langues, les cultures et les dictionnaires. Ce, en attendant la confection d'ouvrages de type bilingue bidirectionnel et monolingue, facilités qu'ils seraient par des dictionnaires bilingues de type monodirectionnel moins naïfs et plus performants (Kalonji 1993: 25). Nous allons illustrer nos propos en nous référant au «Galley» (1964) avec le traitement des items lexicaux *manga* (p. 178) */femme* (p. 467).

Partie fang–français

MANGA (h) n.l, pl. banga. Femme
Syn.: mônga, pl. bônga.

Partie français–fang

FEMME mônga (bh), manga (bh). Femme, épouse, ñgal (h), pl. beyal. Ma

femme, ñga wam (hb), ébo zam (bb). Femme mûre d'environ 40 ans, ésila (b). La femme est successivement: moñe (bm) (enfant), étoñ (b) (jeune fille), évôm (b) (nubile), ôyañgon (h) (30 à 40 ans), ésila (b) (mûre), ôbañ-sila (hh) (50 ans), nnôm (m) (vieille). Femme donnée au gendre en remplacement de celle qui est morte, sans renouveler la dot, ésokh bikî (hb). Jeune femme qui est sous les ordres de sa belle-mère, ñkîa (bm), mbom (h), ou sous les ordres d'une femme du même mari plus âgée. Jeune femme, moneweñya a mônga, moñe a ñgon (bmb). Première femme épousée, ôsu ô mônga, ntôl ô mônga. Femme acquise
avec la dot de notre fille, femme acquise par le beau-père avec la dot reçue de son gendre, mvia (h). Notre fille qui a été échangée ainsi, c'est notre mbala (h). Femme qui a cessé de nourrir l'enfant, qui n'est plus nzensè, nsil (b). Femme enceinte ou bête portante, nyabmume (h).

C'est dans la partie français–fang que l'article du dictionnaire relatif à l'élément lexical *femme/mônga* est le mieux traité. Cela est assez nettement dû au fait dans cette partie, ce traitement s'appuie celui d'un dictionnaire de français et il apparaît que le traitement de *mônga* (femme) n'est pas encore familier à l'activité dictionnairique. D'où, ce déséquilibre assez visible entre les deux traitements du même lemme. Par conformisme et par convention et également pour se donner de l'assurance par rapport à la réussite du projet, les lexicographes se réfèrent aux dictionnaires existants et ce pour avoir des modèles Kavanagh (2002: 276).

En d'autres termes, le «Galley» pouvait d'abord se présenter sans ambiguïtés comme un dictionnaire bilingue monodirectionnel et il aurait largement satisfait les locuteurs ou pas sur les besoins de la connaissance en langue fang. Cela reste toujours possible aujourd'hui dans le cadre d'une révision et d'une réédition de

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

cet ouvrage. C'est à terme et après s'être suffisamment engrangé d'expérience et avoir collecté des données quantitatives et qualitatives que la version bilingue bidirectionnel voire pourra être envisagée avec beaucoup plus de sérénité et de réussite.

2.2 Aspects dans le processus de standardisation des langues locales

La lexicographie scolaire a participé au renforcement de la standardisation de la langue française. Boulanger (2006) en fait une démonstration assez claire à partir de laquelle nous allons simplement faire une transposition dans le contexte du Gabon. Il est important de signaler que l'un des acteurs-clés des débuts de la lexicographie scolaire fut Pierre Larousse. Il est le précurseur de l'iconostructure et celui qui a marqué davantage la vocation encyclopédique des dictionnaires scolaires préparés par ses successeurs, notamment par Claude Augé, cf. Boulanger (2006: 1). Les dictionnaires scolaires vont activement participer à la standardisation des langues locales du Gabon pour diverses raisons qui seront énumérées dans les lignes qui suivent.

2.2.1 Sous l'angle linguistique

Sous l'angle linguistique, les dictionnaires scolaires vont introduire les enfants gabonais dans le monde du lexique et des langues locales du Gabon. Ils vont initier à la lecture fragmentée d'un texte à la grammaire et à la rhétorique particulières. En un mot, les textes dictionnairiques de ces ouvrages vont prolonger, compléter et stabiliser l'enseignement de base des langues locales du Gabon reçu à l'école dans d'autres matières. En effet, le dictionnaire scolaire traite aussi d'autres disciplines en dehors de la langue telles que les mathématiques, la physique, SVT, etc. Ces dictionnaires sont ainsi résolument modernistes, laïcs et synchroniques, du moins pour le vocabulaire commun. Trois (3) caractéristiques sont dénombrables:

1. Les dictionnaires présentent un portrait vivant du vocabulaire d'usage. Les nomenclatures sont normatives en étant accueillantes à quelques nouveautés lexicales.
2. Les terminologies modernes sont prises en compte.
3. Plusieurs articles comportent des développements encyclopédiques consistants.

À la périphérie, les dictionnaires scolaires vont introduire des Vocabulaires Généraux d'Orientation Scientifique (VGOS) dans les langues locales du Gabon. Il s'agit ici des blocs de vocabulaires techniques et scientifiques d'utilité générale, et des tranches de vocabulaires semi-spécialisés, voire spécialisés, sans toutefois pousser du côté des terminologies ultraspécialisées.

L'apprentissage du vocabulaire des langues locales du Gabon que va susciter

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

les dictionnaires scolaires et accompagné de leur usage satisfaisant renferment des objectifs subsidiaires. Les enfants gabonais vont bénéficier du maximum des informations contenues dans ces ouvrages, tant au regard de la compréhension des mots, à savoir de leur décodage, qu'à celui de la production de textes écrits ou oraux, à savoir de leur encodage. À cela, s'ajoute en filigrane l'analyse de quelques fonctions fondamentales du système de la langue. Les habiletés à acquérir du point de vue de l'apprentissage global des langues locales seront les suivantes:

1. Apprendre à dominer l'orthographe des mots, la lecture du dictionnaire conduisant à la maîtrise de l'écriture. Il s'agit en fait de pouvoir répondre à la question: «Ça s'écrit comment?»
2. Apprendre à situer les mots dans les différentes parties du discours. Cet objectif d'ordre grammatical permet de répondre aux questions: «Quelle sorte de mot est-ce?», «C'est quoi ce mot?»
3. Savoir reconnaître les différentes acceptions ou définitions d'un mot, c'est-à-dire les relations sémantiques. Il s'agit en fait de pouvoir répondre à la question: «Qu'est-ce que ça veut dire?»
4. Maîtriser les règles fondamentales du code par le réemploi dans la composition, la rédaction de textes, la préparation de devoirs. Passage du passif à l'actif, du décodage à l'encodage.

Les dictionnaires scolaires seront donc réservés à l'apprentissage des fondements de la langue et cela sera nécessairement utile pour les langues locales du Gabon: vocabulaire, orthographe, écriture, lecture. Cependant, l'apprentissage du vocabulaire sera l'objectif fondamental des dictionnaires scolaires au Gabon. Cette visée sera d'abord d'ordre quantitatif; il s'agira d'apprendre à maîtriser un volume de mots et de sens de plus en plus nombreux afin de s'arrimer au monde de la vie quotidienne et à l'univers des connaissances à travers le prisme des langues locales. Les objectifs secondaires sont de l'ordre de l'orthographe, de l'écriture et de la lecture de ces langues. Chez l'enfant gabonais, la découverte du monde ira de pair avec l'apprentissage des phénomènes langagiers, surtout les mots des langues locales du Gabon.

2.2.2 Sous l'angle social, pédagogique et de la pratique dictionnairique

Sur le plan social, les dictionnaires scolaires occupent à terme une place prépondérante dans le système pédagogique de l'enseignement des langues. Cela est renforcé par la recommandation de l'État de rendre obligatoire le maniement obligatoire du dictionnaire à l'école. Dans cette veine, les directives ministérielles vont stipuler qu'il faut multiplier les occasions de recourir au dictionnaire, d'entraîner les enfants à se servir de ces ouvrages le plus rapidement possible, et surtout d'utiliser des dictionnaires confectionnés pour eux.

À l'école, les dictionnaires scolaires seront des ouvrages qui seront des outils d'information sur les langues locales du Gabon, qui seront adaptés aux

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

objectifs de l'enseignement de ces langues afin de réaliser des activités pédagogiques et de répondre aux exercices programmés par les enseignants. Ce seront des outils de référence, donc de consultation, qui ne vont pas accabler l'élève destinataire avec un excès de matériel linguistique informatif qui dépasse ses fonctions premières d'appréhension de ces langues locales.

En dehors de la classe, les dictionnaires scolaires vont être des instruments de consultation autonome qui continueront à jouer un rôle de nature pédagogique lors des devoirs et des travaux scolaires faits à la maison. Ils deviendront aussi des ouvrages de consultation ponctuelle lors d'activités de jeu ou de lecture des langues locales. Comprendre la structure des dictionnaires, c'est-à-dire savoir reconnaître les diverses composantes d'un article et distinguer le genre textuel dictionnairique des autres genres textuels, sera un atout dans le processus de standardisation de ces langues locales. Cette seconde fonction hors de la classe reste soumise à la première si l'on considère qu'il revient aux enseignants de former les écoliers et les élèves au maniement et au fonctionnement du dictionnaire. Cette pédagogie vitale est de leur ressort.

Les dictionnaires scolaires vont permettre aux jeunes élèves gabonais, un enseignement/apprentissage soutenu des langues locales tout au long de leur période primaire et la période secondaire requise. Cela va aider ces langues à connaître un développement conséquent. Ce seront chez ces jeunes que demain on trouvera des lecteurs et des rédacteurs de la presse et de la littérature à côté des années d'enseignement en langue et culture françaises.

Conclusion

Les dictionnaires scolaires ont impact sur le processus de standardisation d'une langue dans le sens qu'ils aident dès le plus bas âge, c'est-à-dire chez les futurs lecteurs et rédacteurs, à vulgariser et à se familier avec le bon usage de cette langue. Ces ouvrages, ne sont aucunement dans un système d'opposition avec les dictionnaires généraux pour adultes mais plutôt dans un ordre de gradation et de hiérarchisation. En effet, les dictionnaires scolaires sont un état dans le cheminement de l'apprentissage de la langue qui va du bas vers le haut et non l'inverse.

Ces ouvrages doivent donc être nécessairement confectionnés au Gabon. La pratique et la production écrites actuelles des langues locales sont les conditions impératives de leur développement futur. Et cette pratique passera par la préparation des jeunes aujourd'hui afin qu'ils produisent demain une littérature écrite au sens large qui va renforcer le processus de standardisation de ces langues. Ce qui assurerait davantage le relais de l'oral et la diffusion de la culture écrite dans les langues locales au Gabon. Toutefois, les dictionnaires scolaires ne résultent pas d'un simple «copier-coller» des dictionnaires généraux pour adultes. Ils peuvent en résulter par une approche de «copier-coller et suppression des étymologies, des prononciations, des citations, des registres sous-familiers et littéraires ou très techniques».

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Notes

1. Quand bien même elles ne sont pas officialisées dans certains cas, ces langues endogènes remplissent des fonctions réelles dans la communication officielle. En effet, de nombreux agents parlent dans ces langues locales sur leurs lieux de travail pour remplir certaines tâches dans le cadre professionnel (cf. Kalonji 1993: 56).
2. Pour plus d'informations sur ces Constituants Théoriques, cf. Mavoungou (2010: 83-85).
3. Le Constituant Théorique B fait référence à la théorie de l'organisation. Il traite de tout ce qui est planification dictionnairique, notamment le plan de conceptualisation dictionnairique (macrostructure, microstructure, structure d'accès, médio structure, etc.) et le plan d'organisation du dictionnaire (mission du projet de dictionnaire, structures impliquées dans la préparation d'un dictionnaire ainsi qu'une estimation du temps et du capital nécessaire pour le projet).
4. Le Constituant Théorique C, se réfère à la théorie de la recherche lexicographique sur la langue et se compose de trois entités, à savoir la théorie de la collecte des données lexicographiques, la théorie de la saisie des données et la théorie informatique ou encore lexicographie assistée par ordinateur.
5. Cf. Gouws (2001: 113), Nyangone Assam (2002: 292-304) et Mavoungou (2011: 86).
6. Les dictionnaires pédagogiques, eux-mêmes faisant partie de l'une des quatre sous-catégories des dictionnaires monolingues, se divisent en deux sous-catégories, à savoir, les dictionnaires scolaires et les dictionnaires d'apprenant, cf. Gouws (2001: 113) et Boulanger (2001: 3).
7. Á ce propos, Boulanger (2006: 14) indique que techniquement, l'architecture du dictionnaire scolaire obéit à une structure à trois niveaux: la macrostructure, la microstructure et l'iconostructure, la dernière ayant autant d'importance pédagogique que les deux autres.
8. Pour plus d'informations sur la microstructure intégrée et non intégrée, cf. Gouws (2001: 127) et Mavoungou (2010: 65, 87, 178, 203-237, 268, 299, 311, 316). Ce dernier applique ces concepts dans le contexte des langues locales du Gabon.
9. Les renvois d'article externe guident l'utilisateur vers une entrée d'un autre article ou tout autre texte dans le dictionnaire. Avec les renvois d'article interne qui fonctionnent dans les limites d'un article, ils font partie de la médiostructure qui est le système des renvois qui conduit un utilisateur d'une position de référence à une adresse de référence, cf. Gouws (2001: 123).
10. Le co-texte se rapporte à l'environnement syntaxique de l'item lexical traité. Il ne désigne l'environnement d'un fait de langue qu'en tant que zone linguistique, éléments concrets de discours, sans préjuger de l'influence qu'il peut avoir sur le fait de langue en question. Tandis que le contexte donne l'environnement pragmatique de cet item lexical. Il désigne donc l'environnement du fait de langue en tant qu'il est susceptible d'avoir une influence sur ce fait de langue (influence phonétique, morphologique, sémantique, etc.), cf. Mavoungou (2010: 260).

Références Bibliographiques

Boulanger, J.-C. 2006. Du Côté de la Petite Histoire des Dictionnaires Scolaires Modernes. Candel, D. et F. Gaudin (Éds). 2006. *Aspects Diachroniques du Vocabulaire:* 209-240. Mont-Saint-Aignan (France): Publications des Universités de Rouen et du Havre.

Collectif. 2013. *Mon Premier Dictionnaire Bilingue Français–Fang*. Paris: Auzou Editions.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Emejulu, J.D. 2001. Lexicographie Multilingue et Multisectorielle au Gabon: Planification, Stratégie et Enjeux. Emejulu, J.D. (Éd.). 2001. *Éléments de Lexicographie Gabonaise*. *Tome I:* 38-57. New York: Jimacs-Hillman.

Emejulu, J.D. 2002. Défis et Promesses de la Lexicographie Intégrale dans les Pays de Développement. Emejulu, J.D. (Éd.). 2002. *Éléments de Lexicographie Gabonaise*. *Tome II:* 366-381. New York: Jimacs-Hillman.

Gallardo, **A.** 1980. Dictionaries and the Standardization Process. Zgusta, L. (Éd.). 1980. *Theory and Method in Lexicography:* 59-69. Columbia, SC: Hornbeam Press.

Galley, S. 1964. *Dictionnaire Français–Fang et Fang–Français*. Neuchâtel: Henri Messeiller.

Gouws, **R.H.** 2001. Formation lexicographique: Approches et thèmes. Emejulu, J.D. (Éd.). 2001. *Éléments de Lexicographie Gabonaise. Tome I:* 95-134. New York. Jimacs-Hillman.

Kalonji, M.T.Z. 1993. *La Lexicographie Bilingue en Afrique Francophone: L'Exemple Français–Cilubà.* Paris: L'Harmattan.

Kavanagh, **K.** 2002. Pour une Adaptation d'un Dictionnaire Monolingue pour un Usage Local. Emejulu, J.D (Éd.). 2002. *Eléments de lexicographie gabonaise*. *Tome II:* 264-290. New York: Jimacs-Hillman.

Lombard, F.J. 1990. *'n Metaleksikografiese fundering van Afrikaanse skoolwoordeboeke.* Thèse Doctorat. Stellenbosch: Université de Stellenbosch.

Mabika Mbokou, **L.** 2006a. Le Français, Langue Maternelle à Libreville! *CENAREST Infos* 4 Octobre–décembre: 4.

Mabika Mbokou, L. 2006b. *The Dictionnaire Francais–Yipounou / Yipounou–Français* of the Eglise Evangélique du Sud-Gabon (CMA) 1966. *Lexikos* 16: 104-120.

Mavoungou, P.A. 2001. La Mondialisation et la Lexicographie Trilingue ou Plurilingue au Gabon. Emejulu, J.D. (Éd.). 2001. *Éléments de Lexicographie Gabonaise*. *Tome I:* 160-183. New York: Jimacs-Hillman.

Mavoungou, P.A. 2010. *Lexicographie et Confection des Dictionnaires au Gabon.* Stellenbosch: SUN PReSS.

Mavoungou, P.A. 2011. Macro- and Microstructural Issues in Mazuna Lexicography. *Lexikos* 11: 122-138.

Mével, J.P. 2011. *Dictionnaire Hachette Benjamin: CP-CE 5–8 ans.* Paris: Hachette Education.

Mihindou, G.R. 2001. Apports des Missionnaires à la Lexicographie Gabonaise: Dictionnaires Bilingues Fang–Français/Français–Fang; Français–Yipounou/Yipounou–Français; Français/Mpongwé. Emejulu, J.D. (Éd.). 2001. *Éléments de Lexicographie Gabonaise*. *Tome I:* 7-37. New York: Jimacs-Hillman.

Ndinga-Koumba-Binza, H.S. 2006. *Lexique Pove–Français/Français–Pove,* Mickala Manfoumbi: Seconde Note de Lecture. *Lexikos* 16: 293-308.

Nyangone Assam, B. 2001. La Lexicographie dans l'Apprentissage des Langues Gabonaises. Emejulu, J.D. (Éd.). *Éléments de Lexicographie Gabonaise*. *Tome I:* 187-205. New York: Jimacs-Hillman.

Nyangone Assam, B. 2002. Le Dictionnaire Pédagogique Scolaire: Quelques Aspects du Traitement Lexicographique. Emejulu, J.D. (Éd.). 2002. *Éléments de Lexicographie Gabonaise*. *Tome II:* 292-304. New York: Jimacs-Hillman.

Ondo-Mebiame, P. 2005. De la Révision du Dictionnaire de Samuel Galley. *Lexikos* 15: 151-163.

Ondo-Mebiame, P. et E.M. Ella. 2011. Contribution en vue de l'Amélioration du Dictionnaire des Gabonismes d'Eric Dodo Bounguendza. *Humanités Gabonaises* 1: 35-57.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2030 (Article)

Tomba Moussavou, F. 2002. Quelques Aspects de la Lexicographie Monolingue au Gabon. Emejulu, J.D. (Éd.). 2002. *Éléments de Lexicographie Gabonaise*. *Tome II:* 159-173. New York: Jimacs-Hillman.

Unik Afrika. 2013. *Le Dictionnaire des Mwanas «avec Papito et Jolica»*. Libreville: Les Editions Unik Afrika.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Éd.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983:* 13-30. Tübingen: Max Niemeyer.

Wiegand, H.E. 1989. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J et al. (Éds.). 1989. *Wörterbücher*. *Dictionaries. Dictionnaires. Ein Internationales Handbuch zur Lexikographie*. Handbuch zur Sprach-und Kommunikationswissenschaft. Vol. 5.1: 246-280. Berlin/New York: De Gruyter.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

The *Serbian–Russian and Russian–Serbian Idiomatic Dictionary for Expressing Emotions*: Theoretical and Practical Aspects

Milena Mrkaić, *Department of Russian Language and Literature, Faculty of Philosophy, University of Montenegro, Nikšić, Montenegro*
(milena_mrkaic@yahoo.com)
(https://orcid.org/0009-0009-4099-3346)

Abstract: This article presents the development of the *Serbian–Russian and Russian–Serbian Dictionary of Emotion-Expressing Idioms* (SRIDE). The project aims to systematise idioms that convey emotional expression in Serbian and Russian, taking into account their cultural, linguistic, and cognitive aspects. The author designs a specialised dictionary based on data from the Serbian web corpus (srWaC) and the Russian National Corpus (RNC), compiled at the Department of Russian Language and Literature at the University of Montenegro. Material collection has been ongoing since 2020, with the phraseological database continuously expanding. The methodological foundation of the study is a comprehensive analysis of existing monolingual and bilingual dictionaries, which has identified key challenges, such as idiom polysemy and equivalence issues. The project employs a thematic classification of material based on twelve basic emotions defined by Izard's model. Each dictionary entry is meticulously structured, including a thematic heading, adaptive index, transcription, grammatical information, definition, usage examples, etymological notes, equivalents, and antonyms. The implementation of the OnLex electronic platform enhances functionality, offering multimedia storage and automated search capabilities. The findings demonstrate that integrating traditional lexicographic methods with modern technologies enables the creation of a reliable tool for cross-cultural analysis, language teaching, and translation, catering to the needs of linguists, translators, educators, and students.

Keywords: EQUIVALENCE, EMOTIONS, BILINGUAL DICTIONARIES, SERBIAN–RUSSIAN IDIOMS, CORPUS ANALYSIS, COMPARATIVE LEXICOGRAPHY, MULTI-ASPECT CLASSIFICATION

Opsomming: Die *Serbian–Russian and Russian–Serbian Idiomatic Dictionary for Expressing Emotions*: Teoretiese en praktiese aspekte. In hierdie artikel word die ontwikkeling van die *Serbian–Russian and Russian–Serbian Idiomatic Dictionary for Expressing Emotions* bespreek. Die projek het die sistematisering van idiome wat emosionele uitdrukking in Serwies en Rus-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

sies weergee, met inagneming van hul kulturele, linguistiese en kognitiewe eienskappe, ten doel. Die outeur ontwerp 'n gespesialiseerde woordeboek gebaseer op data van die Serwiese webkorpus (srWaC) en die Russiese Nasionale Korpus (RNC) wat by die Departement Russiese Taal en Letterkunde aan die Universiteit van Montenegro saamgestel word. Materiaal word sedert 2020 versamel en die fraseologiese databasis word voortdurend uitgebrei. Die metodologiese grondslag van die studie is 'n omvattende ontleding van bestaande eentalige en tweetalige woordeboeke, wat sleuteluitdagings soos idioompolisemie en ekwivalensiekwessies geïdentifiseer het. 'n Tematiese klassifikasie van materiaal gebaseer op twaalf basiese emosies wat deur Izard se model gedefinieer word, word vir die projek gebruik. Elke woordeboekinskrywing is noukeurig gestruktureer en sluit 'n tematiese opskrif, aanpasbare indeks, transkripsie, grammatikale inligting, definisie, gebruiksvoorbeelde, etimologiese aantekeninge, ekwivalente en antonieme in. Die implementering van die elektroniese platform OnLex verbeter funksionaliteit en bied multimediabergings- en outomatiese soekmoontlikhede. Die bevindings toon dat die integrasie van tradisionele leksikografiese metodes en moderne tegnologie die skep van 'n betroubare hulpmiddel vir kruiskulturele analise, taalonderrig en vertaling, wat voorsiening maak vir die behoeftes van taalkundiges, vertalers, opvoeders en studente, moontlik maak.

Sleutelwoorde: EKWIVALENSIE, EMOSIES, TWEETALIGE WOORDEBOEKE, SERWIES–RUSSIESE IDIOME, KORPUSANALISE, VERGELYKENDE LEKSIKOGRAFIE, MULTI-EIENSKAP-KLASSIFIKASIE

1. Introduction

In the context of global cultural integration and the continuous evolution of idiomatic expressions, the lexicographic task of creating a specialised bilingual dictionary gains particular significance. The present project — *Serbian–Russian and Russian–Serbian Idiomatic Dictionary for Expressing Emotions* (hereinafter SRIDE) — aims to systematise idioms in both languages to capture their emotional and cultural connotations. Lexicography, as the science of dictionary-making, not only documents fixed linguistic expressions but also reflects the mentality, cultural traditions, and historical dynamics of nations. In society, dictionaries serve as authoritative guides for users, functioning as a kind of social institution that regulates word usage and, more broadly, the life of a language within a certain normative framework. This perspective is particularly pertinent to the SRIDE lexicographic project, the relevance of which is determined by several key factors:

1. *Growing communication between Montenegro and Russia*. The increasing interaction between Montenegro and Russia has heightened the need for exchanging expressive phrases and idioms that convey nuances of emotional perception in both cultures. In today's information-driven society, emotional speech plays a crucial role in interpersonal communication, affecting both its effectiveness and intercultural understanding. Idioms

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

expressing feelings and emotions not only serve as a means of expressiveness but also reflect deep-rooted cultural and historical features of a nation's linguistic consciousness. Such a dictionary will be a valuable resource for linguists, translators, educators, students, and anyone interested in the richness of Serbian and Russian phraseology. There is a pressing need for a new dictionary, as no specialised collections of emotion-expressing idioms currently exist. Moreover, both Serbian and Russian are undergoing constant renewal of their idiomatic inventories: some expressions lose relevance, others acquire new emotional connotations, and entirely new phraseological units emerge due to shifts in culture and social life. The dynamic nature of modern culture necessitates an updated lexical resource capable of documenting both established and emerging expressions.

2. *Declining proficiency in phraseological resources*. Over the past two decades, command of phraseological units in Serbian and Russian has significantly declined. Annual testing of first- and second-year philology students at the University of Montenegro — assessing their knowledge of Serbian and Russian (as a foreign language) idiomatic expressions — reveals a stark contrast with previous generations. Tasks requiring the reconstruction of idioms based on their initial or final fragments, which students easily completed two decades ago, now prove challenging for most. Notably, when these same students conduct similar tests with their older relatives, they find that the older generation faces no difficulties in restoring the original form of fixed expressions. This trend underscores the importance of explanatory phraseological dictionaries in contemporary lexicography. Such dictionaries not only record stable expressions but also provide commentary on their usage, ensuring their preservation and accurate transmission to future generations. The SRIDE project thus responds to both linguistic and sociocultural demands, offering a comprehensive tool for cross-linguistic and cross-cultural analysis while addressing the challenges posed by the evolving nature of idiomatic language.

3. *Practical experience and target audience*. Experience from seminars on Serbian–Russian phraseology has shown that students exhibit strong interest in tasks involving the identification of idioms in their native language that are semantically close to Russian equivalents. Such exercises allow learners to compare the imagery and situational contexts underlying idiomatic expressions across languages. For example: *Ljut kao pas* (Serbian) — *Злой как собака* (Russian) — angry as a bear (English) — meaning extremely angry and easily irritated. A specialised bilingual dictionary of idiomatic expressions will thus serve as an effective tool for both students and educators.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

This report presents the SRIDE — a specialised bilingual project under development at the Department of Russian Language and Literature, the University of Montenegro. Work on its database, initiated in 2020, has so far compiled over 500 idioms. The creation of the SRIDE represents a significant scholarly and practical endeavour, contributing to the in-depth study of idiomatic inventories in both languages, as well as the advancement of intercultural communication and lexicography as a whole. The project is designed to serve as a reliable resource for linguists, translators, and educators, while laying a solid foundation for future research in phraseology and cross-linguistic analysis.

2. Theoretical and practical background

The development of the SRIDE builds upon prior research, existing lexicographic traditions, and established methods of lexical description. Lexicographic works form a cohesive academic framework in which new projects expand upon and refine earlier data, ensuring continuity and comprehensiveness. As noted by Dubichinskiy (2009: 50): "It is not uncommon for a dictionary that initially fails to meet its creators' expectations to later prove valuable as a source for compiling other dictionaries". This project draws on: (1) corpus data — Serbian media texts, the Serbian web corpus (srWaC 2025), and the Russian National Corpus (RNC 2025); and (2) phraseological lexicographic sources — existing monolingual and bilingual dictionaries.

A key terminological clarification is required here. The operational unit of this study is the idiom. In line with English linguistic tradition, the terms *phraseological unit (PU)* and *idiom* are treated as synonymous: "PU as an opaque, invariable unit related to the word and its syntactic function in the Russian school, often referred to as 'idiom' in Anglo-Saxon linguistics" (Veisbergs 2013: 110).

The monolingual and bilingual lexicographic works which were considered and analysed when compiling the SRIDE card index include four Serbian and five Russian publications. The Serbian works examined are as follows:

1. *Phraseological Dictionary of the Croatian or Serbian Language* (Matešić 1982), which is intended for linguists, educators, translators, and anyone interested in the phraseology of these languages. The dictionary employs an alphabetical classification of idioms, with partial thematic grouping, for example, expressions related to emotions are structured but not always systematically. The material includes expressions from literary works, oral speech, and journalism, such as: *Srce u petama* — to be terrified, experience intense fear; and *Pući od muke* — to feel extreme irritation or anger. Some examples are supported by classical literary sources: *Skočiti od radosti* — to experience sudden, intense joy; and *Biti na sedmom nebu* — to be overjoyed (lit. to be on the seventh heaven). A notable strength of this dictionary is its regional differentiation between Serbian and Croatian. However, it lacks clear label-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

ling of emotionally charged idioms, and some usage examples are presented without context.

2. *Phraseological Verb–Nominal Syntagms in Contemporary Serbo-Croatian Language* (Mršević-Radović 1987), which focuses on stable idiomatic constructions where the primary element is a verb–noun combination. The classification is based on: (1) syntactic structure — identifying fixed verb–noun syntagms; (2) semantic load — thematic grouping, with significant attention to emotionally expressive phrases; and (3) stylistic and pragmatic features — distinguishing between colloquial, literary, and archaic forms. The material was collected through corpus analysis and native speaker interviews, reflecting real usage patterns in Serbian: *Izbečiti (iskolačiti) oči* — to goggle one's eyes (i.e., to be astonished); *Bojati se vlastite sjene* — to be afraid of one's own shadow (i.e., to be extremely fearful); and *Plakati kao kiša* — to cry like rain (i.e., to be deeply upset). The dictionary includes etymological notes, though they serve only as a general overview — a common feature in specialised phraseological works.

3. *Idioms in the Serbian Language* (Milenković 2006), a reference guide focused on Serbian idioms. Unlike traditional academic works, this is a practical manual for Serbian learners and those seeking to enrich their speech. It emphasises colloquial usage, featuring numerous informal expressions: *Zračiti srećom* — to radiate happiness (i.e., to be very joyful); *Slediti se od straha* — to freeze in fear; and *Grč u stomaku* — to have a knot in one's stomach (i.e., to feel intense anxiety). The material is drawn from literature and media. Idioms are categorised by syntactic, morphological, semantic, expressive, and usage-based criteria. However, the dictionary's limited scope precludes in-depth semantic analysis or extensive examples for each category, with no more than two pages devoted to any classification. Consequently, despite its varied systematisation principles, the content remains largely introductory and lacks systematic rigor — a deliberate choice, as the dictionary prioritises practical spoken-language application over specialised emotional idiom categorisation.

4. *Phraseological Dictionary of the Serbian Language* (Otašević 2012) comprises 25 000 entries, covering both common and specialised idiomatic units. Each entry provides a concise definition, emphasising stylistic nuances and contextual usage. Though academically oriented, the dictionary remains accessible to a broad audience due to its detailed explanations — a distinguishing feature compared to more specialised publications.

This analysis underscores the methodological foundations of the SRIDE, which synthesises these works' strengths while addressing their limitations through

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

systematic emotional categorisation and corpus-based validation. The compiled data are presented in Table 1.

Table 1: Serbian phraseological dictionaries

Criteria	Matešić (1982)	Mršević-Radović (1987)	Milenković (2006)	Otašević (2012)
Principle of Organisation	Alphabetical order	Thematic (verb–noun phrases)	Thematic-alphabetical order	Alphabetical order
Material	Literary examples	Contextualised examples	Mainly conversational examples	Literary, colloquial examples
Etymology	Extensive etymological analysis	Limited etymological data	—	Partial etymological references
Target Audience	Linguists, teachers, philologists, researchers	Philologists, syntax experts	Students learning Serbian, non-native speakers	Linguists, translators, advanced learners
Lexicographic Approach	Descriptive, with detailed definitions and examples	Analytical, focusing on structural features	Practical, emphasising use in daily communication	Descriptive and comparative
Additional Comments	Most detailed in terms of etymology	Highly specialised, focuses on syntactic structures	Focuses on spoken language, adapted for practical use	Most extensive, wide coverage of styles and contexts

As for the relevant Russian monolingual dictionaries, the following were investigated:

1. *Phraseological Units in Russian Speech. Dictionary-Reference Book* (Basko 2002), which focuses on contemporary usage while also incorporating expressions with historical background. The dictionary covers colloquial, literary, professional, vernacular, and archaic idioms. It also documents new phraseological units that have gained prominence in modern media language. The author employs a descriptive method for presenting idiom meanings, supplemented with explanatory notes. Synonymous series are frequently used, and possible expression variations are indicated. Unlike academic dic-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

tionaries, this work contains more authentic examples from spoken language.

2. *Phraseological Dictionary: Cultural and Cognitive Space of Russian Idiomatic Expressions* (Alefirenko and Zolotyh 2008) represents a systematic attempt to describe Russian phraseology with emphasis on reflecting cultural-cognitive space. The dictionary selects culturally marked phraseological units including: (a) idioms reflecting elements of classical culture (e.g., *Аркадская идилия* — *Arcadian Idyll* — a blissful life; and *Медеевская ярость* — Medea's fury — an intense feeling of anger and vindictiveness, associated with the legendary Medea); and (b) idioms containing references to specifically Russian cultural-historical realities (e.g., *Вот так клюква!* — What a cranberry! — an expression of surprise at an unpleasant unexpected situation; and *Сесть в калошу* — to be in a tight spot — to experience a feeling of shame and embarrassment); (c) idioms borrowed from other languages (e.g., *Иметь зуб* — to hold a grudge — a feeling of hostility, or a desire for revenge against someone); and (d) idioms of biblical origin (e.g., *Испытывать благодать* — to experience grace — to feel a sense of spiritual inspiration, bliss, or divine favour, often in a religious context). The dictionary also includes phraseological units based on Russian fairy tales (e.g., *Как за каменной стеной* — like behind a stone wall — to feel safe, confidence in protection) and anecdotes (e.g., *Как слону дробина* — like a grain of sand to an elephant — feeling of calmness or indifference, regardless of the circumstances). Dictionary entries follow established Russian phraseographic principles while incorporating distinctive features, providing comprehensive information about phraseological synonyms and antonyms, grammatical characteristics, syntactic functions, cognitive-cultural information, etymological sources and conceptual frameworks.

3. *Dictionary of Contemporary Russian Phraseology* (Zhukova and Zhukov 2023), which is based entirely on modern newspaper material. It presents not only semantic, stylistic and grammatical characteristics of phraseological units, but also includes historical-linguistic, cultural and etymological commentary when necessary. A significant aspect is the information about paradigmatic, syntagmatic and derivational features of phraseological units. The dictionary's uniqueness lies in its use of illustrative material from prominent Russian newspapers (2006–2018). Notably, it includes phraseological units absent from other dictionaries, incorporating youth slang and colloquialisms (e.g., *Вогнать в краску* — make someone blush — to embarrass or cause someone to turn red from shame or shyness; and *Крыша поехала* — lost one's marbles — to lose one's mind, behave in an irrational or erratic manner, or show signs of insanity). This linguistic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

snapshot demonstrates just one facet of phraseology reflected in modern newspaper discourse.

4. *Dictionary of Russian Idioms* (Subbotina 2024) stands out for its exceptional level of detail and the relevance of its material. This work combines alphabetical organisation with a thematic index, facilitating efficient search and analysis of idioms — particularly those expressing emotional states. The dictionary's value for linguists, educators, translators, and general users is enhanced by the following: abundant usage examples, comprehensive cultural and etymological commentary, and detailed stylistic markers. The dictionary draws on an extensive corpus of contemporary texts including literary works, journalism, media documents, and colloquial speech examples (e.g., *Дрожать как осиновый лист* — to tremble like a poplar leaf — denoting intense agitation or fear).

5. *The Large Phraseological Dictionary of the Russian Language* (Telija 2024) presents an exhaustive collection of citations from all written genres, including internet idioms. Developed in collaboration with the Institute of Linguistics (Russian Academy of Sciences), its selection criteria include: (a) 16 thematic or ideographic fields; (b) archetypal cultural layers; and (c) diversity of cultural codes in phraseological units. The macrostructure comprises three dialogically complementary sections: "meaning", "usage" and "cultural commentary" (Bayramova 2010: 28-29). Telija identifies four groundbreaking features in the preface: (1) first implementation of a "compressed" situational interpretation zone; (2) aspectual "nests" or idiom paradigms (containing up to eleven components); (3) first Russian phraseographic work to incorporate cultural commentary; and (4) systematic revelation of cultural meanings through analysis of idioms as elements of cultural semiotics.

Thus, these idiom dictionaries represent a new generation of lexicographic works that supplement traditional information with culturological data tied to cultural codes and cognitive characteristics. These dictionaries document emerging idioms with current media illustrations — the primary contemporary sources (see Table 2).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

Table 2: Key features of Russian monolingual phraseological dictionaries

Criteria	Basko (2002)	Alefirenko and Zolotyh (2008)	Zhukova and Zhukov (2023)	Subbotina (2024)	Telija (2024)
Principle of organisation	Alphabetical order with thematic sections	Thematic classification (cultural and historical aspects)	Alphabetical, with a thematic index	Alphabetical with additional semantic classification	Alphabetical with conceptual fields and ideographic categories
Material	Contemporary spoken and written Russian, including informal speech	Historical, literary, and cultural idioms, selected based on cognitive criteria	Examples from modern Russian media (2006–2018)	Extensive corpus-based research including literature, media	Comprehensive literary, journalistic, and digital sources
Etymology	Limited etymological references	Strong emphasis on cultural and historical origins	Includes historical and linguistic commentary when relevant	Provides cultural and etymological commentary	In-depth etymological analysis linked to cultural semantics
Target audience	Students, linguists	Researchers, cultural linguists	Journalists, media specialists	Linguists, translators, educators	Academics, linguists, students
Lexicographic approach	Descriptive, with synonymic series and variation tracking	Cognitive and cultural analysis of idioms	Usage-focused, tracing phraseological trends	Comprehensive, detailed usage examples	Semantic and situational clustering of idioms
Additional comments	Integrate informal, colloquial, and slang idioms with descriptive explanations	A unique attempt to categorise idioms according to cultural-cognitive principles	Incorporates mass-media phraseology	Focus on idioms expressing emotional and psychological states, including corpus analysis	Introduces an ideographic approach to phraseology and establishes phraseological "conceptual nests"

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

The next logical stage of the study is the examination of bilingual dictionaries. The study of Serbian–Russian and Russian–Serbian lexicography shows that the tradition of compiling bilingual idiomatic dictionaries has a relatively recent history. Unlike general lexical dictionaries, the development of which can be traced back to the 19th century, phraseographic work in this direction began only in the second half of the 20th century. The formation of dictionaries of this type comprises two editions, each of which made its own contribution to the development of lexicography:

1. *Russian–Croatian or Serbian Phraseological Dictionary*, a systematic collection of fixed expressions emphasising their semantic load and context of use (Menac et al. 1979). This edition focuses primarily on idioms used in the Serbo–Croatian linguistic space, with parallel presentation of equivalents in Russian. Some dictionary entries could have been more detailed: in certain cases, a literal translation of Russian idioms is provided without literary illustrations or stylistic characteristics. The authors adopted a broad approach to phraseology, including only fixed expressions as well as terms and collocations in the process of phraseologisation. At the same time, expressivity is not considered as a key feature of idioms, so the dictionary presents both traditional proverbs and sayings (e.g., *Daleko od očiju, daleko od srca* — out of sight, out of mind) and word combinations that do not have a pronounced imagery, such as *Razrovani put* — *Разбитая дорога* — rough road.

2. *Serbo–Croatian–Russian Phraseological Dictionary*, which is characterised by updated methodology and current data, allowing for a better reflection of the modern linguistic landscape (Trofimkina 2005). It features Serbian phraseological units (in the narrow sense of the term), accompanied by Russian-language semantic descriptions, syntactic characteristics, and phraseological equivalents in Russian. A particularly valuable aspect of the work is the commentary on culture-specific words, elucidating their cultural and linguocultural context. For example, the expression *Tražiti / Htjeti preko hljeba pogače* (to be dissatisfied with what is available; to demand more than necessary) is accompanied by an explanation of the word *pogače* — a traditional wheat bread in Serbian cuisine. The dictionary includes an extensive corpus of quotations from literary, popular science, memoir, and journalistic texts, significantly expanding its utility in research and translation practice.

Despite the modest lexicographic practice in the field of Serbian–Russian idiomatics, it can be noted that this direction is developing promisingly. A number of works continue to emerge: Serbian–Russian idioms of different thematic groups are compared (Dzonic 2016), theoretical aspects are developed (Štrbac 2018), work is underway on compiling a new large bilingual dictionary (Dragičević et

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

al. 2024), and problems in the selection of clauses for the Serbian–Russian dictionary are analysed (Dragičević 2024). The analysis of these works indicates a growing interest in the study of Serbian–Russian idiomatic connections. It is, however, noteworthy that works dedicated to describing idioms expressing emotions are absent in contemporary lexicography. This category of phraseology presents significant challenges for learners of Russian as a foreign language, as it involves complex metaphors, cultural-historical realities, and lexico-grammatical structures that differ from Serbian.

3. Principles and methodology

The development of the SRIDE database employed comprehensive methodological approaches combining lexicographic analysis, corpus linguistics, cultural linguistics commentary, and cognitive research (Knežević et al. 2021). According to Fraser (2008: 2), the organisational principles of bilingual dictionaries depend fundamentally on three kinds of information: "form, meaning and context. Form includes spelling, accentuation and quantities, and inflections. Meaning includes translations: single-word equivalents of the lemma in the exit language. Context covers both the real-world and the linguistic environments." The core principles governing material selection and dictionary organisation include:

— *Frequency and stability.* Phraseological units were selected based on their regular occurrence in literature, media, and spoken discourse. National corpus data (the srWaC for Serbian, and the RNC for Russian) helped identify stable expressions with high representativeness in both languages. Selection criteria prioritised quantitative analysis of phraseological frequency across genres and functional significance in daily communication. The primary criterion was current usage by native Serbian speakers and/or presence in the srWaC, complemented by verification in the RNC.

— *Normative compliance*. Only expressions conforming to contemporary literary standards were included. This principle ensures reliability and practical applicability by excluding taboo or non-normative expressions and regional or dialectal phraseologisms. Exceptions were made for widely adopted non-standard expressions, which were marked with appropriate stylistic labels (e.g., *Izbečiti, iskolačiti oči* — *Шары на лоб полезли* — eyes popped out of one's head — marked as colloquial). As a descriptive dictionary, the SRIDE documents actual usage while maintaining language cultivation standards — particularly crucial for learners mastering literary norms.

— *Linguocultural relevance.* Units must reflect culture-specific emotional realities. Selected idioms embody characteristic imagery, symbolism, and ste-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

reotypes embedded in Serbian and Russian linguistic cultures. This principle facilitates not only accurate translation but also deeper intercultural understanding of emotional expression.

— *Contextual grounding.* Each unit is evaluated through authentic textual examples to ensure accurate representation of usage patterns. Contextual analysis determines the following: register (formal or informal), domain (literature, conversation), and emotional-expressive load. This approach helps users grasp appropriate situational usage.

— *Equivalence principle.* For each Serbian unit, the dictionary provides: (1) direct Russian equivalents where available; (2) descriptive translations for non-equivalent items; and (3) culturally analogous constructions with explanatory notes. This systematic comparison reveals both convergences and divergences in emotional expression across cultures.

— *Standardised presentation.* All entries follow a uniform structural template ensuring consistency in orthography, grammatical annotation, definition formatting, and example presentation.

— *Functional-pragmatic relevance.* The dictionary covers idioms expressing emotions fundamental to modern communication: (a) basic emotions (joy, surprise, fear, anger, sadness); (b) complex affective states (nostalgia, ambivalence); and (c) social emotions (shame, pride). The selection prioritises units with high communicative utility in contemporary contexts.

These methodological principles collectively ensure the SRIDE's scholarly rigor, cultural precision, and practical utility for language learners, translators, cross-cultural researchers and bilingual speakers. The approach bridges theoretical lexicography with applied language pedagogy, addressing a critical gap in Slavic phraseological resources while establishing replicable standards for emotion-focused dictionary compilation. Subsequent sections will demonstrate how these principles manifest in the dictionary's macrostructure and individual entry design.

4. The dictionary

4.1 Macrostructure

The dictionary's macrostructure is currently being developed in Microsoft Word as a manuscript intended for subsequent print publication. Simultaneously, completed materials are being transferred to the OnLex electronic lexicographic plat-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

form (Makarov et al. 2022), which significantly enhances the SRIDE's functionality through several key advantages:

— *Space optimisation:* Unlike print editions, the digital format eliminates spatial constraints, allowing for more extensive illustrative examples, richer contextual information, and comprehensive usage explanations.

— *Advanced search capabilities*. The platform's modern design supports full paradigm representations (enabling searches by any word form), multimedia integration (particularly audio recordings for stress patterns), and multi-criteria searching (including tags, equivalents, and usage examples).

— *Editorial efficiency*. The platform provides online article creation tools, automated format conversion, unified presentation standards, collaborative work environment, etc.

The dictionary features a dual-component structure with complementary sections. The first part — the Serbian–Russian Dictionary — comprises a selection of Serbian idioms, grouped by emotional categories, with detailed descriptions, culturological commentaries, and translations into Russian.

The second part contains a corresponding collection of Russian idioms presented according to the same emotional categories, with comprehensive analyses and translations into Serbian. The bilingual dictionary, constructed on the principle of two equal parts, offers several advantages over a monolingual version:

— *Symmetry of information and ease of search* makes it easy to find similar expressions in both languages, since each section is organised according to one thematic scheme (emotions).

— *Dual lexicographic referentiality* enables contrastive analysis of emotional expression, reveals cultural-linguistic parallels and divergences, and supports bidirectional translation workflows. The inclusion of both sections allows users to see the translation from one language to another and to compare how different idioms convey emotional connotations in both language cultures.

— *Contrastive analysis* documents equivalence types (full, partial, or lacunary), highlights culture-specific emotional conceptualisation, and provides research data for cultural linguistics.

— *Pedagogical utility* serves both Serbian and Russian native speakers, reduces reference source switching, and enhances learning efficiency through systematic presentation.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

While the design principle maintains structural parity, practical implementation reveals quantitative asymmetries across languages, particularly in emotion domains with differential phraseological development. According to the plan, both parts of the dictionary should be structured equally in terms of the volume and quality of the presented material, which ensures a symmetrical reflection of the phraseological system of both languages. However, in practice there are cases when the number of idioms in one part exceeds their number in the other. This circumstance is determined by the peculiarities of the linguistic space and the degree of development of the corresponding phraseology in one of the languages. For example, Serbian idioms reflecting joy or emotional intensity are characterised by great diversity, whereas Russian contains more idioms for melancholy or sadness (reflecting cultural-historical factors). To address such imbalances, the project employs descriptive translation strategies, functional equivalents, explicit lacuna marking, and culturally contextualised commentary. In this regard, the dictionary places higher demands on the classification of polysemous expressions and the selection of precise equivalents, which remains its main challenge. Current development focuses on maintaining methodological consistency while accommodating natural language asymmetries, ensuring the dictionary remains both a reliable reference tool and a valid research instrument for comparative phraseology.

4.2 Material organisation

We employ a combined approach integrating alphabetic-thematic classification principles. This dual methodology incorporates: (1) formal-alphabetic organisation based on graphic sign systems, that follows standard lexicographic ordering, enables quick reference lookup, and maintains traditional dictionary accessibility; and (2) semantic-thematic organisation grounded in conceptual relationships between lexical units, which utilises categorisation and grouping techniques, and follows Faber et al.'s (2014) principle of organising lexical units through systematic conceptual frameworks.

The organisational structure is fundamentally based on Izard's (2009) differential emotions theory, which identifies twelve primary human emotions. These are arranged in alphabetical order for ease of search and comparison: *bijes* (anger), *gađenje* (disgust), *interes* (interest), *iznenađenje* (surprise), *krivica* (guilt), *ljubav* (love), *prezir* (contempt), *sramota* (shame), *stidljivost* (shyness), *strah* (fear), *tuga* (sadness), and *uživanje* (enjoyment). Idioms are systematically linked to these lexical hubs, and within each emotion category, idioms are arranged alphabetically. The model of primary emotions is illustrated in Figure 1.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

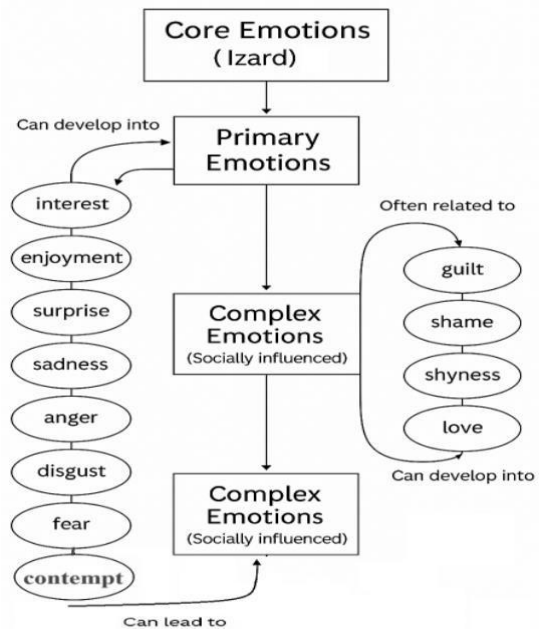


Figure 1: Primary human emotions in Izard's model

In other words, all idioms are grouped alphabetically by the main emotions, each of which represents a separate dictionary entry or section (e.g., *Enjoyment*, *Surprise*, *Sadness*, and *Contempt*). A sample of phraseological grouping based on the central emotion *Enjoyment* is presented in Table 3.

Table 3: Examples of Serbian and Russian equivalents

Serbian idiom	Russian equivalents	Literal translation and meaning
Biti lud (pijan) od sreće	*Без ума от счастья. Быть опьяненным радостью.*	To be crazy (drunk) from happiness — to be extremely happy.
Biti srećne ruke	*Счастливая рука. Легкая рука. Коснуться Мидаса.*	To have a lucky hand — to be fortunate, to have good luck. The expression comes from gambling and card games.
Čista sreća	*Чистое везение. Чистая удача. Везет как никогда. Полоса везения. Везет на все сто. Счастье на голову свалиться.*	Pure happiness or sheer luck — absolute happiness or luck. The word ***sreća*** in Serbian denotes both luck and happiness. Its origins can be traced back to ProtoSlavic roots, where it signified fortune and well-being.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

Osmjeh od uha do uha	*Улыбаться от уха до уха.* *Улыбка до ушей.* *Улыбаться во весь рот.* *Улыбаться во все щеки.* *Расплываться в улыбке.* *Сиять от счастья.*	A smile from ear to ear — a big, wide smile indicating great joy; to shine with happiness. Although the specific wording may vary from language to language, the concept is universal: when someone is happy, their smile becomes wide.
Roditi se u košuljici	*Родиться в рубашке.* *Баловень судьбы.* *Родиться с золотой ложкой во рту.* *Родиться под счастливой звездой.*	To be born in a little shirt — to be born under a lucky star. The idiom originates from an ancient folk belief that a baby born with the amniotic sac was considered fortunate.
Srce je zaigralo od sreće	*Сердце трепещет от счастья.* *Сердце радуется.*	The heart started dancing from happiness — to feel overwhelming joy.

This study critically evaluates the interpretation of polysemantic linguistic units, where the same idiom can have multiple meanings depending on the context, and Figure 1 confirms the cluster model of emotions. The polysemy of emotion-expressing idioms poses a challenge for their unambiguous classification, as many fixed expressions can simultaneously convey a complex range of feelings. For example, the expression *Srce je sišlo u pete* — *Сердце ушло в пятки* — one's heart dropped to one's stomach — may denote fear, anxiety, or surprise. It should also be noted that within the "Anger" cluster, emotions such as dissatisfaction, irritation, indignation, resentment, anger, and rage manifest simultaneously. This complicates the selection of the primary emotional category for inclusion in an alphabetically and thematically organised dictionary. To address the issue of idiom polysemy, the following solutions are proposed:

— *Determining the dominant emotion.* By analysing the frequency and intensity of emotional nuances in a text corpus, it is possible to identify the dominant emotion for a given idiom. This allows for the selection of a primary category, while secondary ones can be retained as supplementary information.

— *Using a multifaceted classification.* A single idiom may be included in multiple thematic entries, with all its emotional nuances indicated. The primary (dominant) nuance is marked with a special label, while secondary ones are reflected through synonyms or additional notes.

— *Cross-referencing.* Each dictionary entry may include cross-references to other categories. For instance, if an expression conveys both joy and sadness, it can be listed under the entry "Enjoyment" with a note "see also Sadness," and vice versa. If similar expressions exist in both sections, cross-references

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

should also be created, enabling users to quickly navigate between related idioms and consider comparative analysis.

— *Employing an adaptive index.* Incorporating a general index section for both parts of the dictionary allows for keyword-based searches. An adaptive index, serving as a consolidated reference, compiles keywords, themes, or categories, enabling users to locate relevant expressions regardless of their placement in the dictionary. This navigational tool offers the following advantages: (a) coverage of the entire dictionary rather than individual entries; (b) inclusion of both idioms and key elements (e.g., searching for the word "fear" in the index displays all related expressions); (c) grouping of idioms with similar meanings, even if they lack direct equivalents.

— *Aiding contextual analysis.* When compiling the dictionary, it is essential to provide usage examples that help users understand the contexts in which a particular emotion predominates.

Thus, the issue of idiom polysemy is mitigated through multifaceted classification with the identification of a dominant emotional nuance, the use of cross-references, and the provision of contextual examples. Even if an idiom conveys a complex range of emotions, users will be able to discern the context and emotional emphasis of its usage. An example of an idiom expressing multiple emotions: *Steže mi se srce* — *Сердце сжимается* — one's heart sinks — a feeling of disappointment, anxiety, or despondency. Here the dominant emotion is sadness, and a possible secondary emotion is fear. This leads to a classification of the entry in the dictionary article "Sadness/Fear", with a cross-reference to "Fear". Each meaning is illustrated with specific contexts demonstrating the idiom's usage in that particular sense. As a result of this approach, the dictionary becomes more flexible and informative.

4.3 Microstructure

The basic unit of the dictionary is the dictionary entry, which includes mandatory and optional lexicographic elements. The microstructure design, with authorial additions and adjustments, is based on the Academic Dictionary of Russian Phraseology's dictionary entry (Baranov et al. 2020), which serves as the foundation for the wordlist. The SRIDE dictionary entry consists of the following zones (Figure 2):

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

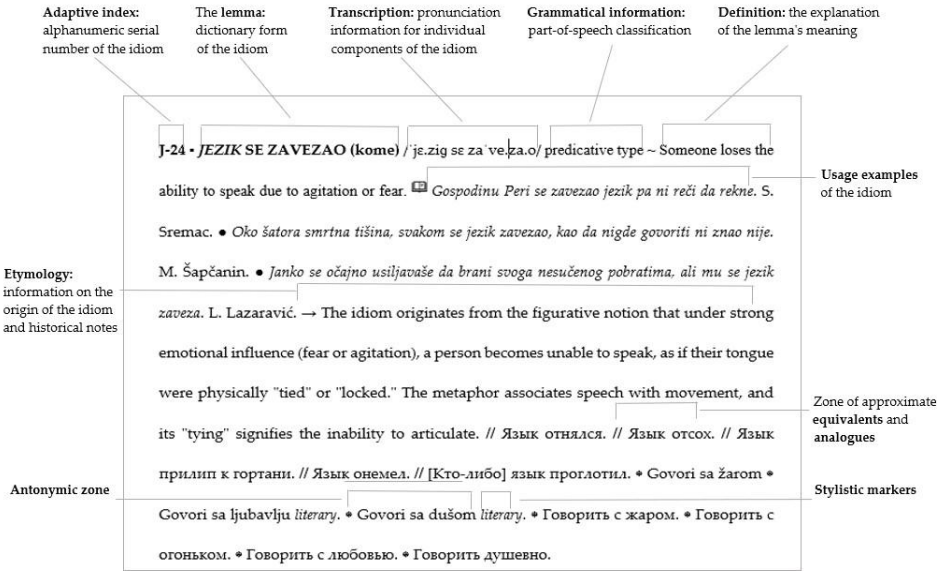


Figure 2: A mock-up of a dictionary entry in the SRIDE

Moreover, if an idiom is polysemous, the sense number is indicated before the definition using Arabic numerals. The zones for stylistic labels, definitions, and usage examples are repeated for each sense. If the stylistic characteristic remains the same for all senses, it is indicated in the lexical entry of the idiom, before the senses section. The structural elements of the dictionary entry include the following:

— *An adaptive index*, which is the serial number of the idiom facilitating its retrieval within the dictionary corpus. It consists of a letter (the initial letter of the core component) and a number indicating its position within that letter. For example, the idiom *Nevolja je natjerala nekoga* — strong dissatisfaction — has the serial number A-14.

— *The lemma* typed in bold uppercase letters, with the symbol "▪" preceding the phraseological unit. The core component is italicised. Features of the morphological realisation of the idiom's valencies are indicated in bold within parentheses: ▪ **IZBEČITI (od nečega)** ***OČI*** **(nečije / kod nekog)** — *to pop one's eyes out*. Variants of expressions are separated by the symbol "/", e.g. **IZBEČITI / ISKOLAČITI**. For an idiom with a verb in the infinitive, the morphological realisation of the subject is not provided. However, since most verbal idioms frequently occur in a specific conjugated form or exclusively in one non-infinitive form, the lexical entry selects the conjugated

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

verb form and specifies the subject's morphological realisation, e.g. **PADA MRAK NA *OČI* (kod nekoga)** — *blinded by rage*. Words and phrases that are optional components of the idiom are enclosed in square brackets, e.g. **NEK BUDE [TRI PUTA] *NESREĆAN*** — strong dissatisfaction, indignation, or irritation. Some idioms may include a negative particle or be used in an affirmative form. If the non-negated form is more common, only that variant is provided. If the expression is used exclusively with negation or the nonnegated form is perceived as less common, the dictionary entry strictly provides the negated form: **NE MOŽE SE *POKRENUTI*** — paralysed with fear.

— *A transcription* providing pronunciation information for individual components of the idiom. A simplified phonetic transcription is used to describe the phonetics of idioms, represented by lowercase symbols in italics enclosed within slash symbols "/". The transcription details the pronunciation features of each component, which are crucial for correct oral usage and comprehension of the idiom, e.g.: ▪ **NA *KRV* I NOŽ** — /na krv i nɔ̝ʒ/ — cruelly and mercilessly.

— *Grammatical information*, specifically the *part-of-speech classification*, which provides details on the grammatical category of the idiom, helping to determine its syntactic role in a sentence. Information on the part of speech aids in understanding the morphological characteristics and functional usage of the expression. Examples of phraseological units include: **KRSTITI SE LIJEVOM *RUKOM*** — to cross oneself with the left hand — predicative type; ***LJUT* KAO PAS** — furious as a dog — adjectival type; and **HLADNO JE OKO *SRCA* NEKOME** — it feels cold around the heart — adverbial type.

— *Stylistic markers* reflecting the tone, register, and contextual nuances of the idiom. Expressions are accompanied by stylistic markers such as *bookish*, *neutral*, *colloquial*, and *vernacular*. Each variant of an idiom in the lexical entry may be assigned a label, set in a smaller font and italics. The bookish register indicates that the idiom is typical of formal written speech, often found in scientific or journalistic texts: **OSTATI BEZ *DAHA*** *literary* — take one's breath away — surprise or admiration. The neutral register means the idiom is stylistically neutral and used across various spoken and written styles: **OD SVEG *SRCA*** *standard* — with all one's heart — sincerely, with deep devotion and full emotional commitment. A vernacular expression indicates that the idiom has a lowered stylistic tone and is typically used in everyday speech: **PUKNUTI OD *BIJESA*** *informal* — outburst of anger or fury. It is worth noting that the stylistic labels do not include *colloquial*, because most idioms inherently belong to the colloquial register. Thus, labelling them as *colloquial* would apply to the majority of entries. To avoid redundancy, colloquial idioms in the dictionary remain unmarked.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

— *The definition*, constituting a semantic explanation of an idiom's meaning, revealing its substantive content. It describes nuances of meaning, variability of interpretations, and specific connotations inherent to the given idiomatic unit, which is a key element for comprehensive lexicographic analysis. In cases of polysemy, the meanings are numbered, with each interpretation presented under the corresponding number and preceded by the symbol "∼". For example: ▪ **KAO MUVA BEZ *GLAVE* HODATI [TUMARATI]** ∼ **1.** To be confused, agitated, or disoriented. **2.** To not know what to do. **3.** To wander aimlessly; to act chaotically and without purpose.

— *Cross-references*, which are optional elements and are indicated when an idiom appears in multiple sections. They direct the user to related entries where each meaning is examined in detail. The reference is marked with "see", facilitating a comprehensive approach to polysemy and preventing information overload in the main entry by avoiding excessive duplication.

— *Usage examples*, marked with the symbol "🖳". Illustrative material is drawn from corpora of modern Serbian and Russian texts, as well as lexicographic sources. Idioms in examples are italicised and differentiated by the symbol ●. The number of examples typically does not exceed five. Their arrangement follows two principles: (1) from literary examples to media materials (prose first, then poetic); and (2) from primary forms of the idiom to secondary non-basic forms. Exceptions are permitted if an example more vividly illustrates the idiom's meaning.

— *Etymology and historical information*, providing data on the idiom's origin and historical development. It concisely describes roots, possible metaphorical transformations, and the evolution of the idiom's meaning, allowing for tracing the cultural-historical aspects of its formation. This zone is introduced by the symbol "→" and is obligatory.

— *A zone of approximate equivalents and analogues*, listing words, phraseological units, or expressions that are semantically or emotionally close to the given idiom. Equivalents are introduced by the symbol "//". For example, the expression *Eto ti ga na! — Well, there you go!* — depending on context, may convey surprise or disgust, such as in the following case:

> ▪ **ETO TI GA *NA*!** /ˈɛ.tɔ ti ga na/ interjectional type ∼ **1.** An expression of surprise, bewilderment, or irritation in response to an unexpected, undesirable, or inexplicable situation. // Вот тебе и на! // Вот так сюрприз! // Вот так номер! // Ну и дела! **2.** An expression of irritation and disappointment. // Ну и ну! // Ну вот, пожалуйста!

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

— *The equivalent selection zone*, which is optional as languages often contain non-equivalent examples. However, when linguistic correspondences exist, they are recorded in the dictionary without fail.

— *The antonymic zone* highlights opposing meanings of the given idiom. This element facilitates a deeper understanding of the language's semantic structure by demonstrating how contrasting emotional states are expressed. Antonyms are marked with "∗". Serbian antonyms are listed first, followed by Russian. Examples of antonyms for the idiom with the meaning *to faint from fear*:

> ▪ **PASTI U *NESVIJEST*** ∗ Biti smiren kao stijena ∗ Быть спокойным как удав. ∗ Олимпийское спокойствие. ∗ В ус не дует. ∗ Не подавать виду. ∗ Хоть трава не расти! ∗ Все трын-трава! ∗ Все параллельно! *informal*. ∗ По барабану *informal*. ∗ До фонаря *informal*.

It is important to note that the project deliberately omits a separate synonym zone, as thematic organisation renders it redundant: phraseological units are grouped by shared emotion, automatically placing synonymous expressions in the same section.

Thus, the microstructure of the SRIDE constitutes a system integrating obligatory and optional elements to provide a comprehensive representation of idiomatic units. Each dictionary entry is structured according to the functional and semantic features of expressions: from the thematic heading, which determines emotional colouring, to the adaptive index for organisation and quick retrieval. The entry includes components such as lexical input, transcription, grammatical information, and stylistic labels, capturing not only the base form of the idiom but also its phonetic, morphological, and stylistic characteristics.

Special attention is given to semantic diversity: in cases of polysemy, meanings are numbered, and corresponding zones of interpretation, usage examples, and stylistic labels are repeated for each, ensuring detailed semantic and pragmatic perspective. Additional elements such as cross-references, etymological data, and zones of equivalents and antonyms contribute to a comprehensive lexicographic analysis, helping users grasp the full spectrum of semantic and emotional nuances.

5. Conclusion

The SRIDE project is based on the integration of traditional lexicographic approaches with modern methods of corpus and cognitive analysis, which ensures not only the description of set expressions but also a deep understanding of their semantic, stylistic, and cultural characteristics. The methodology is grounded in the principle of thematic grouping of idioms into twelve key emotions, enabling contextual and emotional differentiation of the material. Problematic areas, such as the

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

polysemy of idioms, are addressed through multifaceted classification with the identification of dominant emotional nuances, the use of cross-references, and adaptive indexing, facilitating precise matching of expressions in both language systems.

The microstructure of dictionary entries has been developed with consideration of mandatory and optional components, including lexical entry, transcription, grammatical information, stylistic labels, definitions, usage examples, etymological data, and equivalent and antonym zones. Particular attention is paid to maintaining the integrity and functional homogeneity of each entry: a uniform structure ensures quick search, comparison, and matching of idiomatic units, which holds significant practical value for translators, linguists, teachers, and students.

The use of extensive linguocultural sources, such as national language corpora, as well as materials from media and literary texts, allows for the documentation of both established and emerging idioms. This is particularly relevant in the context of the dynamic evolution of idiomatic inventory and continuous cultural interaction, where emotional speech coloring plays a key role in intercultural communication.

The SRIDE project represents a lexicographic work in which methodological principles — based on frequency analysis, normativity, contextualisation, and equivalence — are integrated with modern electronic platforms (OnLex). The systematic approach to designing the microstructure and the well-considered solutions to problematic areas of polysemy contribute to the creation of a reliable tool for cross-linguistic analysis of emotional expression, meeting the needs of both specialists and a broader audience. The SRIDE serves as a specialised resource that contributes to a deeper understanding of emotional expression in language and supports the development of contemporary lexicographic theory and practice.

The first edition of the SRIDE dictionary is scheduled for release in 2026, with final compilation and digital integration expected to be completed by late 2025. Future stages of the project include multilingual expansion, enhanced corpus-based updates, and the refinement of lexicographic annotation techniques based on digital corpora.

References

A. Dictionaries

Alefirenko, N.F. and L.G. Zolotyh (Eds.). 2008. *Frazeologičeskij slovar': Kul'turno-poznavatel'noe prostranstvo russkoj idiomatiki.* Moscow: ELPIS.

Baranov, A.N., M. Voznesenskaja, D.O. Dobrovol'skij, K. Kiseleva and A. Kozerenko (Eds.). 2020. *Akademicheskij slovar' russkoj frazeologii.* Moscow: LEXRUS.

Basko, N. (Ed.). 2002. *Frazeologizmy v russkoj reči.* Slovar'-spravočnik. Moscow: Flint, Nauka.

Matešić, J. (Ed.). 1982. *Frazeološki rječnik hrvatskoga ili srpskog jezika.* Zagreb: Školska knjiga.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2031 (Project)

Menac, A., R. Venturin, T. Korać and M. Popović (Eds.). 1979. *Rusko hrvatski ili srpski frazeoloski rjecnik.* Zagreb: Školska knjiga.

Milenković, T. (Ed.). 2006. *Idiomi u srpskom jeziku.* Beograd: Narodna biblioteka Srbije.

Mršević-Radović, D. 1987. *Frazeoloske glagolsko-imenicke sintagme u savremenom srpskohrvatskom jeziku.* Beograd: Filoloski fakultet Beogradskog univerziteta.ˇ

Otašević, D. (Ed.). 2012. *Frazeološki rečnik srpskog jezika.* Novi Sad: Prometej.

Subbotina, L. (Ed.). 2024. *Slovar' frazeologizmov russkogo jazyka.* Moscow: AST.

Telija, V. (Ed.). 2024. *Bol'šoj frazeologičeskij slovar' russkogo jazyka.* Moscow: AST-Press Shkola, Gramota.

Trofimkina, O. (Ed.). 2005. *Serbokhorvatsko–russkij frazeologicheskij slovar'.* Moscow: Vostok-Zapad.

Zhukova, M. and A. Zhukov (Eds.). 2023. *Slovar' sovremennoj russkoj frazeologii.* Moscow: AST-Press Shkola, Gramota.

B. Other literature

Bayramova, L. 2010. Phraseological Dictionaries of the New Generation. *Vestnik Chelyabinskogo Gosudarstvennogo Universiteta* 29(210): 27-32.

Dragičević, R. 2024. Problems in the Selection of Entries in the Serbian–Russian Dictionary. *Stephanos* 63(1): 54-64.

Dragičević, R., M.D. Korolkova, D.A. Ryzhova, Y.L. Shapich and E.I. Yakushkina. 2024. A New Large Serbian–Russian Dictionary (A General Conception and Problems of Lexicographic Description). *Voprosy leksikografii* 32: 43-60.

Dubichinskiy, V. 2009. *Leksikografiya russkogo yazyka*. Moscow: Flint, Nauka.

Dzonic, V. 2016. Russian and Serbian Phraseological Units: Experience of Comparative Analysis. *Vestnik Kemerovskogo gosudarstvennogo universiteta. Philology* 2: 166-170.

Faber, P., P. León-Araúz and A. Reimerink. 2014. Representing Environmental Knowledge in EcoLexicon. Bárcena, E., T. Read and J. Arús (Eds.). 2014. *Languages for Specific Purposes in the Digital Era:* 267-301. Cham: Springer.

Fraser, B.L. 2008. Beyond Definition: Organising Semantic Information in Bilingual Dictionaries. *International Journal of Lexicography* 21(1): 69-93.

Izard, C.E. 2009. *The Psychology of Emotions.* Second edition. Saint-Petersburg: Piter.

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.

Makarov, Y., M. Melenchenko and D. Novokshanov. 2022. Digital Resources for the Shughni Language. Ojha, A.K., S. Ahmadi, C.-H. Liu and J.P. McCrae (Eds.). 2022. *Proceedings of the Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference 2022, Marseille, France, 3–20 June 2022*: 61-64. European Language Resources Association.

Russian National Corpus (RNC). https://ruscorpora.ru/ [6 April 2025]

Serbian Web Corpus (srWaC). https://www.clarin.si/ske/#open [6 April 2025]

Štrbac, G. 2018. *Frazeologija o čoveku i čovek u frazeologiji.* Novi Sad: Filozofski fakultet.

Veisbergs, A. 2013. Phraseological Units and Idioms, East and West and Where Do We Stand. *Baltic Journal of English Language, Literature and Culture* 3(1): 109-121.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

Traitement des Verbes Auxiliaires Modaux Polysémiques dans le Dictionnaire Chinois–Français Basé sur le Prototype et le Schéma

Zhijie Tang, *Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, Guangdong, China (zhijie.tang2024@outlook.com)*
(*https://orcid.org/0009-0009-1324-2483*)

Résumé: En suivant les théories du prototype et du schéma, cette recherche met en place un cadre de traitement adapté aux verbes auxiliaires modaux polysémiques dans le dictionnaire chinois–français, en se basant sur les caractéristiques de ces verbes. Elle propose également des mesures concrètes pour améliorer le traitement dans trois domaines: (1) marquage prototypique de la partie du discours, soit marquer 动 [aux. modal] pour les cas les plus typiques et ajouter les propriétés de la partie du discours; (2) sélection et arrangement des sens, équivalents et exemples autour du prototype et du schéma, soit étendre les sens, les équivalents et les exemples autour des types modaux; (3) discrimination dans la catégorie du prototype, soit renforcer de la discrimination des synonymes au sein de la même catégorie lexicale à l'aide de remarques et d'autres formes. Enfin, les cas optimisés sont présentés à la fin de cet article.

Mots-clés: PROTOTYPE, SCHÉMA, VERBES AUXILIAIRES MODAUX POLYSÉMIQUES, DICTIONNAIRE CHINOIS–FRANÇAIS, DICTIONNAIRE D'APPRENTISSAGE

Abstract: Treatment of Polysemous Modal Auxiliary Verbs in Chinese–French Dictionary Based on Prototype and Schema. Based on the theories of prototype and schema, this paper constructs a suitable processing framework for polysemous modal auxiliary verbs in Chinese–French dictionary according to characteristics of modal auxiliary verbs, and elaborates specific improvement measures from the following three aspects: (1) prototypical annotation of word class, i.e., labelling the most typical cases with 动 [aux. modal] and adding word class properties; (2) selection and arrangement of the senses, equivalents and examples based on prototype and schema, i.e., gradually extending senses, equivalents and examples around modal type; (3) discrimination within the prototype category, i.e., strengthening the synonym discrimination within the same lexical category through boxes and other forms. Finally, the optimised cases presented at the end of the paper.

Keywords: PROTOTYPE, SCHEMA, POLYSEMOUS MODAL AUXILIARY VERBS, CHINESE–FRENCH DICTIONARY, LEARNER'S DICTIONARY

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

1. Introduction

Comme outil important pour les apprenants et les enseignants, le dictionnaire bilingue s'avère crucial dans la performance linguistique des apprenants. Fondamentalement, le but du dictionnaire bilingue est de trouver des équivalents entre les unités lexicales d'une langue et celles d'une autre (Zgusta 1971: 294). Or, l'équivalence complète se révèle extrêmement rare. Quant aux verbes auxiliaires modaux polysémiques, ils se recoupent souvent un peu dans les différents systèmes linguistiques, ce qui rend leur définition dans les dictionnaires bilingues plus compliquée.

Dans cette étude, nous nous concentrons sur l'aspect de la compilation lexicographique des dictionnaires bilingues, en particulier sur les dictionnaires d'apprentissage chinois–français (monodirectionnel). Un dictionnaire d'apprentissage fournit aux apprenants non seulement des équivalents lexicaux, mais également une compréhension nuancée des significations et des usages des mots dans la langue cible. Ces dictionnaires sont spécifiquement conçus pour répondre aux besoins des apprenants, avec une attention particulière à la sélection du vocabulaire, à la présentation des définitions, à l'inclusion d'exemples de phrases et à l'indication des informations grammaticales essentielles. Pour les verbes auxiliaires modaux polysémiques, cette exigence est particulièrement cruciale, car leur interprétation dépend fortement du contexte et de la connaissance des schémas situationnels associés. Or, les approches lexicographiques traditionnelles peinent parfois à rendre compte de cette complexité sémantique et pragmatique de manière accessible aux apprenants. Une enquête a été menée sur deux dictionnaires bilingues chinois–français existants: le *Grand Dictionnaire Chinois–Français Contemporain* (GDCFC) et le *Dictionnaire Ricci Chinois–Français* (DRCF), publiés en 2014. Bien que ces dictionnaires soient influents, nous constatons qu'il n'existe pas de modes de marquage et de définition uniformes pour les entrées des verbes auxiliaires modaux polysémiques et que les informations définitoires sont parfois incomplètes ou imprécises[1].

En réponse à ces insuffisances, l'intégration de modèles théoriques du prototype et du schéma, pourrait offrir des pistes prometteuses pour améliorer la clarté et l'efficacité des dictionnaires d'apprentissage, cette étude propose une nouvelle méthode d'optimisation, fondée sur les théories des types modaux, du prototype et du schéma. L'objectif principal est d'élaborer un cadre unifié pour traiter la polysémie des verbes auxiliaires modaux, en offrant des solutions pratiques pour améliorer leur présentation dans la microstructure des dictionnaires d'apprentissage chinois–français.

2. Caractéristiques des verbes auxiliaires modaux en chinois et en français

Les verbes auxiliaires modaux, comme leur nom l'indique, peuvent exprimer des significations modales d'un point de vue sémantique et peuvent être utilisés comme verbes auxiliaires d'un point de vue grammatical.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

Sur le plan sémantique, les significations modales des verbes auxiliaires modaux relèvent essentiellement de trois principaux types modaux: modalité épistémique, déontique et dynamique (Palmer 1990; Peng 2007; Nuyts et Van der Auwera 2016). Les verbes auxiliaires ayant au moins deux types modaux peuvent être appelés verbes auxiliaires modaux polysémiques, tels que *应该 (yīnggāi)*, *应当 (yīngdāng)*, *要 (yào)*, *会 (huì)*, *能 (néng)* en chinois, *pouvoir* et *devo*ir en français.

Dans le domaine grammatical, les verbes auxiliaires modaux en chinois et en français apparaissent tous dans la construction «verbe auxiliaire modal + verbe principal», mais les verbes auxiliaires modaux chinois peuvent aussi être suivis d'adjectifs. Du point de vue des parties du discours, les verbes auxiliaires modaux dans les deux langues ont des caractéristiques en commun avec d'autres parties du discours, car les frontières ne sont pas toujours très claires, ce qui rend plus compliqué de savoir exactement de quoi il s'agit.

En examinant les recherches précédentes en linguistique chinoise, nous avons identifié les cinq verbes auxiliaires modaux polysémiques les plus consensuels: *应该 (yīnggāi)*、*应当 (yīngdāng)*、*要 (yào)*、*会 (huì)*、*能 (néng)*. La complexité et les insuffisances de leur définition ont déjà été révélées par notre enquête. Dans ce travail, les verbes auxiliaires modaux moins typiques, atypiques ou controversés ne sont pas abordés. Cependant, compte tenu des différences entre les systèmes linguistiques chinois et français, lors de la recherche des expressions françaises correspondantes dans ce qui suit, nous prendrons en compte divers marqueurs modaux tels que les verbes auxiliaires modaux, les adverbes, les locutions, les temps et les flexions.

3. Théories du prototype et du schéma

Dans les années 1950, Wittgenstein met en avant la notion de ressemblances de famille, qui produit une grande impression dans les domaines de la philosophie, de la psychologie et de la linguistique. Labov (1973) et Rosch (1975) publient les résultats de leurs recherches expérimentales, prouvant que les ressemblances de famille peuvent être appliquées à la description de nombreuses catégories naturelles. Ils appellent ces catégories «catégories du prototype» et proposent la théorie du prototype.

Selon la théorie du prototype, tous les membres d'une catégorie ne sont pas également représentatifs. Certains sont perçus comme plus typiques que d'autres. Parmi eux, le prototype est considéré comme l'exemplaire le plus représentatif, celui qui incarne le mieux les caractéristiques centrales de la catégorie. La catégorie s'établit autour du prototype, et sa frontière s'étend vers l'extérieur selon le degré de représentativité, formant ainsi une catégorie plus large dont les limites sont difficiles à déterminer. En même temps, il convient de noter que le prototype n'est pas nécessairement unique. La plupart des unités linguistiques présentent non pas une structure monocentrique mais une structure polycentrique, soit une catégorie avec de multiples représentations du prototype, ce

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

qui correspond à la polysémie du vocabulaire (Zhang 2009).

Néanmoins, pour certaines catégories comme «meuble» et «petit-déjeuner», il est difficile de donner un exemple largement représentatif parce que les membres ne possèdent que quelques caractéristiques abstraites. Ainsi, nous reconnaissons les limites du prototype: «catégoriser de nouvelles entités avec une observation intuitive et une similarité, cela brouille la hiérarchie et les particularités dans la même catégorie, en ignorant notre capacité d'abstraction et de généralisation dans le processus cognitif» (Zhang 2017: 1). De ce fait, les chercheurs essaient de créer un modèle sémantique plus abstrait et plus généralisé, soit le schéma, pour décrire la relation entre les prototypes et les instances, démontrer les différences entre les instances et faciliter la définition basée sur la catégorisation.

Revenant à nos objets d'étude, les théories du prototype et du schéma nous aideront à améliorer le traitement des verbes auxiliaires modaux polysémiques, sous les aspects de la partie du discours, des équivalents, des sens, des exemples et de la discrimination. Ces théories facilitent la détermination des parties du discours, des équivalents et des exemples, la division et l'arrangement des sens, ainsi que la discrimination des mots dans la même catégorie.

4. Application du prototype et du schéma

A l'aide du prototype et du schéma, nous proposons le mode de définition optimisé et quelques mesures d'amélioration sous trois aspects: marquage prototypique de la partie du discours; sens, équivalents et exemples autour du prototype et du schéma; discrimination dans la catégorie du prototype.

4.1 Mode de définition basé sur le prototype et le schéma

En fonction de la classification de la modalité et de l'analyse de la signification modale (Kronning 2001; Le Querler 2004; Sueur1979; Lin 2019; Lü 2001), les verbes auxiliaires modaux polysémiques couvrent au moins deux types modaux (déontique et épistémique), comme les mots chinois *应该 (yīnggāi)*, *应当 (yīngdāng)* et le mot français *devoir;* mais les cas comprenant trois modalités tels que *要 (yào)*, *会 (huì)*, *能 (néng)* et *pouvoir*, sont également courants. De cela, nous pouvons extraire un modèle sémantique idéal des verbes auxiliaires modaux polysémiques (voir Figure 1).

Dans le processus de construction de la catégorie des verbes auxiliaires modaux polysémiques autour du modèle idéal, le marquage de la partie du discours pour les cas typiques peut constituer notre première étape. Ensuite, nous remarquons qu'il existe, au sein des modalités, quelques éléments sémantiques relativement centraux et originaux qui contiennent des changements d'intensité et qui peuvent s'étendre aux points sémantiques plus forts, plus faibles ou plus raffinés (voir Figure 2). Si nous prenons certains points sémantiques dans deux ou trois modalités comme «entités centrales», les significations des

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

verbes auxiliaires modaux polysémiques peuvent se construire autour de plusieurs sens prototypes.

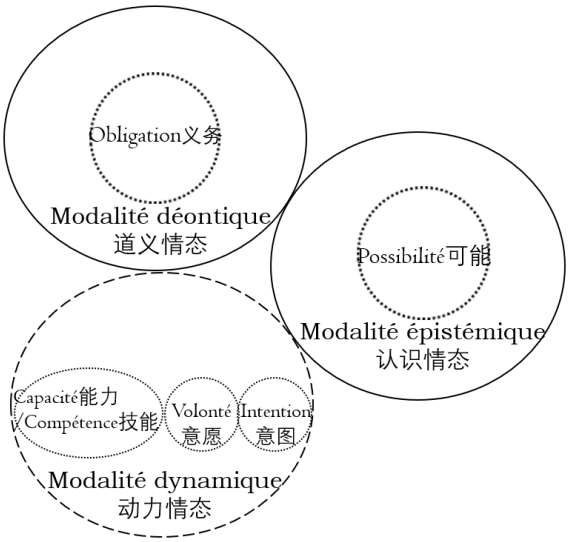


Figure 1: Modèle sémantique idéal des verbes auxiliaires modaux polysémiques

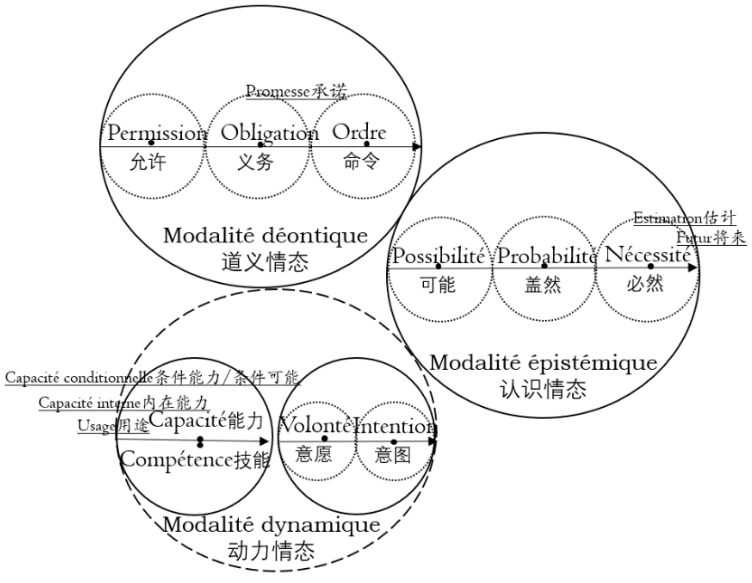


Figure 2: Points sémantiques des verbes auxiliaires modaux polysémiques

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

Selon la *Lexicographie contemporaine*, l'équivalence partielle se produit souvent à un certain point sémantique ou position sémantique. Dans la plupart des cas, elle est en fait la correspondance croisée de points sémantiques (Zhang et Yong 2007: 156-157). De cela, nous prenons les points sémantiques comme repères en étendant les significations des verbes auxiliaires modaux polysémiques chinois et des équivalents français. En même temps, nous veillons à assurer la précision des équivalents en les restreignant par leur usage. Ainsi, les sens, les équivalents et les exemples sont naturellement étendus autour des points sémantiques des types modaux dans un réseau de définition unifié (voir Figure 3), l'exhaustivité et la précision de la définition sont donc améliorées. Par ailleurs, les différences sémantiques des verbes auxiliaires modaux polysémiques sont évidentes dans la Figure 3, ce qui favorise la discrimination des cas dans la catégorie modale.

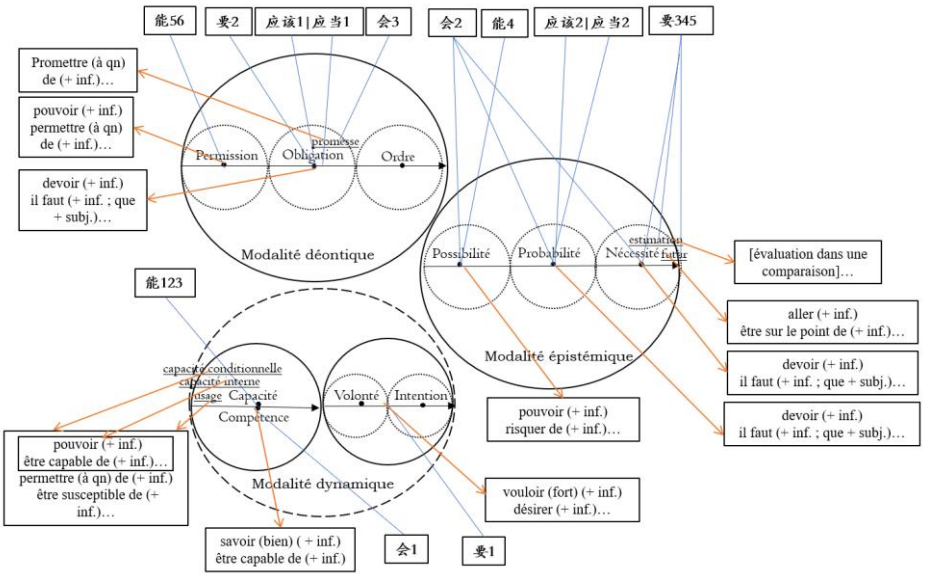


Figure 3: Réseau de définition unifié des verbes auxiliaires modaux polysémiques

4.2 Marquage prototypique de la partie du discours

La théorie du prototype est appliquée à la division des parties du discours en chinois depuis longtemps. En vue de régler l'incertitude des classes de mots, avec plus de 9 000 mots couramment utilisés comme objets d'étude, Yuan (1995: 170; 2010: 9) présente le système de parties du discours chinois basé sur le prototype, puis il propose la classification floue[2] par laquelle nous comprenons que certains mots n'ont pas à être complètement classés dans une certaine partie du discours, ils peuvent avoir différents degrés d'appartenance dans différentes

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

parties du discours, ce qui est bénéfique pour traiter le problème du marquage de la partie du discours dans les dictionnaires chinois–français.

En faisant référence aux recherches dans le domaine chinois–anglais, le marquage de la partie du discours des entrées est propice à (1) l'explication des usages; (2) la détermination des sens; (3) la précision de la définition; (4) la coordination des exemples et des parties du discours; (5) l'amélioration de l'efficacité de l'utilisation du dictionnaire; (6) la résolution du problème des parties du discours du chinois (Wang 2006). En outre, du point de vue de la psycholinguistique, les parties du discours sont étroitement liées au sens abstrait des mots dans le lexique mental, et des «étiquettes» de partie du discours sont automatiquement attachées au sens des mots (Wang et Zhang 2006: 139). Par conséquent, dans l'attente d'optimiser la définition des dictionnaires chinois–français, le manque de partie du discours est un problème urgent à résoudre. Cependant, quelle partie du discours devons-nous marquer pour les verbes auxiliaires modaux?

Pour répondre à cette question, nous faisons référence à deux dictionnaires qui marquent de manière exhaustive les parties du discours dans leur révision: la cinquième édition du *Dictionnaire Chinois Contemporain*[3] et la troisième édition de *A Chinese–English Dictionary*. Les traitements des deux dictionnaires sont similaires. Les équivalents sont principalement divisés selon différentes parties du discours du mots chinois, et la partie du discours est marquée avant le sens correspondant. Pour *应该 (yīnggāi)*、*应当 (yīngdāng)*, le *Dictionnaire Chinois Contemporain* marque 动助动词, *A Chinese–English Dictionary* marque 动 [as aux.]; concernant *要 (yào)*、*会 (huì)*、*能 (néng)* en tant que verbes auxiliaires modaux, le *Dictionnaire Chinois Contemporain* marque 动助动词 mais *A Chinese–English Dictionary* ne marque que 动. Parmi les formes, la marque de 动 [aux.] avant les sens modaux ou après le mot défini semble être un bon choix car il permet de ne pas marquer la partie du discours de chaque équivalent, mais l'emploi d'auxiliaire (semi-auxiliaire) des expressions françaises et du mot chinois peuvent encore être indiqué. Cependant, pour notre recherche, nous essayons de les marquer davantage comme 动 [aux. modal] puisque leur statut représentatif et leur haute typicité sont déterminés. De cette façon, leur appartenance à la partie du discours en chinois peut être présentée plus de façon plus précise, ce qui est propice à l'établissement de la catégorie des verbes auxiliaires modaux.

Pour les verbes auxiliaires modaux, nous avons tendance à marquer 动 [aux. modal] pour les cas les plus typiques, puis à ajouter les propriétés de la partie du discours afin que les utilisateurs puissent avoir une compréhension plus systématique. Le degré d'appartenance, à nos yeux, pourrait être utile pour les chercheurs mais n'est pas une information nécessaire pour les apprenants ordinaires. Compte tenu du fait que cette étude s'inscrit dans la perspective d'apprenants ordinaires et de dictionnaires d'apprentissage chinois–français, nous n'ajoutons pas de degré d'appartenance spécifique lors du marquage de la partie du discours.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

4.3 Sens, équivalents et exemples autour du prototype et du schéma

Selon le mode de définition optimisé, nous organisons les sens, les équivalents et les exemples autour du prototype et du schéma.

4.3.1 Sens

En fonction de la Figure 2 et de la signification modale des verbes auxiliaires modaux chinois, nous pouvons mettre en relief les sens essentiels des cas par des mots guides:

应该：动 [aux. modal] [OBLIGATION] [PROBABILITÉ]
应当：动 [aux. modal] [OBLIGATION] [PROBABILITÉ]
要：动 [aux. modal] [VOLONTÉ] [OBLIGATION] [NÉCESSITÉ]
会：动 [aux. modal] [COMPÉTENCE] [PROMESSE] [POSSIBILITÉ↔NÉCESSITÉ]
能：动 [aux. modal] [CAPACITÉ] [PERMISSION] [POSSIBILITÉ]

Avec les mots guides, les différents sens dans les types modaux sont mieux distingués et divisés. Pour les utilisateurs, la mise en évidence des sens facilite la navigation dans la définition et la compréhension des sens principaux des verbes auxiliaires modaux polysémiques.

En ce qui concerne l'arrangement des sens, les études antérieures ont révélé qu'il existe un processus de développement de la modalité radicale (déontique et dynamique) à la modalité épistémique. Tout d'abord, nous approuvons la prise en compte de l'évolution des modalités. Afin d'obtenir le meilleur effet déclencheur du schéma cognitif des verbes auxiliaires modaux polysémiques chez les utilisateurs, il vaut mieux indiquer la relation entre les sens représentatifs dans l'évolution des significations des mots (Fang et Chen 2008: 74). Selon les recherches sur l'acquisition des verbes auxiliaires modaux en Chine et à l'étranger, l'acquisition des types modaux chez les enfants est parallèle au développement diachronique des types sémantiques des verbes auxiliaires modaux (Peng 2007: 58), et les élèves internationaux suivent également la loi cognitive naturelle de la modalité radicale à la modalité épistémique (Lai 2012: 16-17). Alors, quel est l'ordre des deux modalités (dynamique et déontique) dans la modalité radicale? Selon les recherches sur l'évolution sémantique des verbes auxiliaires modaux (Jia 2020; Yang 2017), une partie considérable des verbes auxiliaires épistémiques et déontiques évoluent à partir de la modalité dynamique. Nous rangeons ainsi les sens en ordre de la modalité dynamique → déontique → épistémique. L'Appliquons à nos objets d'étude:

应该：动 [aux. modal] ① [OBLIGATION] ② [PROBABILITÉ]
应当：动 [aux. modal] ① [OBLIGATION] ② [PROBABILITÉ]
要：动 [aux. modal] ① [VOLONTÉ] ② [OBLIGATION] ③ [NÉCESSITÉ]
会：动 [aux. modal] ① [COMPÉTENCE] ② [PROMESSE]
③ [POSSIBILITÉ↔NÉCESSITÉ]
能：动 [aux. modal] ① [CAPACITÉ] ② [PERMISSION] ③ [POSSIBILITÉ]

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

L'ordre de la modalité radicale avant la modalité épistémique, se conforme non seulement à la loi cognitive universelle, mais construit aussi une liaison sémantique naturelle avec d'autres parties du discours comme noms et verbes des mots définis.

4.3.2 Équivalents

Sur le plan des équivalents, nous nous inspirons des études en traductologie. Dans ce travail, l'équivalent français qui chevauche le plus possible la signification du mot chinois et qui s'utilise dans tous les contextes, peut être sélectionné comme équivalent fixe, tandis que les expressions françaises appartenant à une modalité spécifique sont considérées comme équivalents temporaires[4]. En outre, étant donné la spécificité de la catégorie des verbes auxiliaires modaux, il est recommandé de prendre pour équivalents fixes des mots français qui peuvent être utilisés comme verbes auxiliaires modaux. Selon les critères ci-dessus, nous sélectionnons *devoir (+ inf.)* comme l'équivalent fixe de *应该 (yīnggāi)* et de *应当 (yīngdāng)*, *pouvoir (+ inf.)* comme l'équivalent fixe de *能 (néng)*. Parmi tous les équivalents dans les sens, nous avons tendance à mettre en premier les équivalents fixes, par exemple:

> 应该：动 [aux. modal] ① [OBLIGATION] devoir (+ *inf.*) … ② [PROBABILITÉ] devoir (+ *inf.*) …
> 应当：动 [aux. modal] ① [OBLIGATION] devoir (+ *inf.*) … ② [PROBABILITÉ] devoir (+ *inf.*) …
> 能：动 [aux. modal] ① [CAPACITÉ] pouvoir (+ *inf.*) … ② [PERMISSION] pouvoir (+ *inf.*) … ③ [POSSIBILITÉ] pouvoir (+ *inf.*) …

Le mode avec les équivalents fixes en tête, améliore l'efficacité de la consultation et s'adapte aussi au modèle concis du dictionnaire électronique.

Quant aux équivalents temporaires, nous avons vérifié leur fréquence par l'intermédiaire de Sketch Engine. En général, plus la fréquence est élevée, plus le mot est typique. Selon les résultats, nous avons choisi d'abord les équivalents les plus typiques pour le sens spécifique, puis les classons dans le sens par fréquence. Prenant les équivalents du sens de possibilité de *能 (néng)* comme exemple, nous obtenons l'ordre suivant: pouvoir (+ *inf.*) 5,634,956 > risquer de (+ *inf.*) 495,984 > avoir la possibilité de (+ *inf.*) 154,717 > être susceptible de (+ *inf.*) 130,574 … Pour éviter un nombre excessif d'équivalents et assurer leur typicité, nous ne retenons essentiellement que les expressions françaises dont la fréquence est supérieure ou égale à six chiffres. Dans ce processus, il faut tenir compte de la correspondance entre les équivalents et le mot chinois spécifique afin d'exclure les expressions inappropriées et atypiques pour le mot défini. D'autres marqueurs modaux, s'ils existent, peuvent être ajoutés à la fin des équivalents ou dans les exemples.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

4.3.3 Exemples

A l'égard des exemples, nous étendons d'abord les exemples autour des différents types modaux et fournissons le contexte spécifique qui s'accorde avec les sens modaux. Par exemple, pour *应该 (yīnggāi)*, il convient de fournir des exemples du sens d'obligation (*你不应该打他！Tu ne dois pas le frapper!*) ainsi que du sens de probabilité (*事情的经过应该就是这样。C'est ainsi que les choses ont dû se passer.*). Les exemples sont répartis de manière équilibrée dans différentes modalités et le problème de manque dans certaines modalités est ainsi résolu.

Ensuite, compte tenu des propriétés de la catégorie des verbes auxiliaires modaux polysémiques, il est nécessaire d'assurer la typicité des exemples correspondant aux sens différents et de présenter quelques phrases d'exemple reflétant les caractéristiques syntaxiques comme «*我能不能借用您的词典? Puis-je emprunter votre dictionnaire?*», afin que les apprenants puissent avoir une compréhension précise et systématique des verbes auxiliaires modaux polysémiques tant du point de vue sémantique que syntaxique.

Enfin, d'autres formes d'expression modale que les verbes auxiliaires modaux, comme *probablement* dans «*她应该已经走了。Elle est probablement partie.*», le conditionnel dans «*今晚能下雨吗? Pourrait-il pleuvoir ce soir?*» et le futur dans «*她会不会来? Viendra-t-elle?*», peuvent également être ajoutées aux exemples selon les besoins.

4.3.4 Discrimination dans la catégorie du prototype

Etendue à la dimension des relations lexicales, la théorie du prototype est utilisée dans la discrimination des membres de la même catégorie. Sémantiquement parlant, les verbes auxiliaires modaux polysémiques possèdent non seulement les mêmes caractéristiques prototypiques telles que la modalité déontique et la modalité épistémique, mais également leurs propres caractéristiques non-prototypiques, soit différents sens dans le type modal, tels que la permission et l'obligation dans la modalité déontique, la possibilité et la nécessité dans la modalité épistémique. Les différences manifestées par les traits non-prototypiques sont aussi claires dans les Figures 2 et 3 en jetant ainsi la base de la discrimination sémantique des verbes auxiliaires modaux polysémiques.

Les caractéristiques non-prototypiques en syntaxe peuvent être plus évidentes. Grammaticalement parlant, les verbes auxiliaires modaux ont plusieurs propriétés communes, mais chacun a sa collocation ou utilisation unique. Les particularités syntaxiques communes de cette catégorie sont considérées comme des caractéristiques prototypiques, alors que les structures spécifiques à certains mots comme «*être compatible avec 一定(yīdìng)*» deviennent la clé de la discrimination. Quant à la forme spécifique de la discrimination, nous pouvons ajouter une case de remarque à la fin de l'article dans le dictionnaire.

En un mot, le prototype, le schéma et la catégorisation facilitent la construction d'un cadre de définition unifié en réduisant la confusion entre les membres d'une même catégorie.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

5. Représentation de cas

A travers tous les travaux d'analyse, nous proposons des mesures d'amélioration de la définition sous la direction des théories du prototype et du schéma, en établissant un réseau de définition unifié pour les verbes auxiliaires modaux polysémiques (voir Figure 3), afin d'améliorer l'efficacité, l'exhaustivité et la précision de la définition, ainsi que la catégorisation de cette partie du discours.

D'abord, en termes de partie du discours, nous marquons 动 [aux. modal] pour les cas typiques dans notre recherche. Ensuite, pour les sens et les équivalents, nous prenons les mots guides comme points d'origine pour étendre les expressions françaises correspondantes de différents types modaux, parmi lesquelles les expressions avec la coïncidence sémantique la plus élevée ou celles de la modalité radicale avec le plus de fréquence ont la priorité; les usages des équivalents sont aussi ajoutés. Puis, sur le plan des exemples, en reflétant la signification et l'usage, nous tentons d'organiser des exemples de différentes modalités, de diversifier le modèle de la phrase, le temps et l'aspect. Enfin, nous effectuons une discrimination des synonymes si nécessaire.

Voici les définitions optimisées des cas, au même format que le *Grand Dictionnaire Chinois–Français Contemporain* (GDCFC). Selon son mode d'emploi, les entrées multi-mots sont placées entre « 【】 ». Pour les entrées à un seul mot, l'écriture standard est adoptée, tandis que les caractères traditionnels et les variantes sont placés entre «( )» après l'écriture standard. S'il existe deux sens ou plus, ils sont marqués par ❶❷❸, etc. A l'égard des exemples, ils sont précédés de « ▷ » et séparés par «||». Le mot défini dans les exemples est remplacé par «~», mais l'entrée à un seul mot n'est pas remplacée. Lorsqu'il existe deux traductions d'un exemple dans le même sens, les deux phrases complètes sont séparées par «|». Les phrases d'exemple suivantes sont forgées par nous-mêmes ou adaptées des *Huit cents mots en chinois moderne (édition augmentée)* (Lü 2001), du corpus de BCC et des dictionnaires: le *GDCFC*, Le Robert et Larousse. De surcroît, il faut noter le soulignement des équivalents que nous faisons dans les exemples, qui permet aux utilisateurs de localiser les équivalents rapidement, d'améliorer ainsi l'efficacité de la consultation et d'éviter les malentendus causés par la méconnaissance de la conjugaison des verbes. Quant aux informations accessoires, les marques grammaticales et rhétoriques sont indiquées par des abréviations entre « 〈 〉 », par exemple, 〈litt.〉 signifiant «littéraire». « • » est le symbole du renvoi. Les explications ou la discrimination d'utilisation des entrées sont précédées de «**REM**». Cela permet aux utilisateurs de comprendre la signification ou les usages plus nuancés, mettant en lumière les différences entre les synonymes et renforçant la pertinence entre les entrées. De plus, pour inciter les utilisateurs à étudier systématiquement les verbes auxiliaires modaux polysémiques, nous ajoutons «**voir aussi:** bilan des verbes auxiliaires modaux polysémiques dans les annexes» à la fin des entrées de la même manière que Le Robert le fait pour les semi-auxiliaires.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

【应该】yīnggāi 动 [aux. modal] ❶ **[OBLIGATION]** devoir (+ *inf.*); il faut (+ *inf.*; que + *subj.*); avoir à (+ *inf.*); il convient de (+ *inf.*); il est nécessaire de (+ *inf.*); se devoir (de) (+ *inf.*); être tenu(e) de (+ *inf.*); devoir (qch. abstrait, affectif ou sentimental) à (qn) ▷ 你不~打他！Tu ne dois pas le frapper! | Il ne faut pas le frapper! || ~立即通知他/她。Il faut l'avertir tout de suite. **△注意区分：**他需要休息。Il lui faut du repos. || 他们今晚~完成这项工作。Ils ont à terminer ce travail ce soir. || ~参加这场会议。Il convient d'assister à cette réunion. || 我们~以身作则。Nous nous devons de donner l'exemple. || 我~向您道歉/尊敬您。Je vous dois des excuses/le respect. ❷ **[PROBABILITÉ]** devoir (+ *inf.*); il faut (+ *inf.*; que + *subj.*); probablement; apparemment; il est probable que; pouvoir (+ *inf.*) ▷ 事情的经过~就是这样。C'est ainsi que les choses ont dû se passer. || 他~是疯了才会这么做。Il faut qu'il soit fou pour agir comme ça. || 她~已经走了。Elle est probablement partie. | Il est probable qu'elle est partie. | Elle a dû partir. || 这位女士~有40岁了。Cette dame peut avoir 40 ans. • ≈应当
Voir aussi: bilan des verbes auxiliaires modaux polysémiques dans les annexes

【应当】yīngdāng 动 [aux. modal] 〈litt.〉 ❶ **[OBLIGATION]** 〈多见于法律文件〉 devoir (+ *inf.*); il faut (+ *inf.*; que + *subj.*); avoir à (+ *inf.*); il convient de (+ *inf.*); il est nécessaire de (+ *inf.*); se devoir (de) (+ *inf.*); être tenu(e) de (+ *inf.*); devoir (qch. abstrait, affectif ou sentimental) à (qn) ▷ 法院~设定隔离的具体时限。Le tribunal doit fixer une durée précise pour l'isolement. || 用人单位与劳动者~履行劳动合同约定的义务。L'employeur et le travailleur doivent exécuter les obligations stipulées dans le contrat de travail. || ~遵守官方文件的标准。Il est nécessaire de respecter les normes des documents officiels. ❷ **[PROBABILITÉ]** devoir (+ *inf.*); il faut (+ *inf.*; que + *subj.*); probablement; apparemment; il est probable que; pouvoir (+ *inf.*) ▷ 现在已经很晚了，她~在家。Il est tard, elle devrait être à la maison. || 这个目标，经过努力~可以达到。Nous pourrions atteindre cet objectif après bien des efforts. • ≈应该

REM «应该» et «应当» sont similaires sémantiquement, mais «应该» est utilisé plus fréquemment dans la pratique, «应当» est plus formel et plus fort dans le degré d'obligation, souvent utilisé dans les documents juridiques, les normes sociales et les règlements. La fréquence de «应当» exprimant la probabilité est très faible, beaucoup plus basse que celle de «应该».

Voir aussi: bilan des verbes auxiliaires modaux polysémiques dans les annexes

要 yào 动 [aux. modal] ❶ **[VOLONTÉ]** vouloir (fort) (+ *inf.*); désirer (+ *inf.*); entendre (+ *inf.*); souhaiter (+ *inf.*); tenir à (+ *inf.*); prétendre (+ *inf.*) ▷ 孩子们要学游泳。Les enfants veulent apprendre à nager. || 我要和你谈谈。Je désire m'entretenir avec toi. ❷ **[OBLIGATION]** devoir (+ *inf.*); il faut (+ *inf.*; que + *subj.*); avoir à (+ *inf.*); il convient de (+ *inf.*); il est nécessaire de (+ *inf.*); se devoir (de) (+ *inf.*); être tenu(e) de (+ *inf.*) ▷ 我们一定要时刻关心群众的生活。Nous devons [sommes tenus de] se préoccuper constamment des conditions de vie des masses. || 学习要刻苦认真。L'étude exige une application persévérante. | Il faut faire les études avec assiduité. || 要加强不同文明之间的交流。Il convient de renforcer les échanges entre les différentes civilisations. ❸ **[NÉCESSITÉ]** devoir (+ *inf.*); il faut (+ *inf.*; que + *subj.*); [Futur] aller (+ *inf.*); être sur le point de (+ *inf.*) ▷ 他今晚要走，所以现在来看你。Devant partir ce soir, il vient te voir maintenant. || 要下雨了。Il va pleuvoir. | Il pleuvra. | Il faut pleuvoir. || 火车要开了。Le train est sur le point de partir. || [Signifiant l'évaluation dans une comparaison] 她比我要高一些。Elle est plus grande que moi.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

REM Bien que «应该» et «要» soient similaires sur la signification d'obligation, leur usage est légèrement différent: «一定», «必须», «得» sont compatibles avec «要» mais ne vont pas avec «应该». Par exemple, 我们一定要（≠应该）时刻关心群众的生活。

Voir aussi: bilan des verbes auxiliaires modaux polysémiques dans les annexes

会（會） huì 动 [aux. modal] ❶ **[COMPÉTENCE]** savoir (+ *inf.*); être capable de (+ *inf.*); être habile à (+ *inf.*; qch.); être expert(e) en [dans] (qch.); être fort(e) en (qch.) ▷ 我不会游泳。Je ne sais pas nager. (= Je n'ai pas appris à nager) || 售货员能说会道。Le vendeur sait bien manier la parole. | Le vendeur a la langue agile. || 她很会画画。Elle est forte en peinture. ❷ **[PROMESSE]** promettre (à qn) de (+ *inf.*) ▷ 我不会离开您的。Je promets de ne pas vous quitter. ❸ **[POSSIBILITÉ↔NÉCESSITÉ]** il se peut [pourrait] que (+ *subj.*); être susceptible de (+ *inf.*); devoir (+ *inf.*); aller (+ *inf.*) ▷ 她会不会来？-可能她不会来。Viendra-t-elle? — Il se peut qu'elle ne vienne pas. | Elle ne viendrait pas. || 你在那儿会冷的。Tu dois avoir froid là-bas. 我们都会参加这场会议。Nous allons tous assister à cette réunion. | Nous assisterons tous à cette réunion.
Voir aussi: bilan des verbes auxiliaires modaux polysémiques dans les annexes

能 néng 动 [aux. modal] ❶ **[CAPACITÉ]** pouvoir (+ *inf.*); être capable de (+ *inf.*); être en mesure de (+ *inf.*); permettre (à qn) de (+ *inf.*); avoir la possibilité de (+ *inf.*); être susceptible de (+ *inf.*); être apte à (+ *inf.*; qch.) ▷ 他好多了，能下床了。Il va beaucoup mieux et peut se lever maintenant. || 大厅能容纳数百人。La salle peut contenir des centaines de personnes. || 他一分钟能打150个字。Il peut taper 150 mots par minute. || 生态创新能促进经济健康发展。L'éco-innovation permet de promouvoir un développement économique sain. || 这种药能治支气管炎吗？Ce médicament est-il susceptible de guérir la bronchite? ❷ **[PERMISSION]** pouvoir (+ *inf.*); permettre (à qn) de (+ *inf.*); avoir le droit de (+ *inf.*); être autorisé(e) à (+ *inf.*; qch.); avoir la permission de (+ *inf.*) ▷ 我能不能借用您的词典？Puis-je emprunter votre dictionnaire? || 这儿能停车吗？-能停。/不能停。Il est permis de se garer ici? | Pouvons-nous nous garer ici? — Oui. / Non. Il est interdit de s'y garer. | Aucun stationnement! || 学生们不能在课间休息的时候外出。Les élèves ne sont pas autorisés à sortir pendant l'interclasse. ❸ **[POSSIBILITÉ]** pouvoir (+ *inf.*); risquer de (+ *inf.*); avoir la possibilité de (+ *inf.*); être susceptible de (+ *inf.*); il se peut [pourrait] que (+ *subj.*); peut-être ▷ 今晚能下雨吗？Pourrait-il pleuvoir ce soir? || 如果能去的话我就会去。J'y irai si j'en ai la possibilité. ❹ **[pouvoir (+ *inf.*) entre 不...不 pour exprimer l'obligation]** ▷ 我当时不能不接受。Je n'ai pas pu faire autrement que d'accepter. | J'étais dans l'obligation d'accepter. | Je ne pouvais pas refuser.

REM «能» et «会» indiquent tous la capacité ou la compétence du sujet, mais en fait la portée de capacité de «能» est plus large que celle de «会»: «能» peut indiquer la capacité ou la possibilité lorsque les conditions le permettent, l'usage et une certaine efficacité que «会» ne peut pas exprimer. De plus, «会» se penche sur une capacité constante, une compétence et une technique acquise, mais «能» marque souvent une capacité dans la contexte spécifique. Apprendre une action ou une technique pour la première fois, «能» et «会» sont tous possibles, mais «会» est utilisé plus souvent; retrouver une certaine capacité, utilisons «能» mais pas «会». Par exemple, 他好多了，能（≠会）下床了。

Voir aussi: bilan des verbes auxiliaires modaux polysémiques dans les annexes

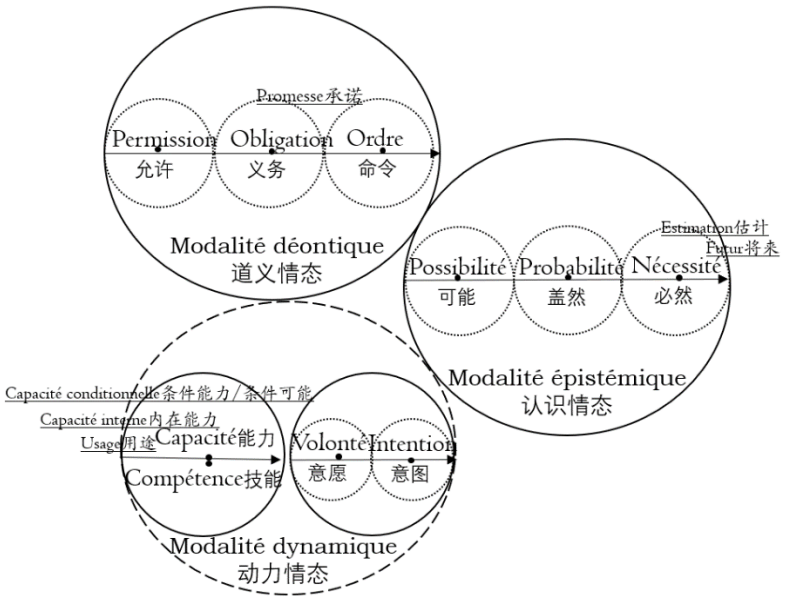
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

En plus des définitions, les utilisateurs peuvent trouver, dans les annexes du dictionnaire chinois–français, le bilan des verbes auxiliaires modaux polysémiques, y compris les caractéristiques sémantiques et syntaxiques, ainsi que les tableaux du vocabulaire chinois–français correspondant à chaque type modal. Les caractéristiques permettent aux apprenants d'acquérir une connaissance plus systématique de l'ensemble de cette catégorie, tandis que les tableaux aident les utilisateurs à trouver facilement les sens modaux et les expressions correspondantes en chinois–français, ce qui permet de relier les trois éléments et de remarquer en même temps leur nuances de signification et d'intensité au sein d'une même modalité. De plus, il est nécessaire de rappeler que nos tableaux ne sont basés que sur cinq verbes auxiliaires modaux polysémiques. Ainsi, l'analyse sémantique plus complète et plus approfondie des verbes auxiliaires modaux reste à réaliser à l'avenir dans le cadre d'étude plus vaste avec les identités lexicales mieux définies.

多义情态助动词小结
Bilan des verbes auxiliaires modaux polysémiques

1. 语义和句法特点 Caractéristiques sémantiques et syntaxiques

典型多义情态助动词（应该、应当、要、会、能）语义特点
Caractéristiques sémantiques des verbes auxiliaires modaux polysémiques
(Cas typiques: «应该», «应当», «要», «会», «能»)



汉语情态助动词句法特点 (cf. Song 2007: 38-39)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

Caractéristiques syntaxiques des verbes auxiliaires modaux chinois

(1) 只能带谓词宾语，不能带体词宾语。即在“主语+情态动词+动词（形容词）”的形式里出现。Ils sont compatibles seulement avec un régime prédicatif et excluent absolument un régime nominal. C'est-à-dire qu'ils apparaissent sous la forme «sujet + verbe auxiliaire modal + verbe (adjectif)».
这个大厅**能容纳**1500人。Cette salle peut contenir quinze cents personnes.
*这个大厅**能**1500人。
他比我**要高**一些。Il est plus grand que moi.

(2) 可以在“X不X”形式里出现，表示疑问。Ils peuvent apparaître sous la forme de «X 不 X» pour poser les questions.
她**会不会**来？Viendra-t-elle?
我**应不应该**原谅他？Dois-je lui pardonner?
我**能不能**借用您的词典？Puis-je emprunter votre dictionnaire?

(3) 不能重叠。Ils ne peuvent pas se répéter.
我要**学学**跳舞。Je veux apprendre à danser.
*我**要要**学跳舞。

(4) 不能带体标记“了”、“着”、“过”。Ils ne sont pas compatibles avec les particules aspectuelles «了», «着», «过».
她**应该吃了**饭。Elle aurait dû manger.
*她**应该了**吃饭。

(5) “情态动词+趋向补语（来、去）”的形式不成立。Ils ne sont pas compatibles avec les régimes directionnels («来», «去»).
他们**上山去**了。Ils sont allés sur la colline.
*他们**应该去**了。（此处“去”是动词而不是趋向补语。Ici «去» est un verbe plutôt qu'un régime directionnel.）

2. 情态类型汉法词汇对应表 Tableaux du vocabulaire chinois–français correspondant à chaque type modal

<table>
<tr><td colspan="4">道义情态
Modalité déontique</td></tr>
<tr><td>允许
Permission</td><td colspan="2">义务
Obligation</td><td>命令
Ordre</td></tr>
<tr><td rowspan="2">能❷</td><td rowspan="2">要❷
应该❶、应当❶</td><td>承诺
Promesse</td><td rowspan="2">…</td></tr>
<tr><td>会❷</td></tr>
<tr><td>pouvoir (+ inf.)
permettre (à qn) de (+ inf.)
avoir le droit de (+ inf.)
être autorisé(e) à (+ inf.; qch.)
avoir la permission de (+ inf.)…</td><td colspan="2">devoir (+ inf.)
il faut (+ inf.; que + subj.)
avoir à (+ inf.)
il convient de (+ inf.)
il est nécessaire de (+ inf.)
se devoir (de) (+ inf.)
être tenu(e) de (+ inf.)
devoir (qch. abstrait, affectif ou sentimental) à (qn)…</td><td>promettre (à qn) de (+ inf.)…</td><td>…</td></tr>
</table>

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

认识情态 Modalité épistémique				
可能 Possibilité	盖然 Probabilité	必然 Nécessité		
会❸ 能❸	会❸ 应该❷、应当❷	会❸ 要❸	将来 Futur	估计 Estimation
			要❸	要❸
pouvoir (+ inf.) risquer de (+ inf.) avoir la possibilité de (+ inf.) être susceptible de (+ inf.) il se peut [pourrait] que (+ subj.) peut-être *conditionnel*…	devoir (+ inf.) il faut (+ inf.; que + subj.) probablement apparemment il est probable que pouvoir (+ inf.)…	devoir (+ inf.) il faut (+ inf.; que + subj.) aller (+ inf.)…	aller (+ inf.) être sur le point de (+ inf.) *futur simple*…	[signifiant l'évaluation dans une comparaison]…

动力情态 Modalité dynamique					
能力 Capacité			技能 Compétence	（强）意愿 (Forte) Volonté	意图 Intention
条件能力/条件可能 Capacité conditionnelle	内在能力 Capacité interne	用途 Usage			
能❶	能❶	能❶	会❶	要❶	…
pouvoir (+ inf.) permettre (à qn) de (+ inf.) être capable de (+ inf.) être en mesure de (+ inf.) avoir la possibilité de (+ inf.) être susceptible de (+ inf.) être apte à (+ inf.; qch.)…	pouvoir (+ inf.) être capable de (+ inf.) être en mesure de (+ inf.) être apte à (+ inf.; qch.)…	pouvoir (+ inf.) permettre (à qn) de (+ inf.) être capable de (+ inf.) être en mesure de (+ inf.) avoir la possibilité de (+ inf.) être susceptible de (+ inf.) être apte à (+ inf.; qch.)…	savoir (+ inf.) être capable de (+ inf.) être habile à (+ inf.; qch.) être expert(e) en [dans] (qch.) être fort(e) en (qch.)…	vouloir (fort) (+ inf.) désirer (+ inf.) entendre (+ inf.) souhaiter (+ inf.) tenir à (+ inf.) prétendre (+ inf.)…	…

Au final, la présentation multimodale mérite également notre attention et notre considération au fur et à mesure des progrès techniques (Zhang 2021). Par exemple, les images et les textes peuvent se compléter dans la définition, les figures précédentes sur les verbes auxiliaires modaux polysémiques peuvent même être animées de manière interactive, présentant les équivalents au clic ou à la voix de l'utilisateur.

6. Conclusion

A l'aide des théories des types modaux, du prototype et du schéma, nous avons proposé un mode optimisé et des mesures d'amélioration pour le traitement des verbes auxiliaires modaux polysémiques. Autour des points sémantiques dans les modalités comme des sens prototypiques, nous étendons la signification des verbes auxiliaires modaux polysémiques en construisant un réseau de définition unifié qui nous permettra d'éviter le manque d'informations et l'imprécision de la correspondance au cours de la compilation. Dans l'ensemble, comme com-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

posante langagière universelle mais particulière, la catégorie des verbes auxiliaires modaux dans les dictionnaires chinois–français reste un sujet qui mérite d'être exploré. Nous espérons élargir davantage la portée de l'étude pour tester l'adaptabilité des méthodes de définition, confirmer les erreurs des apprenants et vérifier la validité de la définition optimisée par des enquêtes à l'avenir.

Endnotes

1. Les résultats détaillés de l'enquête sur les avantages et les insuffisances des dictionnaires seront présentés dans un autre article, en raison des contraintes de longueur de cet article.
2. Les mots spécifiques sont notés en fonction de leur degré d'appartenance aux parties du discours. Les typiques sont sélectionnés comme prototypes avec la note 1; plus les mots sont typiques, plus la note est proche de 1.
3. A l'heure actuelle, le *Dictionnaire Chinois Contemporain* a été mis à jour à la septième édition. Par rapport à la cinquième édition, il n'existe pas de différence significative sur les parties du discours des verbes auxiliaires modaux polysémiques *应该(yīnggāi)*、*应当(yīngdāng)*、*要(yào)*、*会(huì)*、*能(néng)* dans la septième édition.
4. Du point de la traductologie, «il existe deux sortes d'équivalents de nature différente dans la traduction, l'une consiste aux équivalents fixes établis à l'avance, et l'autre comprend les équivalents temporaires établis en fonction du contexte spécifique», voir Li et Hu (2015: 75).

Références

Dictionnaires et corpus

Association Ricci du grand dictionnaire français de la langue chinoise — Commercial Press. 2014. *Dictionnaire Ricci Chinois–Français*. Beijing: Commercial Press.

BCC: BLCU Corpus Center. http://bcc.blcu.edu.cn/zh/cid/1

Bureau de rédaction de dictionnaires, Institut de linguistique, Académie chinoise des sciences sociales. 2005. *Dictionnaire Chinois Contemporain.* 5e version. Beijing: Commercial Press.

Huang, J. (dir.). 2014. *Grand Dictionnaire Chinois–Français Contemporain*. Beijing: Foreign Language Teaching and Research Press.

Larousse: Larousse.fr: encyclopédie et dictionnaires gratuits en ligne. https://www.larousse.fr/

Le Robert: Dico en ligne Le Robert. https://dictionnaire.lerobert.com/

Sketch Engine: https://app.sketchengine.eu/

Yao, X. (dir.). 2009. *A Chinese–English Dictionary*. Beijing: Foreign Language Teaching and Research Press.

Autre littérature

Fang, L. et C. Chen. 2008. Modèle d'arrangement des verbes auxiliaires modaux dans le dictionnaire d'apprentissage anglais–chinois — une étude de 'must'. *Lexicographical Studies* 6: 70-76.

Jia, C. 2020. Évolution diachronique du verbe modal polysémique 'neng'. *Journal of Qiqihar University (Philosophy & Social Science Edition)* 12: 121-124+128.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2033 (Article)

Kronning, H. 2001. Pour une tripartition des emplois du modal 'devoir'. Dendale, P. et J. van der Auwera (Éds.). 2001. *Les verbes modaux.* Cahiers Chronos 8: 67-84. Amsterdam/Atlanta: Rodopi.

Labov, W. 1973. The Boundaries of Words and Their Meanings. Baily, C.-J. et R. Shuy (Éds.). 1973. *New Ways of Analyzing Variation in English*: 340-373. Washington, DC: Georgetown University Press.

Lai, P. 2012. Étude de cas longitudinale du processus d'acquisition des modalités radicale et épistémique. *Journal of Yunnan Normal University (Teaching & Studying Chinese as a Foreign Language Edition)* 10(3): 9-17.

Le Querler, N. 2004. Les modalités en français. *Revue belge de Philologie et d'Histoire* 82(3): 643-656.

Li, X. et K. Hu. 2015. Une étude basée sur le corpus de l'interprétabilité de l'interprétation de conférences de presse Chinois–Anglais. *Foreign Language Learning Theory and Practice* 3: 74-79+96-97.

Lin, L. 2019. *Étude sur l'intensité de la modalité chinoise*. Beijing: Social Sciences Academic Press.

Lv, S. 2001. *Huit cents mots en chinois moderne (édition augmentée).* Beijing: Commercial Press.

Nuyts, J. et J. van der Auwera (Éds.). 2016. *The Oxford Handbook of Modality and Mood.* Oxford Handbooks. Oxford: Oxford University Press.

Palmer, F.R. 1990. *Modality and the English Modals.* Second edition. London: Routledge.

Peng, L. 2007. *Étude de la modalité en chinois moderne.* Beijing: China Social Sciences Press.

Rosch, E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104(3): 192-233.

Song, Y. 2007. *Étude sur la négation des verbes modaux en chinois moderne.* Beijing: China Social Sciences Press.

Sueur, J.-P. 1979. Une analyse sémantique des verbes *devoir* et *pouvoir*. *Le français moderne* 47(2): 97-120.

Wang, R. 2006. *Une étude empirique du marquage des parties du discours dans les dictionnaires Chinois–Anglais: une étude cognitive.* Shanghai: Shanghai Translation Publishing House.

Wang, R. et Y. Zhang. 2006. Une étude empirique de l'influence du marquage des parties du discours sur la précision de la traduction dans les dictionnaires Chinois–Anglais. *Modern Foreign Languages* 29(2): 139-146+218-219.

Yang, L. 2017. *Subjectivité et subjectivation des verbes auxiliaires modaux chinois.* Guangzhou: World Publishing Corporation.

Yuan, Y. 1995. Ressemblance familiale de la catégorie des parties du discours. *Social Sciences in China* 1: 154-170.

Yuan, Y. 2010. *Investigation cognitive et classification floue des classes de mots en chinois mandarin.* Shanghai: Shanghai Educational Publishing House.

Zgusta, L. 1971. *Manual of Lexicography*. Prague: Academia.

Zhang, Y. 2009. *Sémantique, Cognition, Définition*. Shanghai: Shanghai Foreign Language Education Press.

Zhang, Y. 2017. Relation lexicale et modes de définition basés sur les théories des prototypes et des catégories de schéma: Un cas de définition basée sur la catégorisation. *Lexicographical Studies* 5: 1-12+93.

Zhang, Y. 2021. Conception du design des textes multimodaux-lexicographiques du point de vue des médias convergents. *Lexicographical Studies* 2: 20-32+125.

Zhang, Y. et H. Yong. 2007. *Lexicographie contemporaine*. Beijing: Commercial Press.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

Prompts for Language Learners: A Practical Guide to Using DeepSeek as a Dictionary

Ai Zhong, *College of Foreign Languages and Literature, Fudan University, Shanghai, China; and College of Foreign Languages, Shanghai Maritime University, Shanghai, China (zhongai@shmtu.edu.cn) (https://orcid.org/0000-0003-3882-7129)*

Abstract: Nowadays, AI models are capable of generating lexicographic content that is as good as that produced by traditional dictionaries. Meanwhile, using different prompts can return disparate results. Considering this, this article tests several prompts to demonstrate to language learners how to use a generative AI model as a dictionary. From a practical point of view, learners should improve their prompt crafting and fine-tuning skills in order to obtain well-reasoned and well-structured outputs from AI.

Keywords: GENERATIVE AI, LANGUAGE MODEL, DEEPSEEK, LEARNER'S DICTIONARY, LEXICOGRAPHIC PRODUCTIVITY, PROMPT ENGINEERING

Opsomming: Prompe vir taalleerders: 'n Praktiese gids tot die gebruik van DeepSeek as 'n woordeboek. KI-modelle is deesdae daartoe in staat om leksikografiese inhoud te genereer wat net so goed is soos dié wat deur tradisionele woordeboeke geproduseer word. Terselfdertyd kan uiteenlopende resultate verkry word deur verskillende prompe te gebruik. Met inagneming hiervan word daar in hierdie artikel verskillende prompe getoets om aan taalleerders te demonstreer hoe om 'n generatiewe KI-model as 'n woordeboek te gebruik. Vanuit 'n praktiese oogpunt behoort leerders hul vaardighede in die skep en verfyn van prompe te verbeter om goed beredeneerde en goed gestruktureerde uitsette van KI te verkry.

Sleutelwoorde: GENERATIEWE KI, TAALMODEL, DEEPSEEK, AANLEERDERSWOORDEBOEK, LEKSIKOGRAFIESE PRODUKTIWITEIT, PROMPONTWIKKELING

1. Introduction

Artificial Intelligence (AI) is not a new concept. However, it was not until the end of 2022 when OpenAI launched ChatGPT that AI had gradually become more involved in the field of lexicography. A couple of studies have explored the possibility of using AI to create dictionary contents. Among the first batch of studies at the interface of lexicography and AI are some detailed comparisons between AI and traditional dictionaries (e.g. Lew 2024, Lew et al. 2024, Ptasznik et al. 2024, Rees and Lew 2024). The results are quite similar: generative AI is capable of pro-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

ducing high-quality contents such as dictionary entries that include definitions and examples. For the moment, AI models outperform both monolingual and bilingual dictionaries in productive lexical tasks (cf. Lew et al. 2024, Ptasznik et al. 2024). Comparisons are also made between different AI models as support tools for language learners in different lexical tasks (cf. Huete-García and Tarp 2024, Li and Tarp 2024). While AI can be used as a useful lexical tool in a range of tasks, its shortcomings have been discussed in several works (e.g. De Schryver 2023, Lew 2023, Fuertes-Olivera 2024).

At the same time, studies have shown that the results generated by AI depend largely on its interaction with users. Users' prompts (i.e. instructions and questions given to an AI model) guide AI to produce relevant outputs, and users, therefore, have to improve and refine their prompts to achieve better results (cf. Lew 2023, Fuertes-Olivera 2024, Li and Tarp 2024, Tarp and Nomdedeu-Rull 2024). In this short article, several prompts instructing AI to undertake some productive lexical tasks will be examined. Rather than evaluating the quality and effectiveness of these prompts, the aims of the present work are fairly practical: to explain how to craft prompts to obtain lexicographic contents; how to modify prompts to achieve more relevant and accurate results; and how to fine-tune prompts to return well-written and well-structured outputs.

2. Materials and methods

For the present study, ten target headwords were randomly selected by choosing the 'Word of the Day' in April from four well-known learner's dictionary sites, namely *Cambridge Advanced Learner's Dictionary & Thesaurus* (henceforth CALD), *Collins COBUILD Advanced Learner's Dictionary* (COBUILD), *Longman Dictionary of Contemporary English* (LDOCE), and *Oxford Advanced Learner's Dictionary* (OALD). Among the ten lexical items are three nouns (*feng shui*, *sunflower*, and *frontispiece*), an adverb (*consciously*), a phrase (*have seen better days*), and five words with more than one part of speech (*elaborate*, *highlight*, *dizzy*, *so-so*, and *marathon*).

Before using an AI chatbot to create lexicographic contents, the ten items are also looked up in the abovementioned learner's dictionaries. Taking a quick glimpse at the entries in these dictionaries, it becomes evident that some principal entry components are the headword, part of speech, pronunciation, definition, examples, and word forms. Other components appearing in learner's dictionaries include etymology, a frequency marker, related words, synonyms, antonyms, collocations, topics, usage notes, and an illustration, depending on each dictionary's microstructure design.

The AI contents were generated between the 23rd and 27th of April 2025. The AI model used was the free version of DeepSeek, a Chinese open-source AI model that has grown in popularity. Compared to its AI rivals mainly from the US, DeepSeek ensures accessibility to Chinese users and excels in both Chinese and English proficiency. Furthermore, DeepSeek has been integrated into the majority of Chinese universities' online platforms. The following section details the tests of crafting and modifying prompts for creating lexicographic contents.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

3. Prompts for use

Prompt #1: Define 'X'.
The very first instruction given to DeepSeek simply used an imperative tone to ask the AI model to define a lexical item 'X', with X being a word, phrase, or even morpheme. The command 'define …' works very well in search engines such as Google and Microsoft Bing, and will return results containing data licensed from a traditional, well-known publisher's dictionary (in these two cases, that of Oxford). However, if Prompt #1 is used with DeepSeek, the responses generated by the AI are quite far from standard dictionary entries, though they do provide a lot of useful information. An example is demonstrated in Figure 1.

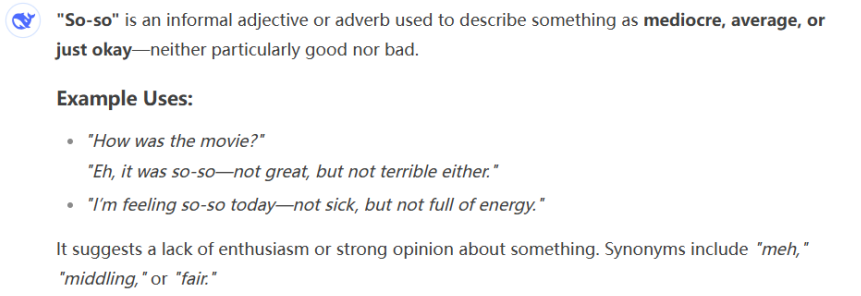
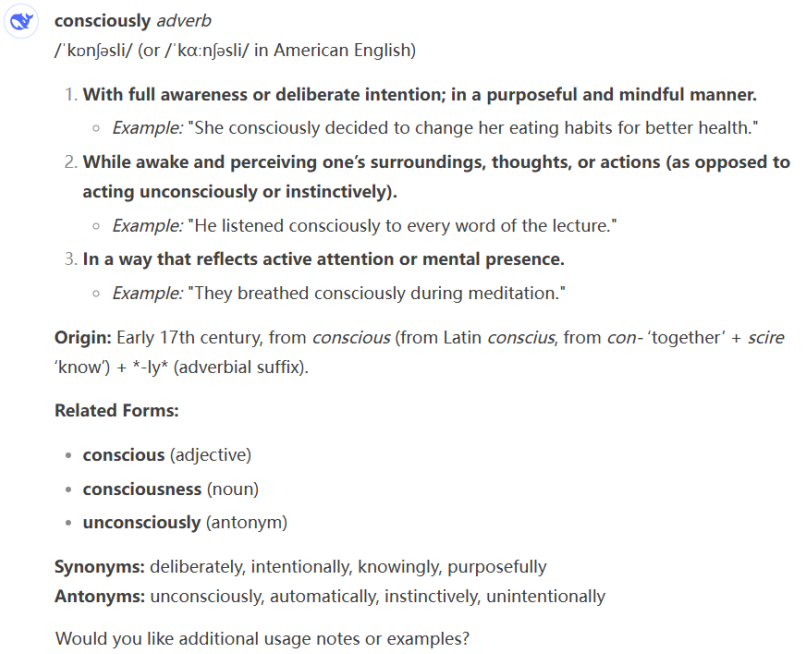


Figure 1: DeepSeek's output for Prompt #1: Define 'so-so'

As evident in Figure 1, DeepSeek's response starts with a sentence rather than a headword, and all of the outputs for the ten test items are formatted in this way. Some lexical information is given in the case of *so-so*, such as parts of speech, an explanation of meaning, three example sentences, and synonyms. But for some words, especially nouns, DeepSeek mixes linguistic and extra-linguistic data in the answers, as this is not specified in the prompt. For example, the responses produced by DeepSeek illustrate some core principles of *feng shui* (e.g. qi, yin and yang, the five elements, the bagua map, and the commanding position), key features of *marathon* (e.g. the distance, types, training, and major marathons), and likewise key characteristics of *sunflower* (e.g. appearance, heliotropism, seeds, cultivation, and symbolism). These outputs of Prompt #1 are very similar to contents provided by an encyclopaedia rather than a dictionary, and the prompt therefore has to be refined to achieve precise and relevant responses, with a particular focus on lexicographic data.

Prompt #2: Define 'X' like a dictionary entry.
Based on Prompt #1, the current prompt is modified by adding a constraint to command DeepSeek to produce dictionary entry-like contents, as shown in Figure 2.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)



consciously *adverb*
/ˈkɒnʃəsli/ (or /ˈkɑːnʃəsli/ in American English)

1. **With full awareness or deliberate intention; in a purposeful and mindful manner.**
   - *Example:* "She consciously decided to change her eating habits for better health."
2. **While awake and perceiving one's surroundings, thoughts, or actions (as opposed to acting unconsciously or instinctively).**
   - *Example:* "He listened consciously to every word of the lecture."
3. **In a way that reflects active attention or mental presence.**
   - *Example:* "They breathed consciously during meditation."

Origin: Early 17th century, from *conscious* (from Latin *conscius*, from *con-* 'together' + *scire* 'know') + *-ly* (adverbial suffix).

Related Forms:

- **conscious** (adjective)
- **consciousness** (noun)
- **unconsciously** (antonym)

Synonyms: deliberately, intentionally, knowingly, purposefully
Antonyms: unconsciously, automatically, instinctively, unintentionally

Would you like additional usage notes or examples?

Figure 2: DeepSeek's output for Prompt #2: Define 'consciously' like a dictionary entry

This time, the output generated by DeepSeek looks very similar to a dictionary entry in terms of format. The headword is given in bold, with its part of speech next to it. All senses are numbered. Several other typical entry components are also provided, including the pronunciation, examples, the word origin, related forms, synonyms, and antonyms. However, the structure of DeepSeek's responses is rather fluid. For instance, the parts of speech for some words are placed in brackets, while for others, they are not, and the origin of *feng shui* appears above its definition. Section names are not consistent either, such as *etymology/origin*, *related terms/forms/see also*, and *examples/example sentences*. Meanwhile, some default entry sections are missing in several responses — for example, the pronunciations of *highlight*, *marathon*, and *frontispiece* are not provided. Table 1 displays the lexicographic components included in DeepSeek's generated answers.

Table 1: Components in DeepSeek's responses for Prompt #2

Headwords	Included components
consciously	part of speech, pronunciation, definition, examples, origin, related terms, synonyms, antonyms
dizzy	part of speech, pronunciation, definition, examples, origin, synonyms, antonyms

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

elaborate	part of speech, pronunciation, definition, examples, synonyms, antonyms, origin, usage notes
feng shui	part of speech, pronunciation, origin, definition, examples, related terms
frontispiece	part of speech, definition, examples, origin, related terms
have seen better days	part of speech, definition, usage notes, examples, synonyms, origin
highlight	part of speech, definition, examples, origin, synonyms, antonyms
marathon	part of speech, definition, examples, origin, related terms
so-so	part of speech, pronunciation, definition, synonyms, antonyms, examples, origin, usage notes
sunflower	part of speech, pronunciation, definition, origin, examples, related terms

Whereas some standard entry sections are excluded in the answers, DeepSeek (and many other AI models) sometimes generates follow-up questions to provide further information to users. For the chosen words, six out of ten responses contain this type of questions, for example: 'Would you like a pronunciation guide or additional examples?', 'Would you like additional usage notes or related terms?', and 'Would you like additional variations or cultural references?'. Users can simply reply 'Yes' to continue the conversation with DeepSeek. If users want specific information, they have to give detailed instructions in another prompt.

At this point, the languages used in the interaction with AI are noteworthy. In fact, DeepSeek is pre-trained on large datasets in both English and Chinese. If the input is in English, the output by AI will also be in English; likewise, if the instruction is in Chinese, the response will be in Chinese as well. This is important to note as learners may also want to use DeepSeek as a bilingual dictionary. In such case, Prompt #2 can be slightly changed into 'Define "X" like a bilingual dictionary entry'. But again, the user's choice of language matters. Through several tests it became clear that, if the prompt is written in English, the default result will be an English–Spanish entry. Only if a Chinese language instruction is given can DeepSeek understand that the response should be a Chinese–English-bilingual-dictionary-style entry.

Prompt #3: Define 'X' like a dictionary entry. Include sections like Y_1, Y_2, Y_3, …
If the output of Prompt #2 does not meet expectations, the prompt can be further refined by providing specific instructions. At this stage, DeepSeek is asked to include relevant information from various areas, whether lexical, grammatical, semantic, pragmatic, or even cultural. Then 'Y' in the prompt can be replaced with a range of sections — some are central while others are peripheral (cf. Atkins and Rundell 2008) — that may appear in a real dictionary, as listed in Table 2.

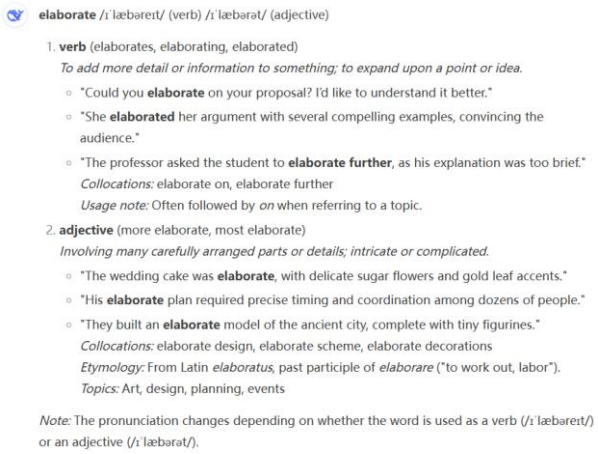
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

Table 2: Possible entry components (represented by 'Y') in Prompt #3

Central	Peripheral
headword	pronunciation
part of speech	etymology
definition	word forms
example	related terms
	frequency
	synonyms and antonyms
	collocations
	idioms
	topics
	usage notes
	illustration

To ensure a well-structured AI 'entry', a good example of a prompt can be found in the addendum of De Schryver's article (2023). A modified version of prompt for this study is demonstrated below (called Prompt #3 Plus), followed by its response shown in Figure 3:

> Define 'elaborate' like a dictionary entry: Start with the headword in bold and its pronunciation at the top. Each sense should be in a numbered block. Each block then starts with the part of speech and the morphological forms of the respective sense. This is followed by a sense definition and sense examples that illustrate both the use and the meaning of each particular sense. For the example sentences, make sure to use different sentence structures, referring to different people; vary long and short example sentences. Include other information if relevant, such as etymology, collocations, topics, and usage notes.



elaborate /ɪˈlæbəreɪt/ (verb) /ɪˈlæbərət/ (adjective)

1. **verb** (elaborates, elaborating, elaborated)
   To add more detail or information to something; to expand upon a point or idea.
   - "Could you **elaborate** on your proposal? I'd like to understand it better."
   - "She **elaborated** her argument with several compelling examples, convincing the audience."
   - "The professor asked the student to **elaborate further**, as his explanation was too brief."

   Collocations: elaborate on, elaborate further
   Usage note: Often followed by *on* when referring to a topic.
2. **adjective** (more elaborate, most elaborate)
   Involving many carefully arranged parts or details; intricate or complicated.
   - "The wedding cake was **elaborate**, with delicate sugar flowers and gold leaf accents."
   - "His **elaborate** plan required precise timing and coordination among dozens of people."
   - "They built an **elaborate** model of the ancient city, complete with tiny figurines."

   Collocations: elaborate design, elaborate scheme, elaborate decorations
   Etymology: From Latin *elaboratus*, past participle of *elaborare* ("to work out, labor").
   Topics: Art, design, planning, events

Note: The pronunciation changes depending on whether the word is used as a verb (/ɪˈlæbəreɪt/) or an adjective (/ɪˈlæbərət/).

Figure 3: DeepSeek's output for Prompt #3 Plus

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

Prompt #4: Give me a Z-style dictionary entry for 'X'.
Rather than listing all possible components in the prompt, a more direct approach is to ask AI models, which are normally trained on publicly accessible data, to imitate the entry from those well-known dictionaries, represented by 'Z' in Prompt #4. Figures 4 and 5 illustrate the comparison between DeepSeek-generated content and that of real dictionary entries.

As the figures show, DeepSeek's responses, to a large extent, follow OALD's and LDOCE's respective formats, such as the word's opposite in OALD and the frequency marker in LDOCE. Furthermore, DeepSeek concludes several major features of the named dictionary at the end of its answer, seen in the final sentences in Figure 4 and the bullet points in Figure 5. In addition to obtaining lexicographic information, learners can use DeepSeek to compare the characteristics of dictionary products in the market by employing this prompt.

Prompt #5: Here is an example entry. Produce a similar entry for 'X'.
So far, the prompts crafted are all zero-shot ones, that is, in the field of AI, prompts without examples. On the contrary, few-shot prompts refer to those containing some examples (or called one-shot if there is only one example). A study by DeepSeek-AI (2025) recommends that users use zero-shot prompting rather than few-shot, as the latter may degrade performance. However, in this article, few-shot prompts are still tested to examine whether they really produce worse results than the prompts above. Lew (2023) has tailored an impressive example of few-shot prompting, in which ChatGPT is presented with two example entries from COBUILD. Since Lew's study is primarily from a lexicographer's perspective and his prompt may be too long for learners, the instruction has been simplified to Prompt #5 for the purpose of this study. Before the result of using this prompt is considered, the DeepSeek-generated responses to the inquiry about producing an entry for *frontispiece*, using Prompt #1, Prompt #2, and Prompt #4, are displayed in Figure 6.

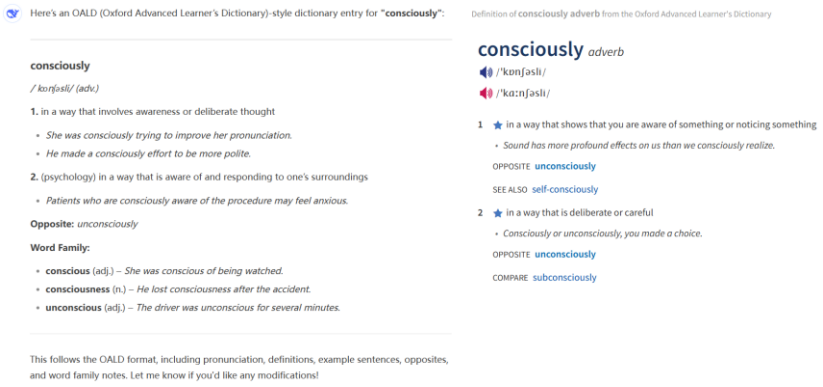
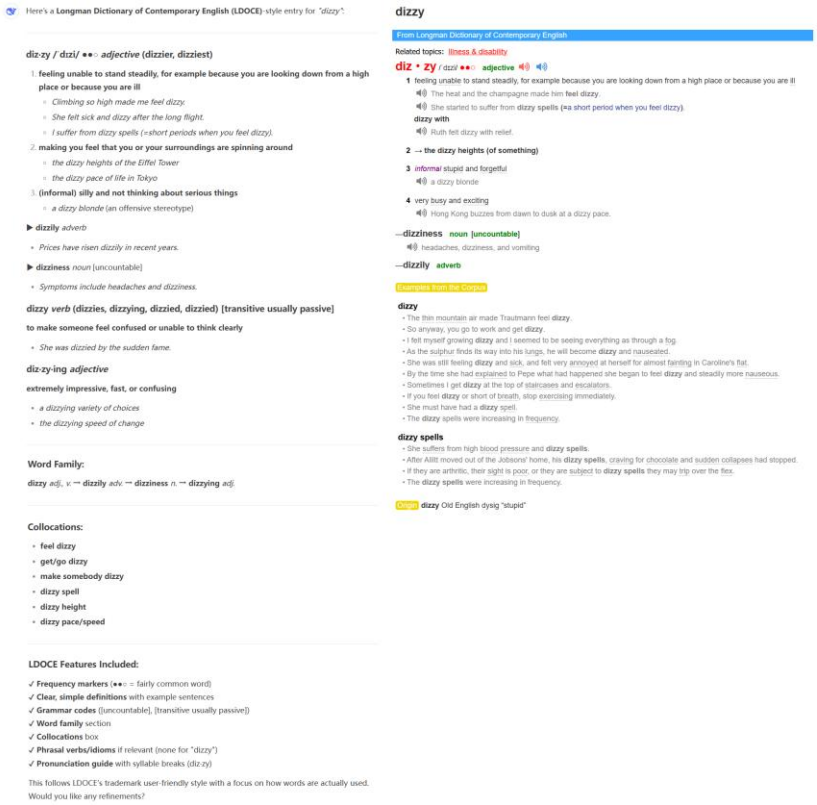


Figure 4: DeepSeek's output for Prompt #4: Give me an OALD-style dictionary entry for 'consciously' vs. OALD's entry for 'consciously'.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)



Here's a **Longman Dictionary of Contemporary English (LDOCE)**-style entry for *"dizzy"*:

diz-zy / dɪzi/ ●●○ *adjective* (dizzier, dizziest)

1. **feeling unable to stand steadily, for example because you are looking down from a high place or because you are ill**
   - *Climbing so high made me feel dizzy.*
   - *She felt sick and dizzy after the long flight.*
   - *I suffer from dizzy spells (=short periods when you feel dizzy).*
2. **making you feel that you or your surroundings are spinning around**
   - *the dizzy heights of the Eiffel Tower*
   - *the dizzy pace of life in Tokyo*
3. **(informal) silly and not thinking about serious things**
   - *a dizzy blonde* (an offensive stereotype)

▶ **dizzily** *adverb*

- *Prices have risen dizzily in recent years.*

▶ **dizziness** *noun* [uncountable]

- *Symptoms include headaches and dizziness.*

dizzy *verb* **(dizzies, dizzying, dizzied, dizzied) [transitive usually passive]**

to make someone feel confused or unable to think clearly

- *She was dizzied by the sudden fame.*

diz-zy-ing *adjective*

extremely impressive, fast, or confusing

- *a dizzying variety of choices*
- *the dizzying speed of change*

Word Family:

dizzy *adj., v.* → **dizzily** *adv.* → **dizziness** *n.* → **dizzying** *adj.*

Collocations:

- **feel dizzy**
- **get/go dizzy**
- **make somebody dizzy**
- **dizzy spell**
- **dizzy height**
- **dizzy pace/speed**

LDOCE Features Included:

✓ **Frequency markers** (●●○ = fairly common word)
✓ **Clear, simple definitions** with example sentences
✓ **Grammar codes** ([uncountable], [transitive usually passive])
✓ **Word family** section
✓ **Collocations** box
✓ **Phrasal verbs/idioms** if relevant (none for "dizzy")
✓ **Pronunciation guide** with syllable breaks (diz·zy)

This follows LDOCE's trademark user-friendly style with a focus on how words are actually used.
Would you like any refinements?

dizzy

From Longman Dictionary of Contemporary English

Related topics: Illness & disability

diz • zy /ˈdɪzi/ ●●○ **adjective**

1 feeling unable to stand steadily, for example because you are looking down from a high place or because you are ill
The heat and the champagne made him **feel dizzy**.
She started to suffer from **dizzy spells** (=a short period when you feel dizzy).
dizzy with
Ruth felt dizzy with relief.

2 → **the dizzy heights (of something)**

3 informal stupid and forgetful
a dizzy blonde

4 very busy and exciting
Hong Kong buzzes from dawn to dusk at a dizzy pace.

—dizziness noun [uncountable]
headaches, dizziness, and vomiting

—dizzily adverb

Examples from the Corpus

dizzy
- The thin mountain air made Trautmann feel **dizzy**.
- So anyway, you go to work and get **dizzy**.
- I felt myself growing **dizzy** and I seemed to be seeing everything as through a fog.
- As the sulphur finds its way into his lungs, he will become **dizzy** and nauseated.
- She was still feeling **dizzy** and sick, and felt very annoyed at herself for almost fainting in Caroline's flat.
- By the time she had explained to Pepe what had happened she began to feel **dizzy** and steadily more nauseous.
- Sometimes I get **dizzy** at the top of staircases and escalators.
- If you feel **dizzy** or short of breath, stop exercising immediately.
- She must have had a **dizzy** spell.
- The **dizzy** spells were increasing in frequency.

dizzy spells
- She suffers from high blood pressure and **dizzy spells**.
- After Alitt moved out of the Jobsons' home, his **dizzy spells**, craving for chocolate and sudden collapses had stopped.
- If they are arthritic, their sight is poor, or they are subject to **dizzy spells** they may trip over the flex.
- The **dizzy spells** were increasing in frequency.

Origin **dizzy** Old English *dysig* "stupid"

Figure 5: DeepSeek's output for Prompt #4: Give me an LDOCE-style dictionary entry for 'dizzy' vs. LDOCE's entry for 'dizzy'.

DeepSeek's answer to Prompt #1 provides a lengthy definition of the word — if that can be called a 'definition' — and no other lexical information. The answer to Prompt #2 is much better, with several major entry components, though it lacks the pronunciation and only offers one example for each sense. Similarly, the AI's response to Prompt #4 is not as satisfactory as those in Figures 4 and 5. Based on numerous tests, AI's output changes every time and its format is quite fluid if not specified in the prompt. Therefore, few-shot prompting gives examples to the AI model and ensures users the output in their preferred format. Figure 7 illustrates a conversation with DeepSeek on the topic of producing an entry for *frontispiece*. Firstly, a screenshot of *dizzy* from LDOCE (see the right side of Figure 5) is given to DeepSeek. Then Prompt #5 asks the AI model to follow the instruction to carry out the lexical task. At first glance, the output format is much better than that in Figure 6. Although the response is slightly flawed (e.g. the play symbols are replaced and the frequency marker is unchecked), it is well-reasoned in many ways, suggesting that few-shot prompting is still effective in such a task.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

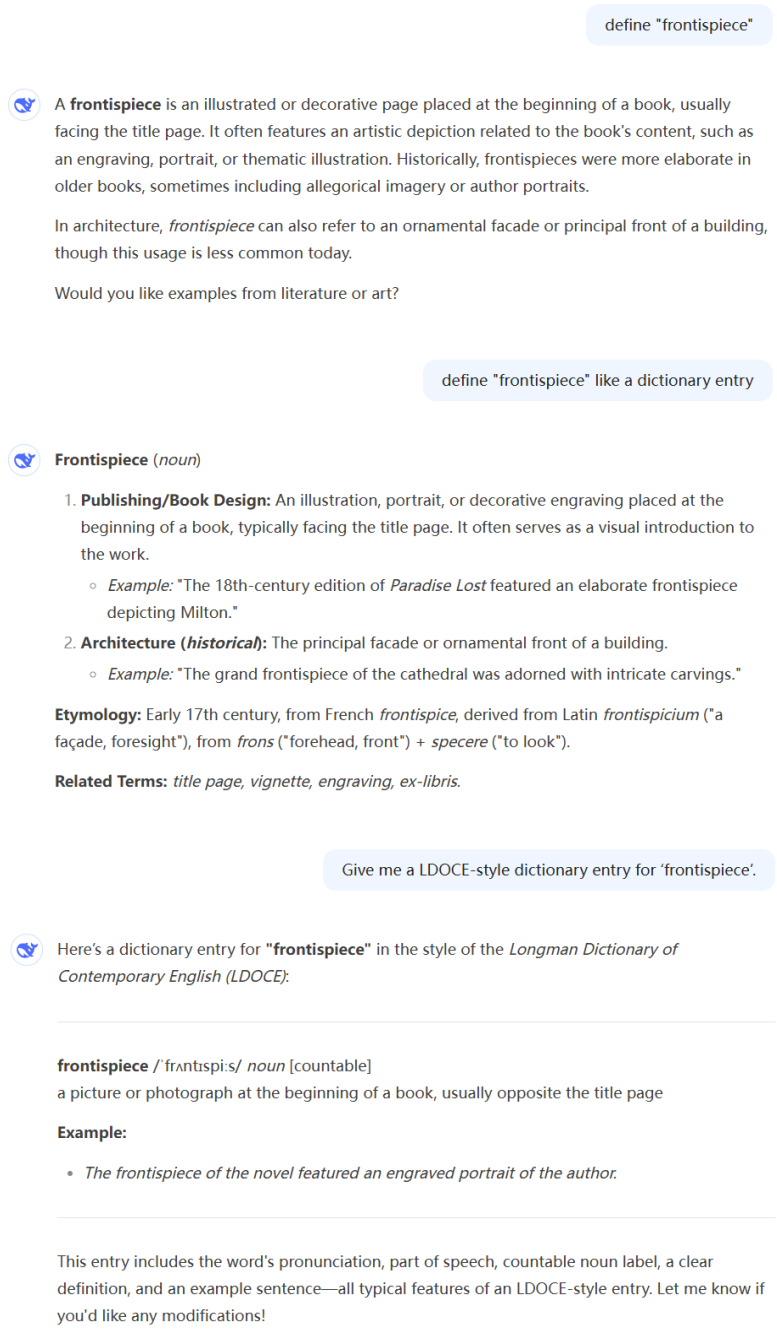


Figure 6: DeepSeek's outputs for Prompt #1, Prompt #2, and Prompt #4 for the word *frontispiece*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

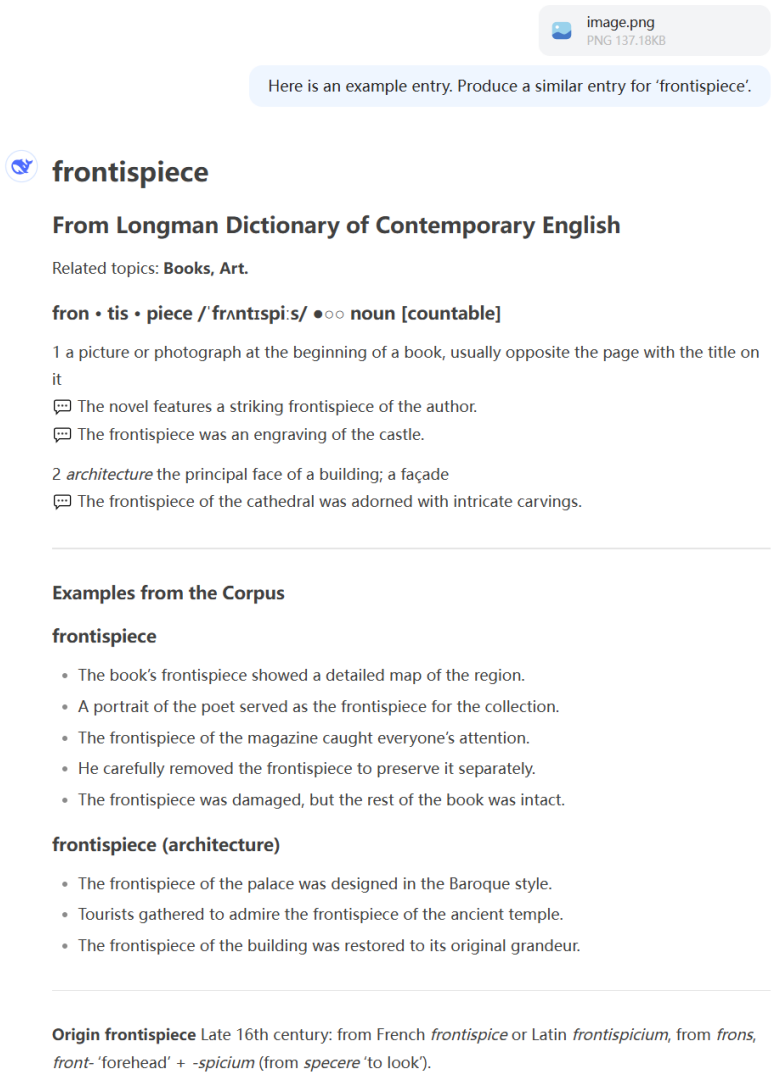


Figure 7: DeepSeek's output for Prompt #5: Here is an example entry. Produce a similar entry for 'frontispiece'.

It might seem paradoxical for users to ask the AI model to produce a dictionary-style entry when they can more directly look up the lexical item in a dictionary. However, consider a more extreme situation: if a word is not included in the dictionaries they usually use, where can language learners obtain information on its meaning and usage? AI might be a valuable tool in such instances. For example, *Yorkiepoo* (a cross-breed between a Yorkshire terrier and miniature or toy poodle) is a newly added word in the *Oxford English Dictionary* (OED), and has

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

not been included in any of the four learner's dictionaries examined. Those who do not have access to the OED, which needs a subscription, can instruct an AI model to create a LDOCE-style or OALD-style entry, for example, for *Yorkiepoo*. In this case, *Yorkiepoo* is an existing word in the English lexicon. However, if a non-existent word is tried, DeepSeek (as well as other models) will still generate prolific outputs, which are known as AI hallucinations (cf. Fuertes-Olivera 2024). This raises other challenges.

4. Concluding remarks

The aforementioned prompts attempt to help users maximise the efficiency and accuracy of AI-generated outputs while crafting minimal instructions. However, there are some limitations to this pilot study. Firstly, the ten words examined represent only a small sample. The present study only covers content words, not function words that primarily serve a grammatical function and express little semantic meaning, making them more difficult to define. In addition, the usage of DeepSeek as a bilingual dictionary is not fully explored. DeepSeek's performance in other lexical tasks also needs future research. Furthermore, the 'entries' generated by AI for these words are not all very lengthy and the prompts crafted do not involve complex logical reasoning. The overall interactions with DeepSeek are therefore quite smooth, with no technical problems. It should be noted again that the AI's responses may vary every time prompted. Therefore, users are encouraged to improve their prompting skills by continuously fine-tuning their prompts to ensure better outputs in both content and format.

References

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

CALD. *Cambridge Advanced Learner's Dictionary & Thesaurus*. https://dictionary.cambridge.org/ [27 April 2025]

COBUILD. *Collins COBUILD Advanced Learner's Dictionary*. https://www.collinsdictionary.com/ [27 April 2025]

DeepSeek. https://chat.deepseek.com/ [23 April 2025]

DeepSeek-AI. 2025. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *arXiv*. https://arxiv.org/abs/2501.12948/.

De Schryver, G.-M. 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.

Fuertes-Olivera, P.A. 2024. Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos* 34(1): 123-140.

Huete-García, A. and S. Tarp. 2024. Training AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos* 34(1): 21-40.

LDOCE. *Longman Dictionary of Contemporary English*. https://www.ldoceonline.com/dictionary/ [27 April 2025]

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2037 (Lexiconote)

Lew, R. 2023. ChatGPT as a COBUILD Lexicographer. *Humanities and Social Sciences Communications* 10: 704.

Lew, R. 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11: 426.

Lew, R., B. Ptasznik and S. Wolfer. 2024. The Effectiveness of ChatGPT as a Lexical Tool for English, Compared with a Bilingual Dictionary and a Monolingual Learner's Dictionary. *Humanities and Social Sciences Communications* 11: 1324.

Li, Q. and S. Tarp. 2024. Using Generative AI to Provide High-Quality Lexicographic Assistance to Chinese Learners of English. *Lexikos* 34(1): 397-418.

OALD. *Oxford Advanced Learner's Dictionary*. https://www.oxfordlearnersdictionaries.com/ [27 April 2025]

OED. *Oxford English Dictionary*. https://www.oed.com/ [28 April 2025]

Ptasznik, B., S. Wolfer and R. Lew. 2024. A Learners' Dictionary Versus ChatGPT in Receptive and Productive Lexical Tasks. *International Journal of Lexicography* 37(3): 322-336.

Rees, G.P. and R. Lew. 2024. The Effectiveness of OpenAI GPT-Generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Orientated Reading Task. *International Journal of Lexicography* 37(1): 50-74.

Tarp, S. and A. Nomdedeu-Rull. 2024. Who Has the Last Word? Lessons from Using ChatGPT to Develop an AI-based Spanish Writing Assistant. *Círculo de Lingüística Aplicada a la Comunicación* 97: 309-321.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

Introduction to the Special Section of Papers from the 6th GLOBALEX Workshop on Lexicography and Neology

Annette Klosa-Kückelhaus, *Leibniz-Institut für Deutsche Sprache, Mannheim, Germany (klosa@ids-mannheim.de) (https://orcid.org/0000-0003-4529-9432)*
and
Ilan Kernerman, *Lexicala by K Dictionaries, Israel (ilan@lexicala.com) (https://orcid.org/0000-0002-9052-8217)*

1. Background

This special section of *Lexikos* focuses on lexicographic neology and neological lexicography, featuring seven of the papers presented at the sixth Globalex Workshop on Lexicography and Neology (GWLN-6),[1] which was held in conjunction with the 28th International Conference of the African Association for Lexicography, AFRILEX 2024, at the University of Pretoria, South Africa.[2]

The GWLN series began as a one-time event conjugated with the 22nd Biennial Meeting of the Dictionary Society of North America (DSNA) at Indiana University, Bloomington in 2019.[3] The first event included thirteen invited papers from around the world, relating to the detection, selection, representation, and research of neologisms in lexicographic contexts. Of these, eight formed a special issue of the DSNA's journal *Dictionaries,* published the following year as volume 41, issue 1 (2020).[4] However, in view of the successful workshop and corresponding publication, we decided to continue regularly on an annual basis, relocating each iteration with the conference of a different continental association for lexicography. GWLN-2[5] was planned to be held in conjunction with the congress of the European Association for Lexicography (Euralex 2020, Alexandropoulos, Greece),[6] but due to the COVID-19 pandemic the onsite event was cancelled. Instead, some of the papers were presented in a standalone online workshop that year,[7] while others were included in another online session of Euralex 2020 the following year,[8] with selected papers from both online sessions published in volume 34, issue 3 as a special issue of *International Journal of Lexicography* (2021).[9] GWLN-3 was also held online, as a pre-conference workshop of the Australasian Association for Lexicography (Australex 2021),[10] focusing on neologisms arising in relation to COVID-19, with thirteen papers published as a volume of *Lexicographica. Series Maior* entitled *Lexicography of Coronavirus-*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

related Neologisms (2022).[11] The same year, GWLN-4[12] was held at the Euralex 2022 Congress (Mannheim, Germany),[13] with all six papers published as part of the conference proceedings.[14] GWLN-5[15] took place in conjunction with the 16th International Conference of the Asian Association for Lexicography (Asialex 2023, Seoul, South Korea),[16] with the main topic of unregistered words. Following the conference, six of the papers were published that year as a special issue of *Lexicography. Journal of Asialex* in volume 10, issue 2 (2023).[17]

As stated in our Introduction to the *Lexicographica, Series Major* volume (Klosa-Kückelhaus and Kernerman 2022), lexicography has existed for millennia and has continuously adapted to societal and linguistic developments, particularly in the recent decades with their intensively profound transformations. The phenomenon of neology, the creation of new lexical items, has been constantly present throughout history, playing a pivotal role in the evolution of language and, by extension, in the development of lexicography. In the last decade, neology has garnered increased attention from research communities, as evidenced also by the establishment of the COST Action European Network on Lexical Innovation (ENEOLI, CA22126, 2023).[18] This heightened focus has led to the emergence of novel practical studies and applications in various domains, including academia, language institutes and observatories, the language technology industry, and the general public. The pace of novelty in daily life is accelerating, and the number of innovations is growing exponentially — all of which are defined by language and affect language in turn. Consequently, there is a growing interest in neologisms and in the role of lexicographic resources in recording and disseminating them to the world.

The overarching objective of GWLN meetings and publications is to investigate the nexus of neology and lexicography on a global scale. This endeavor entails the identification of common factors and the exposition of distinctive characteristics. Furthermore, it involves the dissemination of findings among diverse members, with the aim of fostering mutual exchange and comprehension, enhancing professional proficiency, and attaining user satisfaction. The primary issues under consideration are initiated with the detection of neologisms and are subsequently pursued through their categorisation, lexicographic treatment, and representation. As such, the description in our Introduction to the special issue of *Dictionaries* (Klosa-Kückelhaus and Kernerman 2020) is overall appropriate here too and we reproduce it with slight adjustments:

"Neology constitutes a natural, dynamic, and multilateral part of all living human languages, whether as a reflection or for facilitation of linguistic communication, and lexicographic interest in neologisms is at least as old as dictionaries themselves. There is a vast field of research of neologisms, pertaining to their origin (stemming from the given language as in new word formation, or loan words from other languages including the dominance of English today, as well as combining both), distribution (in general language and in domain-specific language, that is terminology), identification (applying corpus linguistics methods, editorial methods, user generated candidates, and comparison of different methods), evaluation (such as in blogs and chats), and more. The gen-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

eral definition of neologisms as applied here refers to new words, new multiword units, new elements of word formation, and new meanings of either of them, and addresses lexicography-driven or -oriented aspects, including:

— How to interoperate lexicographic datasets with online resources and incorporate neologisms into dictionaries (the media, formatting, labelling, etc.)
— How to deal with grammatical/orthographic/pronunciation variation (descriptive vs. prescriptive approaches)
— How to explain meaning with/without encyclopaedic information, and how to use illustrations and audio-visual media
— How well are neologisms that are integrated in dictionaries accepted by the community (issues of rejection of new words and language purism)
— How differently, if at all, should neologisms be treated in different dictionary types (e.g. in historical comprehensive ones as opposed to those focusing on current usage; in monolingual vs. bilingual dictionaries; in special dictionaries of neologisms; in special domain dictionaries)
— How to deal with neologisms that are no longer *new* and with those no longer used
— How can dictionary users help with finding and informing about neologisms."

In addition to the substantial growth in international and cross-lingual research in neology and its implications for lexicography during the short five-year span since the description quoted above was published, we would like to highlight the rapid and fundamental game-changing advent of generative AI and large language models (LLMs), whose massive impacts have had only their first glimpses in this special section. However, at the time of writing this Introduction (April 2025), the submissions to the seventh iteration of GWLN,[19] which will take place in conjunction with the eLex 2025 conference,[20] already indicate a surge in the interest in the roles of AI in finding, describing, and generating neologisms and their interaction with lexicography.

2. This volume

The first paper in this section offers an overview of the aims and work of the new European Network in Lexical Innovation (ENEOLI), which brings together researchers, practitioners, teachers, students, translators, and other language professionals in the field of neology. Among other things, the network concerns terminology in the field and carries out comparative studies of lexical innovation in different languages. Its presentation here leads directly to issues that are explored in more detail in the other six papers, which concern six languages: English, Georgian, German, Modern Greek, Korean, and Spanish. Two of the papers deal with methods for detecting neologisms (in Korean and Georgian), using both automated and editorial means. Two other papers look at specific elements of word formation (combining forms in English and the

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

suffix *-ing* in Modern Greek) that are used to create new words. One paper presents an in-depth study of the lexicographic treatment of two new verbs in Spanish, and the paper on German neologisms focuses on their role in critical discourse.

Giovanni Tallarico, Rute Costa, Ana Salgado, Judit Freixa, Kris Heylen, John Humbley, Ilan Kernerman, Annette Klosa-Kückelhaus, Ana Ostroški Anić, Federica Vezzani, Vincent Balnat, Petar Božović, Špela Arhar Holdt, and Weronika Woźniak-Żak (**Lexicography and Neology: Building Bridges through a Multilingual Approach with ENEOLI)** examine the work of the European Network on Lexical Innovation (ENEOLI) COST Action, launched in 2023 to address key gaps in the study of lexical innovation, in the context of lexicography. The authors' "focus is on the description and the methodology developed within the Action for creating a glossary that compiles and defines the meta-terminology of neology". After presenting the aims and structure of ENEOLI, the "state of neology research at the intersection with lexicography" is discussed. In particular, research and methods for determining which of the new words, senses, and multiword units "will eventually become part of the language and, therefore, should be included in the general dictionary" are presented. Neologism dictionaries face other challenges, as "neologisms in a dictionary of neologisms are doomed to disappear either by their incorporation into the general dictionary (and the consequent loss of their neology status) or by disappearance from use". One of the main tasks of ENEOLI is the development of a multilingual glossary of the meta-terminology of neology, based on corpora specifically created for the project. The objective is "to create a platform that enables users to access, explore, and understand terminology used in the field of neology across different languages efficiently". The workflow for this specific terminology project is not different from that of lexicographic projects, as three main phases can be identified: the selection of candidates for inclusion in the database, their description on the basis of specific microstructural components, and the follow-up and possible updating of the entries. The glossary is intended to "serve as a practical resource while also advancing the understanding of how neology is approached within lexicography, both in practice and in theory."

In her discussion of **Neologisms and their Functions in Critical Discourse**, Petra Storjohann takes a discourse-oriented rather than a terminological approach. As "neologisms introduce new ways of thinking by embodying novel conceptual knowledge and influencing collective thought processes", these functions challenge lexicographers. Traditionally, key discourse aspects of new words are not part of the lexicographic description. However, neologisms do not appear in isolation, but within discourses. Thus, a usage-based approach (based on corpus linguistics) can support the analysis of "lexical patterns that define the immediate context of a given term". By presenting detailed information on the German neologism *Greenflation*, the author discusses two methods that "should be employed complementarily to investigate discourse, neologisms, and their meanings, roles, and functions": the corpus-based analysis of collocations as well as syntagmatic patterns and methods for identifying lexical–semantic relations

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

at the paradigmatic level. The author concludes: "*Greenflation* and its associated discursive potential are integral to its semantics. Accordingly, this paper posits that lexicographic descriptions should take this into account." Detailed information is given on a proposal for a discourse-oriented type of lexicographic treatment in a new resource on German neologisms of the 2020s, while also discussing the limitations of discursive descriptions: "Still, any attempt to document discourse within a dictionary remains fragmentary, just as any corpus represents only a portion of the broader discourse it reflects." Nevertheless, this paper provides valuable food for thought for lexicographers working on neologisms and will hopefully inspire the inclusion of information on discourse-related aspects in other dictionaries.

In **The Role of Combining Forms in Creating New English Compounds: Data-Driven and Construction Approaches**, Jin-Hong Huang and Yongwei Gao focus on combining forms (CFs), such as *auto-* or *-preneur*, that are very productive in contemporary English in forming new lexical items. They are defined as "a form of a word normally used in compounds in combination with another element to form a word" and can have either a native or neoclassical origin. The authors use data-driven and constructional approaches to study the material and they also discuss their treatment in dictionaries. Their approach to analyse the formational mechanisms of new compounds by using the method of construction morphology (CxM) is fruitful; in this theory, "a word is conceptualised as a pairing of form and meaning, with words in the lexicon understood as systematic form-meaning correspondences." It is used to show how the "systematic nature of syntactic constraints in CxM shows that CFs combine only bases that conform to specific word class patterns". Productivity is defined quantitatively, i.e., "the number of new compounds included in dictionaries over specific periods" is used to analyse the productivity of a particular CF, and it is classified as "highly productive, moderately productive, or low in productivity". Three English–Chinese neologism dictionaries and *Oxford English Dictionary* are the data sources for the study of 11 CFs of Classical Greek or Latin origin. The CFs's dates of first use, their formal properties, the word classes of new compounds formed with the CF, and the semantic properties of the new compounds, as well as the semantic changes of the CFs in new compounds, are all analysed in detail. Of particular interest to practical lexicographers is the discussion of the results concerning "demarcation between compounds and others and their lexicographic treatment" (cf. 6.3). The authors invite dictionary compilers "to propose a standardised lexicographical annotation for CF compounds, distinguishing them clearly from derivatives and blends to ensure the authority and accuracy of dictionaries".

New meanings emerge not only for some of the words used as CFs mentioned above, but also for lexemes. In their article **Semantic Verbal Neology in Spanish Dictionaries: Case Study of the Lexicographic Representation of *anestesiar* and *coreografiar***, Marina Berri and Victoria de los Ángeles Boschiroli discuss two Spanish verbs that are in fact semantic neologisms, but are not treated as such in "widely used online general dictionaries". The problem with treating seman-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

tic neologisms in general dictionaries is that "they require the modification of an existing article, either by adding a sense, in the case of the emergence of a new meaning, or by modifying an existing sense, which may be necessary in the case of extensions, restrictions or other types of displacements". The authors also note that "in dictionaries of neologisms, defining a semantic neologism implies referring in some way to a previous meaning that this dictionary does not include". Based on Patrick Hank's Theory of Norms and Exploitations (TNE), new senses or new meanings can be "discriminated by identifying phraseological patterns" and are then recorded in a dictionary according to their frequency. TNE is also "relevant to study neologicity", since the exploitation of norms is a gradual phenomenon. The case studies of *anestesiar* and *coreografíar* show in detail how TNE can be used to find information about new meanings as well as pragmatic values that should be recorded in a dictionary entry. One of the results of the study is that the structural and semantic reality of semantic neologisms differs significantly from their representation in general language dictionaries. Therefore, the authors propose to "adopt a systematic approach for the inclusion of VSN [verbal semantic neologisms] that anticipates how to deal with the grey area between exploitations and secondary/new norms" (i.e. new senses).

Anna Vacalopoulou presents an in-depth study of the treatment of Anglicisms in general dictionaries of Modern Greek in **Digging for *-ings*: A Survey of Selected (Pseudo)-Anglicisms in Dictionaries of Modern Greek**. Anglicisms are defined as "lexical items borrowed from English and integrated into the lexicon of another language", and pseudo-Anglicisms are defined as "words that seem English but do not exist in the same form or meaning in English". Anglicisms (particularly nouns ending in *-ing*) can be "adopted verbatim, spelled out in the Latin alphabet, or may be transliterated into the Greek alphabet". In the early stages of the integration process, both methods are used, but later it often occurs that a "hybrid approach to incorporating new terms" is also common, which is problematic for lexicographic treatment. Five Modern Greek dictionaries were selected for analysis and all nouns ending in *-ing* were extracted and classified (Anglicisms and pseudo-Anglicisms). The lexicographic treatment of these nouns varies in terms of inclusion and treatment — for example, the orthographic information, where "a notable disconnect between prescriptive rules and actual usage" can be found. The study may inspire practical lexicographers of other languages to aim for a more systematic treatment of Anglicisms/ Pseudo-Anglicisms in their dictionaries.

In **Semi-Automatic Detection of New Words in Modern Georgian**, Tamar Laluashvili and Tinatin Margalitadze provide information on the historical background ("rapid socio-political changes in the country [Georgia] after the collapse of the Soviet Union") that has led to an increase in neologisms in Georgian. The new realities of life after gaining independence as well as other events, such as the COVID-19 pandemic or the Russian–Ukrainian war, have left their mark on the language in the form of loanwords and blends (e.g. *რაშიზმი / rashizmi* 'Rushism') and new word formations (e.g. *ინდმეწარმე / indmets'arme* 'individual entrepreneur'). The authors discuss methods and tools that were developed

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

for semi-automatic detection of these neologisms in modern Georgian with the aim of finding candidates for inclusion in a neologism dictionary. As a first step, "a study corpus was composed and two lemmatisers were applied to it"; the corpus "served as an exclusion source" and the lemmatisers were "used for reducing the number of neologism candidates". Using this method, both neologism candidates and "undocumented words" were found, with the latter being cases where words cannot be considered neologisms "but are not documented in dictionaries". In total, "over 1 700 lexical neologisms in modern Georgian" were found, the majority of which are Anglicisms. Another category is "hybrid words, with borrowed and Georgian roots", and some new "word-forming components" that were also identified. Interestingly, some "new polysemous meanings of Georgian words, corresponding to *like*, *share*, *page*, *wall*, *friend*, and *follower*" have emerged. The material collected will form the basis for a new, explanatory online neologism dictionary, and for future monitoring of neologisms in modern Georgian. Overall, this paper provides insights into the methodology of semi-automatic neologism detection and its challenges for lexicographic purposes, which may be helpful to projects for other languages in their search for neologism candidates.

In the last paper of this section, **Detection and Description of Neologisms in Korean Lexicography: Methodological Issues in Corpus Balance, Word Unit Bias and LLM Assistance,** Kilim Nam, Soojin Lee and Hae-Yun Jung look into the future of lexicography, where LLMs can support "neologism extraction and dictionary compilation", even if there are still some limitations at present. The authors present information on the research background of the Korean Neologism Investigation Project and the chosen methodology (a large-scale collection of neologisms from different corpora) before discussing the current state of corpus-based neologism extraction and related corpus linguistic issues. The main challenge for the neologism collection is that it relies mainly on newspaper and news corpora, although "it is evident that news data cannot fully represent the linguistic reality of Korean native speakers". Other sources, such as instant messages or online posts, should also be used to find and describe neologisms. The authors ask "whether LLMs trained on large-scale data encompassing various sources, such as encyclopaedias like Wikipedia, books, academic articles, news, blogs, forums, and other websites, can serve as an alternative". They tested how LLMs performed when asked to identify and describe semantic neologisms and present their findings on how they "perform the task of recommending neologisms and unregistered words." They conclude: "The experimental results provided by ChatGPT and CLOVA X indicate that LLMs show potential as useful tools for detecting unregistered words and neologisms that are often omitted from traditional news-centred corpora. However, issues such as noise in the list of recommended neologisms, limitations in describing semantic neologisms, and duplicate outputs remain challenges to be addressed."

We hope that this closing paper provides food for thought and encourages further studies on the performance of LLMs for neologism detection in other languages as well. Moreover, as noted above, LLMs are already surging as detec-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2038 (Lexicofocus)

tors, describers, and generators of neologisms and are thus expanding our field of research to their impact on neology and lexicography (cf. GWLN-7, 2025).

Endnotes

1. https://globalex.link/gwln-6_2024/
2. https://www.afrilex.co.za/conferences
3. https://dictionarysociety.com/conference/dsna-22-in-bloomington-indiana/
4. https://dictionarysociety.com/wp-content/uploads/2020/05/Dictionaries-41.1-Table-of-Contents.pdf
5. https://globalex2020.globalex.link/gw-euralex2020
6. https://euralex.org/conferences/euralex-2021/
7. https://globalex.link/gw-euralex2020/
8. https://euralex.org/wp-content/uploads/2022/06/EURALEX2021_ScientificProgramme-Preview-2-1.pdf
9. https://academic.oup.com/ijl/issue/34/3
10. https://www.australex.org/our-conferences/
11. https://www.degruyter.com/document/doi/10.1515/9783110798081/html
12. https://globalex.link/globalex2022euralex/
13. https://euralex2022.ids-mannheim.de/
14. https://euralex.org/category/publications/euralex-2022/
15. https://globalex.link/globalex2023asialex/
16. https://korealex.org/asialex_2023
17. https://utppublishing.com/toc/lexi/10/2
18. https://eneoli.eu/
19. https://globalex.link/gwln-7_2025/
20. https://elex.link/elex2025/

References

Klosa-Kückelhaus, A. and I. Kernerman. 2020. Global Viewpoints on Lexicography and Neologisms: An Introduction. *Dictionaries. Journal of the Dictionary Society of North America* 41(1): 1-9.

Klosa-Kückelhaus, A. and I. Kernerman. 2022. Lexicography of Coronavirus-related Neologisms: An Introduction. Klosa-Kückelhaus, A. and I. Kernerman (Eds.). 2022. *Lexicography of Coronavirus-related Neologisms.* Lexicographica. Series Maior 163: 1-10. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110798081-001

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

Lexicography and Neology: Building Bridges through a Multilingual Approach with ENEOLI*

Giovanni Tallarico, *Dipartimento di Lingue e Letterature Straniere, Università di Verona, Verona, Italy (giovanni.tallarico@univr.it) (https://orcid.org/0000-0002-0585-4473)*

Rute Costa, *Centro de Linguística da Universidade NOVA de Lisboa, Lisbon, Portugal (rute.costa@fcsh.unl.pt) (https://orcid.org/0000-0002-3452-7228)*

Ana Salgado, *Centro de Linguística da Universidade NOVA de Lisboa, Lisbon, Portugal (anasalgado@fcsh.unl.pt) (https://orcid.org/0000-0002-6670-3564)*

Judit Freixa, *Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Spain (judit.freixa@upf.edu) (https://orcid.org/0000-0002-1440-5514)*

Kris Heylen, *Instituut voor de Nederlandse Taal, Leiden, the Netherlands, and KU Leuven, Belgium (kris.heylen@ivdnt.org) (https://orcid.org/0000-0002-8361-0669)*

John Humbley, *ALTAE, Université Paris Cité, Paris, France (humbley.john@orange.fr) (https://orcid.org/0000-0002-1759-1156)*

Ilan Kernerman, *Lexicala by K Dictionaries, Israel (ilan@lexicala.com) (https://orcid.org/0000-0002-9052-8217)*

Annette Klosa-Kückelhaus, *Leibniz-Institut für Deutsche Sprache, Mannheim, Germany (klosa@ids-mannheim.de) (https://orcid.org/0000-0003-4529-9432)*

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

Ana Ostroški Anić, *Institute for the Croatian Language, Zagreb, Croatia (aostrosk@ihjj.hr) (https://orcid.org/0000-0001-9999-0750)*

Federica Vezzani, *Dipartimento di Studi Linguistici e Letterari, Università degli Studi di Padova, Padua, Italy (federica.vezzani@unipd.it) (https://orcid.org/0000-0003-2240-6127)*

Vincent Balnat, *Département d'études allemandes/ LiLPa (UR 1339), Université de Strasbourg, Strasbourg, France (balnat@unistra.fr) (https://orcid.org/0000-0003-1023-0574)*

Petar Božović, *Department of English Language and Literature, Faculty of Philology, University of Montenegro, Nikšić, Montenegro (petarb@ucg.ac.me) (https://orcid.org/0000-0002-2921-3881)*

Špela Arhar Holdt, *Faculty of Arts and Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia (spela.arharholdt@ff.uni-lj.si) (https://orcid.org/0000-0003-0565-0531)*
and
Weronika Woźniak-Żak, *Doctoral School of Humanities, Faculty of Philology, University of Lodz, Poland (weronika.wozniak.zak@edu.uni.lodz.pl) (https://orcid.org/0000-0003-0169-2734)*

Abstract: Neology, the study of new word formations and new word meanings, remains an underrepresented yet critical dimension of linguistic research. This paper explores the efforts of the European Network on Lexical Innovation (ENEOLI) COST Action, launched in October 2023, in addressing key gaps in the study of lexical innovation in connection with lexicography. With 343 participants from 51 countries as of March 2025, ENEOLI seeks to build a comprehensive, multilingual approach to neology research. This initiative focuses on developing digital resources, comparative studies, and professional training to elevate the understanding of neology within the linguistic community, including the development of a multilingual glossary of the meta-terminology of neology. By integrating lexicographic and computational methodologies, ENEOLI sets a benchmark for multilingual research. This initiative not only contributes to a deeper understanding of neology but also provides resources that support its application across diverse academic and professional contexts, including translation, education, and digital humanities. Through its collaborative framework, ENEOLI highlights the significance of lexical innovation in understanding language evolution and its cultural and social dimensions.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

Keywords: ENEOLI, NEOLOGY, LEXICOGRAPHY, LEXICAL INNOVATION, LANGUAGE CHANGE, MULTILINGUAL GLOSSARY, TERMINOLOGY

Opsomming: Die leksikografie en neologie: Die smee van bande deur 'n veeltalige benadering met ENEOLI. Neologie, die studie van nuwe woordvormings en nuwe woordbetekenisse, bly 'n onderverteenwoordigde dog kritieke aspek van taalkundige navorsing. Hierdie artikel ondersoek die pogings van die Europese Netwerk vir Leksikale Innovasie (ENEOLI) COST-aksieplan, wat in Oktober 2023 van stapel gestuur is, om sleutelkwessies in die bestudering van leksikale innovasie rakende die leksikografie aan te pak. ENEOLI streef met 343 deelnemers uit 51 lande (teen Maart 2025) daarna om 'n omvattende, veeltalige benadering tot neologienavorsing te volg. Hierdie inisiatief fokus op die ontwikkeling van digitale hulpbronne, vergelykende studies en professionele opleiding om die begrip van neologie binne die taalkundige gemeenskap te bevorder, insluitend die ontwikkeling van 'n veeltalige glossarium van die metaterminologie van neologie. Deur leksikografiese en rekenaargebaseerde metodologieë te integreer, stel ENEOLI 'n maatstaf vir veeltalige navorsing. Dié inisiatief dra nie net by tot 'n dieper begrip van neologie nie, maar voorsien ook hulpbronne wat die toepassing daarvan oor diverse akademiese en professionele kontekste, insluitend vertaling, opvoedkunde, en digitale geesteswetenskappe, ondersteun. Met sy samewerkingstruktuur beklemtoon ENEOLI die belangrikheid van leksikale innovasie in die verstaan van taalontwikkeling en die kulturele en sosiale dimensies daarvan.

Sleutelwoorde: ENEOLI, NEOLOGIE, LEKSIKOGRAFIE, LEKSIKALE INNOVASIE, TAALVERANDERING, VEELTALIGE GLOSSARIUM, TERMINOLOGIE

1. Introduction

Neology is a young discipline with a strong interdisciplinary character.[1] While having its own objectives and theoretical aims, most of its lines of research are at the intersection with other, more established, disciplines. For instance, many studies look at neologisms from a morphological perspective and analyse how lexical innovations modify the rules of word formation according to a given language's grammar, or which describe the new trends in a language according to different formation mechanisms, or which compare these trends among different languages. Some examples of other disciplines involved include sociolinguistics (dialectal or functional variables of neologisms in language contact and borrowing), terminology (lexical items with a specialised meaning), or pragmatics (the function a neologism fulfils in an utterance, e.g., denominative, ludic, or ironic).

The study of neologisms is crucial for understanding not only the evolution of language but also the cultural and social forces shaping lexical innovation. Neology has until recently been the poor relation of linguistic research for several reasons.[2] From a theoretical standpoint, it could have been argued that lexicology itself, which englobes neology, was a peripheral concern, on the assumption that morphology, syntax and semantics could on their own account for lexical innovation. On the national level, scholarly traditions in neology were devel-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

oped, which indeed structured research in individual countries and regions, for example France and Québec for French. However, such different cultures of research made exchanges at the international level difficult to achieve, in spite of flourishing research in individual settings and, partly, in international networks such as the Réseau panlatin de terminologie (REALITER), Red Iberoamericana de Terminología (RITerm), and Réseau Lexicologie, Terminologie, Traduction (LTT). There thus remains a pressing need for cohesive and expansive research networks to unify these efforts and further advance the field.

The ENEOLI COST Action 22126 (European Network on Lexical Innovation)[3] aims to bridge gaps in the study of neology by promoting collaboration among researchers, developing innovative resources, and conducting comprehensive multilingual studies. To achieve these objectives, ENEOLI focuses on four key areas: (1) the creation of a multilingual glossary of the meta-terminology of neology; (2) the development of methods, digital resources, and tools for neological research; (3) diachronic and synchronic comparative studies of lexical innovation; and (4) training initiatives in neology. These four pillars are unified by the shared need for a common meta-terminology that captures the processes and concepts central to lexical innovation.

In this paper, the focus is on the description and methodology developed within the Action for creating a glossary that compiles and defines the meta-terminology of neology. To achieve this, neology studies are interrelated with lexicographic, metalexicographic and terminology studies, along with corpus linguistics theories and methods.

The paper is structured as follows: Section 2 introduces the four Working Groups of ENEOLI and the work carried out in each. In Section 3, the current state of neology research at the intersection with lexicography is reflected upon, while in Section 4, the lexicographic dimension of ENEOLI and the creation of the Multilingual Glossary of Neology are delved into. In Section 5, it is concluded that the glossary will serve not only as a practical resource but will also contribute to advancing the theoretical understanding of how neology is addressed in lexicography.

2. ENEOLI COST Action

2.1 Objectives and structure

ENEOLI's primary goal is to establish a network of neology researchers across Europe and beyond, enabling the sharing of best practices, methodologies, and resources. This section details the objectives of ENEOLI, which are organised into four Working Groups (WGs):

WG1 is dedicated to the creation of an open-access glossary that defines key terms in lexical innovation across multiple languages. To achieve this, the group is organised in seven distinct tasks, each targeting a specific component of glossary development:

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

— Task 1.1 "Exploration of the Lambert-Lucas Corpus" focuses on preliminary term extraction from a French-language starter corpus to establish a foundation for the glossary;

— Task 1.2 "NeoCorpus & NeoVoc" involves the development of NeoVoc, a multilingual neology thesaurus, and NeoCorpus, a comprehensive corpus designed to support multilingual studies and term validation;

— Task 1.3 "Domain & Knowledge Organization" addresses the classification and organisation of terminology by domains and fields of knowledge to ensure conceptual clarity;

— Task 1.4 "Macro & Microstructure Planning" focuses on defining the overall structure of the glossary, including the selection and arrangement of entries and the detailed design of individual entry components;

— Task 1.5 "Data Representation & Presentation" deals with the representation of glossary data, ensuring usability and accessibility in digital formats;

— Task 1.6 "Glossary Editing" oversees the refinement and standardisation of glossary content, including the validation of terms and their linguistic equivalents; and

— Task 1.7 "Documentation & Guidelines" provides comprehensive documentation of methodologies, workflows, and usage guidelines to support the glossary's creation and future updates.

WG2 gathers researchers working or who are interested in developing and refining digital tools, resources and methods for identifying, tracking, and analysing neologisms. This group is organised into six tasks:

— Task 2.1 "Survey on Lexical Innovation Methods and Practices, Available Resources, and Tools" involves the collection of data on the existing lexical innovation methods and practices, as well as available tools and resources, by conducting a wide survey among ENEOLI members and other researchers working in fields such as lexicography, terminology, linguistics, and translation;

— Task 2.2 "Digital Resources and Tools for Neology" focuses on identifying digital resources and tools for work on lexical innovation with a view to compiling an ENEOLI repository;

— Task 2.3 "Guidelines and Best Practices on the Methodology of Neology Extraction, Description, Analysis and Follow-up" involves identifying and synthesising best practices by analysing relevant research literature and gathering experience;

— Task 2.4 "Methods for the Identification and Monitoring of Neologisms" aims to obtain an overview of computational methodologies, tools and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

resources for the (semi-)automatic identification and monitoring of neologisms;

— Task 2.5 "Neology from a Lexicographic Perspective" conducts a comparison of how different lexicographic institutes and projects select, describe, and follow up lexical innovations in their databases and dictionaries; and

— Task 2.6 "Filter Words, Algospeak, and Neologisms" explores strategies to circumvent internet bans on offensive or taboo content through filter words, aiming to create a typology of lexical innovations related to this avoidance and enhance detection methods within large corpora.

The WG2 tasks connect with the results of the survey conducted in T2.1, which underlines the need for fine coordination and inter-task collaboration.

WG3 focuses on comparative aspects of neology in European languages, conducting diachronic and synchronic studies with special attention to technology-driven and contact-induced language changes. The group is divided into six tasks, topics and activities which are carried out largely independently from each other:

— Task 3.1 "Blends" aims to compile monolingual corpora, annotated with blend types in different languages and is divided into two subtasks: T3.1.1 concerns the decoding of blends by speakers, and T3.1.2 relates to their translatability;

— Task 3.2 "Neology and Discourse" undertakes investigations of neologisms through the lens of critical discourse analysis with the initial aim of compiling a bibliography of relevant studies;

— Task 3.3 "Neology and Proper Nouns" deals with neologisms coined from proper nouns and aims at analysing this particular way of creating common nouns, adjectives or verbs, as well as other proper nouns, taking into account the textual dimension of this type of lexical innovation;

— Task 3.4 "Inclusion of Neologisms in the Mental Lexicon" focuses on cognitive aspects of neologisms and studies the phonological, orthographic, grammatical and semantic inclusion of neologisms in the mental lexicon. This task has two subtasks: T3.4.1 looks at how neologisms are mapped into languages that have grammatical gender (and why), while T3.4.2 is dedicated to creating standardised tests for gauging how people and AI systems create and understand neologisms;

— Task 3.5 "Verb Neology" studies the emergence of new verbs, or borrowed verbs mainly in French and Czech, and their usage across different domains such as technology, ecology, and politics. Key areas of research include morphology, semantics, and usage frequency, which help in assessing how new verbs integrate into the existing lexicon and their perception and use by speakers; and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

— Task 3.6 "Neology and Gender Equality in the Languages of Europe" is a recently added task, and the scope, aims and methodological approach are currently being discussed.

The overall goal of WG4 is to develop an innovative methodology and didactic materials for teaching neology in various professional domains, including translation, interpreting, language teaching, and related fields. To achieve this, it will identify and address the linguistic, intercultural, technological, and other competencies required for neology across diverse fields, integrating these into academic curricula to enhance understanding and foster autonomy in managing neologisms. Additionally, WG4 aims to create innovative modules for continuous professional development (CPD) to support translation, language education, and related disciplines. The work is organised into four interconnected tasks, each further divided into subtasks:

— Task 4.1 "Innovating Neology Teaching Methods" focuses on designing innovative modules for teaching neology in higher education settings and within CPD programmes;

— Task 4.2 "Developing Neology Teaching Materials" involves designing didactic materials for teaching neology across various disciplines;

— Task 4.3 "Conducting a CPD Training School" aims to pilot the developed modules and materials; and

— Task 4.4 "Fine-Tuning Teaching Methods and Materials" involves refining the teaching approaches and materials from T4.1 and T4.2 by incorporating the feedback from the training school conducted in T4.3.

2.2 Methodological approach

This section outlines the methodological framework adopted across the ENEOLI WGs to ensure a systematic and collaborative approach to the study of neology and lexicography. Each WG employs tailored methodologies aligned with its specific objectives, combining traditional lexicographic practices with computational tools and emerging technologies. By integrating corpus-based research, multilingual and comparative analysis and advanced digital resources, the initiative seeks to establish a dynamic model that can be adapted and applied across diverse contexts. The methodologies described herein highlight the synergy between theoretical and practical approaches, ensuring that the project outputs are comprehensive and relevant to a wide range of disciplines and professional contexts.

The methodological framework adopted in WG1 ensures a systematic approach to the development of a multilingual glossary of the meta-terminology of neology. This methodology combines corpus analysis, computational tools, and expert validation to ensure accuracy, consistency and applicability.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

The initial phase involved the compilation and analysis of a starter corpus, consisting of six French monographs from Lambert-Lucas Editions. These monographs were selected for their coherent and systematic approach to neology. The corpus was processed using semi-automatic term extraction tools, including AntConc[4] and Sketch Engine[5], to identify high-frequency candidate terms.

To ensure the accuracy and relevance of the extracted terms, a panel of experts in French neology, alongside WG1 members, conducted a rigorous validation process. This resulted in a provisional list of monolexical terms, documented in ENEOLI Deliverable 1.1 (Silva et al. 2024). The validated terms were then catalogued and integrated into NeoVoc[6], a multilingual neology thesaurus hosted on ENEOLI Wikibase, which facilitates the structured linking of concepts and their multilingual equivalents.

The enriched metadata from NeoVoc is stored in NeoCorpus, supplemented by references collected in a Zotero library.[7] This integration supports complex cross-linguistic analyses and the tracking of term usage over time, enhancing the glossary's relevance to multilingual and comparative research.

This methodological approach represents an intersection of lexicographic principles and computational advancements. By synthesising lexicographic methods with digital tools, WG1 ensures that the glossary serves as both a theoretical and practical resource for dictionary-making. Moreover, the methodology is designed to be adaptable to other languages and domains, providing a template for future lexicographic initiatives.

Given the nature of the general aims of WG2, digital resources and computational methods are at the centre of attention of individual tasks. However, T2.1, T2.3 and T2.5 also rely on qualitative analyses of existing documentation on guidelines and practices in describing neology. T2.3 and T2.5 have been particularly devoted to learning about successful existing practices through webinars and workshops given by their members with relevant experience. T2.5 is using the input for developing an ENEOLI typology of selection criteria for including and excluding lexical innovations from lexicographic databases (see Section 4.2). Being entirely organised around investigating and exploiting best methods in lexical innovation research, T2.4 is equally interested in traditional approaches and methodologies that leverage the potential of machine learning and Large Language Models and the use of appropriate metrics (for extraction and evaluation). Applying Linguistic Linked Open Data (LLOD) principles in the representation of neologisms is also of relevance to this Task.

In WG3, each task has specific aims and methodological approaches. T3.1 works mainly with word association tests. T3.2 and T3.5 use corpora to discern formal and semantic patterns, and thematic or communicative focal points, while T3.2 also concentrates on semantic (synchronic and diachronic) analysis, the exploration of corpus methods, and the application of these insights to lexicography. For T3.3 and T3.6, the methodology is still to be worked out. In T3.4, a questionnaire will be created in which neologisms (blends, compounds, and derivations) are presented to participants of various languages; the participants (human

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

and AI systems) will have to choose the most probable meaning in single-choice questions.

As far as WG4 is concerned, T4.1 focuses on developing innovative approaches to teaching neology through a structured, multi-step process. It begins with a status quo analysis that involves conducting surveys and interviews with key stakeholders, such as translators, interpreters, and language educators. This analysis aims to identify the linguistic, intercultural, technological, and other challenges that professionals face when dealing with neologisms. Simultaneously, a literature review is conducted to examine existing teaching methods in relevant fields. Based on these insights, the task defines the core competencies required to perceive and manage neologisms, ensuring they meet professional needs. The final step involves designing innovative, competence-based training methodologies for academic courses and CPD. Building on T4.1, T4.2 will develop practical teaching materials for use in these curricula and programmes, including glossaries, manuals, corpora, videos, online courses, and more, ensuring diverse and accessible resources. T4.3 will apply these methodologies and materials by organising a training school as part of CPD, providing a platform to pilot the innovations developed in earlier tasks. The feedback of the participants will be essential to the iterative process of refining teaching methods and materials. Finally, T4.4 will refine the methodologies and materials based on this feedback, ensuring they are effective, adaptable, and aligned with the needs of various professional fields. This iterative, stakeholder-driven approach aims to ensure that WG4's outputs are the most useful, practical, and applicable across diverse domains.

2.3 Preliminary findings and ongoing work

This section outlines the advancements made during the first year of ENEOLI's activity, highlighting key advancements and laying the foundation for future phases of the project. The efforts of each WG are outlined below:

2.3.1 Working Group 1

WG1 has made remarkable progress in creating resources for multilingual and cross-linguistic neology research. An initial ice-breaking survey (Humbley and Salgado 2024) has engaged ENEOLI members, resulting in a collection of 129 references, including articles, books, and specialised dictionaries, spanning a broad temporal range (1965–2023) across multiple languages. Of these, 70 references are machine-readable, forming a solid foundation for ENEOLI's lexicographic work.

A major milestone is the development of NeoVoc, a multilingual neology thesaurus hosted on the ENEOLI Wikibase, and the assembly of NeoCorpus, a comprehensive corpus designed to support the project's multilingual dimension.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

These resources, along with a meticulously curated Zotero library, represent a significant step toward a standardised and systematic approach to cataloging and analysing new lexical items across languages.

The Wikibase workflow enables the creation and management of concept entries and their multilingual equivalents. Initially, French terms — resulting from T1.1 — were uploaded, and corresponding concept entities were created in other languages. Glosses consisting of short definitions of word meanings were also reviewed and, where relevant, added.

The initial phase of the review process focused on assessing the multilingual equivalents that had been automatically imported from Wikidata for French. To facilitate this, a specialised query system was used to identify and highlight entries that required verification. Reviewers accessed these flagged entries via their unique identifiers and followed a structured validation procedure:

— If the imported equivalent listed in the "statements" section was accurate, the warning tag was removed to confirm its validity.

— If the equivalent was incorrect or needed revision, it was modified accordingly, and the Wikidata warning was deleted before saving the changes.

— If a gloss was present in the "description" section, it was examined and either approved, edited, or replaced as needed.

Additionally, a new validation protocol mandates that each confirmed equivalent must be explicitly annotated with the "validated by" qualifier. This qualifier records the name of the person responsible for validation, ensuring accountability and enabling the system to generate real-time reports on validated terms. If a validator does not yet have a Wikibase entity, a new entry is created following a predefined procedure.

Once validated, these equivalents are integrated into lexicographic entries (also referred to as lexeme entities, identified by IDs beginning with 'L'). These entries are systematically linked to their corresponding concepts and include structured linguistic data describing word senses and forms. Future phases will expand these efforts by validating polylexical terms, integrating additional languages, enriching NeoCorpus, and completing key deliverables such as data representation, glossary editing, and comprehensive documentation. These ongoing activities highlight WG1's dedication to bridging the gap between neology and lexicography, creating tools that not only document lexical innovation but also directly support its application in dictionaries. While the details of other WG1 tasks are beyond this article's scope, their collective goal is to broaden ENEOLI's impact and reinforce its relevance in lexicographic research.

2.3.2 Working Group 2

The greatest achievement of WG2 during the course of the Action's first year has been the survey "Studying and Documenting Neologisms: Survey on Methods,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

Practices, Tools, and Resources (ENEOLI COST Action CA22126)", which was designed to collect information about the methods, practices, tools, and resources used by ENEOLI members as well as other research groups, individual researchers, and language specialists involved in the field of lexical innovation. The survey was organised around specific projects or research topics, such as resource creation in lexicography and terminology, software development for neology tracking, and academic research in lexicology, linguistics, and translation studies. Additionally, a section was dedicated to teaching and training practices related to neology. By targeting these various projects and research topics, the survey provides an invaluable source of data and information for tasks within the entire Action. Preparing the survey was a collaborative effort of members from other tasks and WGs. The results will help shape the future direction of WG2. In addition, research has been carried out within T2.5 to compare the treatment of lexical innovation by lexicographic institutes and projects (see Section 4.2).

2.3.3 Working Group 3

As for WG3, the task on blends has developed a work plan and a set of guidelines that will allow the members to compile monolingual corpora of blends that follow a shared design. The task on neology and discourse started compiling a bibliography entitled "European Discourse Studies and Neologisms". A series of studies have been undertaken, focusing on notable neologisms emerging from European discourses of empowerment, the migration crisis, Covid-19, and the climate crisis, in different languages, including Breton, English, French, German, Icelandic, Polish, and Portuguese. Initial findings have been shared during regular meetings, elucidating the corpora, methods, and discoveries, while also addressing challenges. In the task on verb neology, recent discussions have propelled the project forward: initial themes include the morphological adaptation of new verbs in French and Czech, the syntactic and lexical incorporation of neological verbs in Galician, and the reflection of political and social dynamics through neological verbs in French and Polish contexts. A collective volume in French is already in preparation.

2.3.4 Working Group 4

During the first year of WG4's activities, initial progress has been made across several key tasks. The needs analysis subtask has involved developing the survey with WG2, focusing on the challenges and needs in teaching neology in various professional domains. The survey, designed to gather insights from key stakeholders, has been launched, and the results are pending. Additionally, the literature review has led to the compilation of a list of 49 relevant references, categorised into five key areas: (a) neology in language teaching (L1 or L2), (b) neology in teacher training, (c) neology in teaching specialised languages/terminology/translation, (d) titles of interest, and (e) general works on vocab-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

ulary teaching related to neology. This review has highlighted the fragmented nature of current research on teaching neology, while also contributing to identifying potential avenues for further research to fill this gap. The subtask on defining neology competences has made initial progress by outlining the competences needed in fields such as translation, interpreting, and language teaching. Meanwhile, the task focused on designing innovative training methods for university courses and CPD is in its conceptual phase, with draft modules being reviewed and plans for professional interviews under way to gather practical insights for future training materials. This preliminary work has set the stage for further development, pending the survey results.

2.4 Grant awarding and communication/dissemination activities

Raising awareness of the constant evolution of language, highlighting the importance of neology research, and fostering a high-level, comprehensive, multilingual study of lexical innovation are core aspects supported by the grant awarding. Within the ENEOLI COST Action, this constitutes an integral part of the implementation of its objectives and principles. This networking initiative, which promotes collaboration among researchers, encompasses two primary types of activities: participation in conferences organised by third parties, and mobility of our members.

Firstly, the need to enhance the understanding of neology within the linguistic community is addressed by promoting research in this field at high-level conferences. The active participation in sessions and discussions at such events (in 2024, held in locations such as Croatia, Poland, South Africa, Spain, the UK, and USA) fosters knowledge exchange around neology in European languages and helps to raise awareness about the importance of lexical innovation. Presentations support the dissemination of the work performed as part of research conducted within the WGs, as well as of the results of an individual member's research within the scope of the Action. This, in turn, effectively increases visibility, attracts new members, and provides valuable insights and feedback from the scientific community, institutions, and language professionals on the research undertaken within the Action.

Secondly, considering the scope of the Action reflected in the number of participants with different scientific backgrounds, the mobility of researchers constitutes a prerequisite for the main objective achievement, which is to establish a network of researchers and innovators and to share best practices and methods around the study of lexical innovation. Short-term scientific missions play a crucial role in connecting researchers and ensuring a circulation of ideas and knowledge across countries with different research traditions. Hitherto, the onsite meetings between professionals affiliated with language institutions from Estonia, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Romania, Serbia, and Spain have contributed to sharing effective frameworks and approaches for investigating lexical innovation regarding, for example, blend acquisition,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

neological terminological collocations, or tools for neologism extraction. The face-to-face visits supported by grants significantly facilitated the successful completion of tasks determined and designed within WGs, such as: preparing the ENEOLI survey launch, establishing software infrastructure and workflows for corpus development, advancing neology research, refining digital methodologies for analysing lexical innovation, and preparing standardised tests to study how people — and AI — create and interpret neologisms. The mobility will also support the delivery of joint publications or conference presentations highlighting the Action's outcomes.

Moreover, both networking activities provide young researchers and innovators with opportunities to acquire methodological and theoretical knowledge, boost their careers, and increase their visibility in the scientific community. Grant support also plays an important role in involving members from ITC countries,[8] who may have different research and networking traditions and opportunities.

The primary objective of ENEOLI's communication and dissemination strategy is to showcase the Action's activities and results, and more generally, to highlight the importance of understanding linguistic innovation across cultures and contexts. In this light, the aim is to engage both academic and non-academic audiences, promoting the relevance of neologisms and neology. In its first year, key infrastructures were established to achieve this objective, including a cohesive visual identity for consistent branding, and an interactive website as a central information hub. Social media profiles — on Facebook, LinkedIn, X, and Instagram — were created to broaden engagement, combining strategic content with participatory features. A key ongoing initiative is the "Neologism of the Week", which features one neologism per week from diverse languages. These neologisms are selected based on linguistic research, trends, and suggestions submitted by the public through the ENEOLI website, ensuring a collaborative and inclusive approach. Each entry is presented with a concise explanation and shared as an illustrated post on social media. At the time of writing this paper, nearly 30 neologisms have been highlighted, for example *adulting* (English), the act of managing adult responsibilities; *tongpunter* (Dutch), a word on the tip of one's tongue; *tracances* (French), the practice of working remotely while on vacation; and *Buchendschmerz* (German), the bittersweet feeling of finishing a compelling book. These weekly features are designed to capture interest and entertain as well as facilitate an intuitive understanding of neologisms and neology among the wider public.

3. The state of neology research at the intersection with lexicography

Neology, as a theoretical discipline concerned with the creation and use of new words, has a well-established tradition, particularly in countries such as France. Its origins can be traced to the work of Darmesteter (1887), whose pioneering research on compounding laid the foundation for the systematic study of neolo-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

gisms. In the mid-20th century, Matoré (1952) broadened the scope of the field by incorporating historical and sociological perspectives, emphasising the influence of social and cultural dynamics on lexical innovation.

The discipline gained further prominence in the 1970s, when scholars built upon Matoré's lexicological approach while also investigating neology from a lexicographic perspective. A key contribution was made by Rey (1976), who provided critical insights into the lexicographic treatment of neologisms. Concurrently, Guilbert (1975) established a framework for a comprehensive and interdisciplinary study of neology, integrating synchronic and diachronic perspectives, along with sociolinguistic and grammatical considerations. Quemada (1967, 1971) also made important contributions in the neology field, focusing on the evolution of lexical innovation and the methodological challenges of integrating new words into lexicographic practice, though without being able to establish such a comprehensive approach as did Guilbert or Rey.

Beyond France, the study of neology also advanced significantly, particularly in the United States. Bolinger, a professor of Romance languages at Harvard University, played a crucial role in promoting neological research. As the first editor of the neology supplement *Among the New Words*, published in the journal *American Speech* (Bolinger 1941) from 1941 onward, he systematically documented emerging words in English. This initiative was later continued by John Algeo (Algeo and Algeo 1991), further reinforcing the importance of tracking lexical innovation in real-world usage.

By incorporating multiple perspectives — lexicological, lexicographic, sociolinguistic, and historical — neology has evolved into a robust and interdisciplinary field, reflecting the dynamic nature of language and its continuous adaptation to cultural and technological changes.

However, it was not until the first decade of this millennium that neology showed clear signs of establishing itself as a discipline: 2007 saw the birth of *Neologica*[9] (Humbley and Sablayrolles 2007), the first journal specialised in neology, and 2008 that of CINEO,[10] the first International Congress of Neology in Romance Languages, organised by Teresa Cabré and the IULATERM team[11] at Universitat Pompeu Fabra in Barcelona. *Neologica* is published annually in France, containing articles on neology from a wide range of perspectives and occasionally including works with a lexicographic connection. For example, Humbley (2008) analyses the functions of the borrowing dictionary, and Pruvost (2009) analyses the neologism mark in the general dictionary, and Issue 8 (Humbley and Sablayrolles 2014) devoted the thematic part to 'Neology and Dictionary'. As for CINEO, none of its iterations so far (including the sixth due in Coimbra in 2025)[12] have featured explicitly the intersection with lexicography (as done with many other disciplines, such as translation, terminology, literature, or digital communication), though the topic did occur randomly (cf. Vincent 2024). The focus on French neological studies is explained by theoretical as well as practical considerations. From a theoretical point of view, French language research in neology has a long history and rests on a relatively consensual methodologi-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

cal framework, based on structuralism at its origins with application to major language planning projects, particularly in Québec. From the practical point of view, the French publishing house Lambert-Lucas agreed to grant ENEOLI access to six recent monographs on neology in French. This starter corpus has been used to provide a small but relatively homogenous sample of basic contemporary neology terminology.

Researching the connection between neology and lexicography is of a recent nature, and this intersection is experiencing a growth that is observable in both disciplines. David K. Barnhart (1985) reflected on the difficulty for dictionary editors to know which new words will remain in use and should be recorded, an issue that has generated one of the most fruitful lines of research in the space shared by neology and lexicography, as will be seen below. Neology, on the other hand, has approached lexicography with theoretical questions such as whether the neological character of a word ends once it appears in a general dictionary of a language, or regarding the need to create dictionaries of new words.

The first international event entirely focused on this topic has recently emerged in the form of Globalex Workshop on Lexicography and Neology (GWLN), co-organised by Annette Klosa-Kückelhaus and Ilan Kernerman. It was first held in 2019 as a once-off workshop in conjunction with the 22nd Biennial Meeting of the Dictionary Society of North America, in Bloomington, Indiana (Klosa-Kückelhaus and Kernerman 2020),[13] and sponsored by the Global Alliance for Lexicography (Globalex).[14] Its success, however, turned it into a series of workshops co-located annually with a regular conference of one of the continental associations for lexicography. In its six iterations so far, GWLN has been held on five continents (and one held online due to the COVID-19 pandemic), bringing together specialists combining this dual field of knowledge from all over the world and for different languages and lexicographic traditions. In addition, each iteration has been complemented by a corresponding publication (such as the current paper, which was the highlight of GWLN-6, held in conjunction with the 28th International Conference of the African Association for Lexicography, in Pretoria, South Africa)[15] in an international lexicography journal or conference proceedings.

The contrast and similarity of experiences of African languages such as Lusoga or Swahili, with those of Asian ones such as Chinese or Japanese, or with various European languages present in the different iterations of GWLN, have promoted an important advance in the global study and knowledge of the current phenomena of lexical innovation and its lexicographic investigation and representation. Most of this work is grouped around major lines of research, which are briefly presented below (without aiming to be exhaustive, as it is a field that opens diverse possibilities), considering features such as the size of the language, the time of permanence in the use of neologisms, the passage from a specialised unit to a general language unit, and many more.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

(a) Firstly, there are works (such as Sheidlower 1995) that analyse the presence or treatment of neologisms in a particular dictionary, usually a general reference dictionary of the language. In some studies, the analysis focuses on a specific type of neologisms (such as semantic verbal neologisms, in Berri and Boschiroli 2024, for Spanish), compares the treatment of neologisms in different lexicographic works of the same language (such as Rødningen and Karlsen 2020, for Norwegian), or refers specifically to online dictionaries (such as Porath 2019, for Hebrew). Within this line, works that analyse the presence or treatment of loanwords in dictionaries (such as Humbley 2011) deserve a separate mention. In this case, these are usually English borrowings, and various aspects are analysed or compared (such as pseudo-anglicisms in Modern Greek dictionaries, in Vacalopoulou 2024) or from different perspectives (such as, from a lexicographic diachronic perspective for Korean, in Jun and Jung 2022).

(b) Another connection between lexicography and neology explores the development of dictionaries of neologisms, since they present important specificities in comparison with general lexicography, due to their obsolescent character; neologisms in a dictionary of neologisms are doomed to disappear either by their incorporation into the general dictionary (and the consequent loss of their neology status) or by disappearance from use. Ološtiak and Rešovská (2021) for Slovak, and Krvina (2020) for Slovenian, are examples of work along this line.

(c) The criteria for the inclusion or exclusion of neologisms in dictionaries is another one of the lines with the most contributions, and we devote Section 4 to it.

As mentioned above, the greatest challenge for lexicographers facing the updating of general dictionaries is to determine which of the new words, senses, and polylexical expressions that appear in social usage will eventually become part of the language and, therefore, should be included in the general dictionary. Of course, this depends largely on the function of the dictionary and that language's lexicographic tradition; the lexicography of some languages is more open to the introduction of new words, while the approach in other languages could be more conservative, but at some point, the lexicographic team is faced with the decision of which lexical innovations to include and which to leave out. Logic dictates that neologisms that will become established in usage should be included, whereas those that will prove to be temporary should be excluded, so in the end, the lexicographer must predict to some extent the future of new words.

This quantification of the possibility of survival of a neologism would be useful for the decision-making of lexicographers to determine the inclusion and exclusion of words in the dictionary, but it is also per se a line of research for neology in the attempt to predict the life of words. This approach originated with Renouf (2007), with the attempt to identify the different stages of devel-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

opment of a neologism, and it is gaining importance with more computational proposals through work on lexical emergence (cf. Grieve et al. 2016).

What is certain is that reflection on the criteria for the inclusion and exclusion of neologisms for lexicographic updating has evolved and expanded considerably since the pioneering work of Barnhart (1985), Algeo (1993), and Sheidlower (1995), among others, in which it can be seen that it is an interest shared by lexicographers and linguists. With the arrival of the new millennium, it was Ishikawa (2006) and O'Donovan and O'Neill (2008) who gave a boost to the debate (in addition to the aforementioned work by Metcalf 2002), always from the English language perspective, but a few years later it is also true for other languages, such as Korean (Nam et al. 2016) or Spanish (Adelstein and Freixa 2013), among others. Since the last decade, this line of research has blossomed with works such as those of Klosa-Kückelhaus and Wolfer (2020), the Neómetro team for Spanish (Bernal et al. 2020), and the proposal to automate the criteria for Catalan (Freixa 2022b).

Neology and lexicography are two disciplines whose convergence has proved in recent years to be a mutually fruitful area of knowledge, in both applied and theoretical fields. As will be seen in Section 4, this area of knowledge is one of the strengths of the ENEOLI network.

4. The lexicographic dimension in ENEOLI

4.1 Developing a multilingual glossary of the meta-terminology of neology

One of ENEOLI's most impactful contributions lies in its integration of neology research into lexicographic practices. By addressing the challenges of documenting lexical innovation across languages, ENEOLI bridges a critical gap between neology and lexicography, offering tools, methodologies, and theoretical insights that advance both fields. At the core of this effort is the development of the Multilingual Glossary of Neology, a pioneering resource tailored to meet the diverse needs of academic researchers, translators, lexicographers, and other language professionals.

Developed within WG1, the glossary will be a multilingual online resource built upon corpora meticulously curated for the project, such as NeoCorpus. These corpora are primarily constructed from systematically collected sources, including books, journal articles, conference proceedings, dissertations, and linguistic specialised dictionaries. They encompass a wide range of themes, such as theoretical principles of neology, lexical innovation across languages, neology in specialised domains, and language-related phenomena (e.g., purism, borrowing, and terminology management). Additionally, the corpora incorporate research on the lexicographic treatment of neologisms and methodologies for detecting lexical innovations, combining automated and manual approaches. This comprehensive foundation ensures that the glossary captures the evolving nature of neology while maintaining its relevance for practical applications.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

WG1 adopted a methodological approach grounded in the systematic compilation and analysis of a starter corpus. This corpus was built from six French monographs, selected for their consistent and homogeneous approach to neology, provided by Lambert-Lucas Editions. The corpus formed the basis for semi-automatic term extraction using AntConc and Sketch Engine, tools that facilitated the identification of high-frequency candidate terms. These terms were subsequently validated by a panel of experts in French neology and WG1 members, ensuring their relevance and accuracy. This collaborative validation process culminated in a provisional list of monolexical terms, documented in ENEOLI Deliverable 1.1 (Silva et al. 2024).

The extracted and validated terms were catalogued and integrated into NeoVoc. This controlled vocabulary provides a structured framework for linking concepts and their multilingual equivalents, thereby supporting the project's multilingual and comparative objectives. The enriched metadata from NeoVoc is stored in NeoCorpus, which is further augmented with references collected in a Zotero library, enabling complex cross-linguistic analyses and the tracking of term usage over time. Each term was assigned a unique concept identifier within the Wikibase system. At first, these concept entities contained only French equivalents, serving as the foundation for further multilingual expansion.

To support this expansion, an automated import process was carried out, retrieving existing multilingual equivalents from Wikidata and other external databases. These imported equivalents were then subjected to a manual verification process to ensure their accuracy and consistency: (i) if correct, the equivalent was retained and validated by a designated expert; (ii) if incorrect or inconsistent, the equivalent was edited or replaced, ensuring alignment with the validated concept; and (iii) if missing, new equivalents were manually added by linguistic specialists. The verified and enriched data from NeoVoc that is stored in NeoCorpus is continuously updated with validated terms and multilingual equivalents. This structured integration enables real-time updates, improving the accuracy of linguistic representation and ensuring the glossary remains dynamic and up to date.

To enhance user accessibility and experience, a dedicated task focuses on the development and implementation of the data representation and presentation framework for the Multilingual Glossary of Neology. The goal is to create a platform that enables users to access, explore, and understand terminology used in the field of neology across different languages efficiently. This requires a robust user interface that supports seamless navigation, multiple languages, and efficient information retrieval. Collaboration with the team responsible for the editorial style guide (T1.4) ensures alignment between data structure, terminology presentation, and formatting rules, maintaining consistency between editorial guidelines and visual design. Moreover, close coordination with the editing team (T1.6) guarantees that updated and edited content is accurately and promptly integrated into the platform.

Future phases of the project will focus on expanding multilingual equivalents, validating polylexical terms, and enriching NeoCorpus. Through con-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

tinued methodological refinement and collaborative input, ENEOLI WG1 is establishing a benchmark for the creation of dynamic, comprehensive, and multilingual lexicographic resources.

The platform's implementation will leverage an existing solution, such as FAIRterm 2.0,[16] a flexible, open terminology management system tailored for collecting and managing concept-oriented entries. Designed for describing concepts in specialised domains and their linguistic designations from a multilingual perspective, FAIRterm 2.0 adheres to the latest ISO TC 37[17] standards for terminology management. This tool aligns with the FAIR principles, which means that the data is findable, accessible, interoperable, and reusable (Vezzani 2022). Its features will support the compilation, organisation, and dynamic updating of the glossary, creating a comprehensive and user-friendly resource. This tool offers several features that enhance the compilation, organisation, and updating of the glossary. These include an intuitive web-based interface for collaborative editing, structured data fields for multilingual concept entry management, semi-automated validation processes to ensure terminological consistency, and API integration for interoperability with other lexicographic and linguistic resources. Additionally, it supports version control, user permission management, and export functionalities in standard formats, making it a comprehensive and user-friendly resource for terminology management.

Teams of researchers, lexicographers, and domain experts contribute to glossary development through standardised workflows and shared digital tools. The use of advanced computational methods, including machine learning and LLOD, supports efficient term extraction, validation, and integration. Together, these efforts ensure that the ENEOLI glossary not only serves as a valuable resource for documenting lexical innovation but will also set a benchmark for creating multilingual, dynamic lexicographic platforms.

Organising the editorial work involves task-specific teams, standardised workflows, and shared tools such as ENEOLI Wikibase. Formats and technologies used include metadata standards for interoperability, and web-based editing platforms.

4.2 Neology from a lexicographic perspective

Within WG2, T2.5 compares how different lexicographic institutes and projects (a) select, (b) describe, and (c) follow up lexical innovations in their databases and dictionaries. The focus is on the editorial decisions that dictionary editorial boards and lexicographers make when dealing with neology phenomena and how that decision-making is informed by the data, tools, and other resources at their disposal.

The main objective of T2.5 is the compilation of a publicly available database with a systematic overview of editorial practices in the lexicographic treatment of neology. Additionally, the results will be disseminated through presentations at conferences and publications.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

Over the course of the Action, the activities are thematically organised around three phases that can be distinguished in the lexicographic treatment of neology:

1. selection of lexical innovations for inclusion in the lexicographic resource, based on inclusion and exclusion criteria;
2. description of the lexical innovations using neology-specific microstructure components;
3. follow-up of lexical innovations after they have been included in the resource, with possible updates.

The workflow of T2.5 dedicates one year of work to each of these three phases, starting in the first year with the selection phase. During its first year, the task has been charting the inclusion and exclusion criteria used by different lexicographic resources and in different languages. The concrete aim is to compile an ENEOLI typology of selection criteria for neologisms, which can serve as an overview and point of reference for lexicographers and the developers of lexicographic tools. Work started off with a look at the most elaborate typology found in the literature, developed by Freixa (2022a, 2022b) as state-of-the-art for Catalan. Then other perspectives were added, stemming from different traditions and languages, including Danish (*Den Danske Ordbog*), German (Storjohann 2024), Korean (Nam and Jung 2020), Spanish (Sánchez Ibáñez and Maroto 2021), French (*Le Petit Robert*), and Portuguese (*Dicionário da Língua Portuguesa*). After this fact-finding stage, the task will concentrate on integrating the criteria into a single typology that will be made publicly available in the second half of 2025.

5. Conclusion

The ENEOLI network represents a major advancement in the study of lexical innovation, offering a comprehensive, multilingual, and collaborative approach to understanding language change. By addressing key challenges in the field, it is not only contributing to the development of valuable resources but also fostering a deeper and more nuanced understanding of how language evolves in a globalised world. Through its multifaceted perspective, which encompasses specialised lexicography, digital resources, comparative studies, and didactics, the network is paving the way for innovative research and practical applications that will shape the future of neology and lexicography. This holistic approach ensures that both theoretical insights and practical tools are available to researchers, practitioners, and educators, reinforcing the importance of a collaborative and interdisciplinary effort in advancing the field.

The main outcome aimed for is a specialised glossary centred on terms related to neology and lexical innovation. This glossary contains a meta-terminology derived from corpus data, as well as expert input and validation processes. The methodology used is one typically associated with terminological

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

and specialised lexicography research. By combining different approaches, the glossary will serve as a practical resource while also advancing the understanding of how neology is approached within lexicography, both in practice and in theory.

Consistent with the principles of all COST Actions,[18] ENEOLI is open to researchers, practitioners, and other stakeholders from participating countries, fostering interdisciplinary collaboration and knowledge exchange across Europe and beyond. Designed to be inclusive, these Actions enable participants at various career stages and from diverse backgrounds to engage in networking, training, and joint research activities. The ENEOLI website provides a contact form, contact information of network leaders, and direct links to further guidance on how to join, ensuring accessibility for those interested in contributing to the network's activities.

Acknowledgements

The authors sincerely thank the anonymous referees for their valuable comments that improved the quality of this manuscript. We also extend our gratitude to the dedicated task leaders, whose expertise and commitment have been instrumental to the success of this COST Action.[19]

The research of Rute Costa and Ana Salgado is supported by the FCT — Portuguese Foundation for Science and Technology, I.P. as part of the project UID/03213 — Linguistics Research Centre of NOVA University Lisbon (CLUNL). The research of Federica Vezzani is also part of the initiatives carried out by the Centre for Studies in Computational Terminology (CENTRICO) of the University of Padua. The research of Špela Arhar Holdt is supported by the ARIS research programme Language Resources and Technologies for Slovene (P6-0411).

Endnotes

1. This paper is based on work from COST Action ENEOLI, CA22126, supported by the European Cooperation in Science and Technology (COST), https://www.cost.eu/actions/CA22126/.
2. Recent initiatives aim to bridge gaps in research and practice through workshops, conferences, and publications, such as the Globalex Workshop on Lexicography and Neology (GWLN) that has been held annually since 2019 in conjunction with the international conference of one of the continental lexicography associations. The sixth iteration of GWLN, held at the University of Pretoria in conjunction with the Afrilex 28th International Conference in July 2024, highlighted ENEOLI. https://globalex.link/gwln-6_2024/.
3. https://eneoli.eu [28 February 2025].
4. https://www.laurenceanthony.net/software/antconc/ [1 December 2024].
5. https://www.sketchengine.eu [1 December 2024].
6. https://eneoli.wikibase.cloud/wiki/Main_Page [27 March 2025].
7. https://www.zotero.org/groups/5449136/neocorpus/library [28 February 2025].
8. https://www.cost.eu/about/strategy/excellence-and-inclusiveness/ [1 December 2024].
9. https://classiques-garnier.com/neologica.html [1 December 2024].

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

10. https://iula.upf.edu/agenda/cineo_08/cineo01fr.htm [1 December 2024].
11. https://www.upf.edu/web/iulaterm [1 December 2024].
12. https://ucpages.uc.pt/events/cineo2025/ [1 December 2024].
13. https://dictionarysociety.com/conference/dsna-22-in-bloomington-indiana/ [1 December 2024].
14. https://globalex.link/ [1 December 2024].
15. https://globalex.link/gwln-6_2024/ [1 December 2024].
16. https://shiny.dei.unipd.it/fairterm/compilation20.html [1 December 2024].
17. https://www.iso.org/committee/48104.html [1 December 2024].
18. https://www.cost.eu/cost-actions/what-are-cost-actions/ [28 February 2025].
19. We would like to mention all their names, following the order of the four WGs: Raquel Silva, David Lindemann, Margarida Ramos, Onorina Botezat, Anna Vacalopoulou, Jelena Kallas, Kristina Koppel, Hiwa Asadpour, Anca Gâta, Neslihan Önder-Özdemir, Voula Giouli, Elisa Squadrito, Fahad Khan, Alina Villalva, Sara Pita, Petra Storjohann, Besim Kabashi, Jean-Louis Vaxelaire, Esther Breuer, Radka Mudrochová, Andrzej Napieralski, Sabela Fernández Silva, Sorina Ciobanu, Miguel Sánchez Ibáñez, Daniela Hasa, and Etleva Kondi.

References

Dictionaries

***Den Danske Ordbog*.** Det Danske Sprog- og Litteraturselskab. https://ordnet.dk/ddo [1 December 2024]

***Dicionário da Língua Portuguesa*.** Academia das Ciências de Lisboa. Salgado, A. (Ed.). Lisboa. https://dicionario.acad-ciencias.pt/ [1 December 2024]

***Le Petit Robert de la langue française*.** Dictionnaires Le Robert-SEJER. https://www.lerobert.com [1 December 2024]

Other literature

Adelstein, A. and J. Freixa. 2013. Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Presentation at the International Congress *El diccionario: neología, lenguaje de especialidad, computación, 2013, Mexico City, Mexico, 28–30 October 2013.*

Algeo, J. 1993. Desuetude among New English Words. *International Journal of Lexicography* 6(4): 281-293.

Algeo, J. and A.S. Algeo. 1991. *Fifty Years among the New Words: A Dictionary of Neologisms 1941–1991.* New York: Cambridge University Press.

Barnhart, D.K. 1985. Prizes and Pitfalls of Computerized Searching for New Words for Dictionaries. *Dictionaries* 7: 253-260.

Bernal, E., J. Freixa and S. Torner. 2020. Criterios para la diccionarización de neologismos: de la teoría a la práctica. *Revista Signos* 53(104): 592-618.

Bolinger, D.L. 1941. Among the New Words. *American Speech* 16(2): 144-148.

Berri, M. and V. Boschiroli. 2024. Lexicographic Representation of Verbal Semantic Neology in Spanish Dictionaries. *6th Globalex Workshop on Lexicography and Neology (GWLN-6), 3 July 2024, University of Pretoria, Pretoria, South Africa.*

Darmesteter, A. 1887. *La vie des mots étudiée dans leurs significations.* Paris: Delagrave.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

Freixa, J. 2022a. The Dictionarisation of Neologisms: The NADIC as a Model. Freixa, J., M.I. Guardiola, J. Martines and M.A. Montané (Eds.). 2022. *Dictionarization of Catalan Neologisms*: 15-37. Berlin: Peter Lang.

Freixa, J. 2022b. Garbell: l'avaluador automàtic de neologismes catalans (Garbell: The Automatic Catalans Neologisms Analyser). *Terminàlia* 26: 7-16.

Grieve, J., A. Nini and D. Guo. 2016. Analyzing Lexical Emergence in Modern American English Online. *English Language and Linguistics* 21(1): 99-127.

Guilbert, L. 1975. *La créativité lexicale*. Paris: Librairie Larousse.

Humbley, J. 2008. Le dictionnaire d'emprunts : fonctions descriptives et prospectives. *Neologica* 2: 55-73.

Humbley, J. 2011. Le traitement des anglicismes dans un dictionnaire français et allemand. Steuckardt, A., O. Leclercq, A. Niklas-Salminen and M. Thorel (Eds.). 2020. *Les dictionnaires et l'emprunt (16è–21è siècle)*: 219-233. Aix-en-Provence: Publications de l'Université de Provence.

Humbley, J. and J.-F. Sablayrolles (Eds.). 2007. *Neologica. Revue internationale de néologie* (No. 1). Paris: Garnier.

Humbley, J. and J.-F. Sablayrolles (Eds.). 2014. *Neologica. Revue internationale de néologie* (No. 8). Paris: Garnier.

Humbley, J. and A. Salgado. 2024. *Report 1 — WG1 Ice Breaking Survey. CA22126 — European Network On Lexical Innovation (ENEOLI)*.
https://eneoli.eu/wp-content/uploads/2025/05/ENEOLI_WG1_Report1.pdf

Ishikawa, S. 2006. When a Word Enters the Dictionary: A Data-Based Analysis of Neologism. JACET Society of English Lexicography (Ed.). 2006. *English Lexicography in Japan:* 39-52. Bunkyo-ku: Taishukan.

Jun, C. and H.-Y. Jung. 2022. On Loans in Korean New Word Formation and Lexicography. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs and P. Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX Euralex International Congress, EURALEX 2022, Mannheim, Germany, 12–16 July 2022*: 814-824. Mannheim: IDS-Verlag.

Klosa-Kückelhaus, A. and I. Kernerman (Eds.). 2020. Global Viewpoints on Lexicography and Neologisms. *Dictionaries. Journal of the Dictionary Society of North America* 41(1): 1-10.

Klosa-Kückelhaus, A. and S. Wolfer. 2020. Considerations on the Acceptance of German Neologisms from the 1990s. *International Journal of Lexicography* 33(2): 150-167.

Krvina, D. 2020. The Newest Slovene Lexis in the Growing Dictionary of Slovene Language: Lexicological and Lexicographical Perspectives. *2nd Globalex Workshop on Lexicography and Neology (GWLN-2), 4 November 2020, held in conjunction with the EURALEX 2020 Congress at Alexandroupolis, Greece.*

Matoré, G. 1952. Le néologisme: naissance et diffusion. *Le français dans le monde* 20(2): 87-92.

Metcalf, A.A. 2002. *Predicting New Words: The Secrets of Their Success.* Boston: Houghton Mifflin.

Nam, K., S. Lee and H.-Y. Jung. 2020. The Korean Neologism Investigation Project: Current Status and Key Issues. *Dictionaries: Journal of the Dictionary Society of North America* 41(1): 105-129.

Nam, K., S. Lee, H.-Y. Jung and J. Choi. 2016. The Life and Death of Neologisms: On What Basis Shall We Include Neologisms in the Dictionary? Margalitadze, T. and G. Meladze (Eds.). 2016. *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity,* Ivane Javakhishvili Tbilisi State University, *Tbilisi, Georgia, 6–10 September 2016:* 389-393. Tbilisi: Ivane Javakhishvili Tbilisi State University.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2039 (Lexicofocus)

O'Donovan, R. and M. O'Neill. 2008. A Systematic Approach to the Selection of Neologisms for Inclusion in a Large Monolingual Dictionary. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, Spain, 15–19 July 2008*: 571-579. Barcelona: Institut Universitari de Lingüística Aplicada — Universitat Pompeu Fabra.

Ološtiak, M. and S. Rešovská. 2021. A Dictionary of Slovak Neologisms (A Project Description). *International Journal of Lexicography* 34(3): 302-314.

Porath, N. 2019. Adding Neologisms to the Hebrew Online Dictionary *Rav-Milim*. *1st Globalex Workshop on Lexicography and Neology (GWLN), 8 May 2019, Indiana University, Bloomington, Indiana.*

Pruvost, J. 2009. La marque 'néol' dans le dictionnaire. Retour sur un dictionnaire de synonymes et d'analogies de la fin du XXe siècle. *Neologica* 3: 53-66.

Quemada, B. 1967. *Les dictionnaires du françois moderne. Études sur leur histoire, leurs types et leurs méthodes.* Paris: Didier.

Quemada, B. 1971. À propos de la néologie. Essai de délimitation des objectifs et des moyens d'action. *La Banque des mots* 2: 137-150.

Renouf, A. 2007. Tracing Lexical Productivity and Creativity in the British Media: The Chavs and Chav-Nots. Munat, J. (Ed.). 2007. *Lexical Creativity, Texts and Context*: 61-89. Amsterdam: John Benjamins.

Rey, A. 1976. Néologisme: un pseudo-concept? *Cahiers de Lexicologie* 28: 3-17.

Rødningen, D. and K.E. Karlsen. 2020. The Treatment of Neologisms in Two Norwegian Dictionaries. *2nd Globalex Workshop on Lexicography and Neology (GWLN-2), 4 November 2020, held in conjunction with the EURALEX 2020 Congress at Alexandroupolis, Greece.*

Sánchez Ibáñez, M. and N. Maroto. 2021. Beyond Timelines: The Challenges of Combining Theoretical Premises and Speakers' Insights about the Assessment, Validation and Inclusion of Spanish Neologisms in Dictionaries. *International Journal of Lexicography* 34(3): 358-381.

Sheidlower, J.T. 1995. Principles for the Inclusion of New Words in College Dictionaries. *Dictionaries* 16: 33-44.

Silva, R., A. Salgado and J. Humbley. 2024. *ENEOLI Deliverable 1.1: Results of the Candidate Terms Validation Process — Monolexical Terms. CA22126 — European Network On Lexical Innovation (ENEOLI).* https://eneoli.eu/wp-content/uploads/2025/04/ENEOLI_Deliverable1.1_08.04.25_FINAL.pdf

Storjohann, P. 2024. IDS-*Neo*$^{2020+}$: A Novel Resource for New German Words in Use. *International Journal of Lexicography* 37(4): 389-403.

Vacalopoulou, A. 2024. Digging for *-ings*: A Survey of Selected (Pseudo)Anglicisms in Dictionaries of Modern Greek. *6th Globalex Workshop on Lexicography and Neology (GWLN-6), 3 July 2024, University of Pretoria, Pretoria, South Africa.*

Vezzani, F. 2022. *Terminologie numérique: conception, représentation et gestion.* Bern: Peter Lang.

Vincent, N. 2024. Des néologismes qui dérangent. Analyse de débats médiatiques sur le militantisme de lexicographes du Robert. Rossi, M. (Ed.). 2024. *Néologie, terminologie et variation / Neología, terminología y variación / Neologia, terminologia e variazione*: 153-169. Lausanne: Peter Lang.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

Neologisms and Their Functions in Critical Discourse*

Petra Storjohann, *Leibniz-Institut für Deutsche Sprache, Mannheim, Germany*
(*storjohann@ids-mannheim.de*)
(*https://orcid.org/0000-0001-8752-0874*)

Abstract: The contributions of new words to the construction of a discourse are understood as discourse functions. For example, neologisms introduce new ways of thinking by embodying novel conceptual knowledge and influencing collective thought processes. They play a crucial role in distinguishing emerging social practices by clarifying vague concepts and laying the groundwork for further lexical innovation, such as through word formation. Within crisis-related discourse, neologisms become fundamental in shaping linguistic reality, acting as pivotal nodes within a structured network. They have the potential to expose ideological content and cultural values, as these terms can highlight key issues in debates. Terms such as *Cancel Culture*, *Wokeness*, and *Greenflation* capture a range of views and beliefs in societal dialogues across ideological, cultural, and political spectrums. However, when neologisms are documented lexicographically, their roles in discourse are not specified. This paper demonstrates how the key discourse aspects highlighted by new vocabulary are systematically documented in a recently developed German neologism dictionary. Clearly, the novel resource seeks to transcend the constraints of previous lexicographical methods by adopting specific elements of discourse lexicography (Kämper 2006: 350). This paper also explores whether new methods can be utilised for such an endeavour.

Keywords: NEOLOGISM RESOURCE, DISCOURSE LEXICOGRAPHY, CRITICAL DISCOURSE ANALYSIS, DISCOURSE PATTERNS, DISCOURSE FUNCTIONS, WORD EMBEDDING

Opsomming: Neologismes en hul funksies in die kritiese diskoers. Die bydraes van nuwe woorde tot die konstruksie van 'n diskoers word as diskoersfunksies verstaan. Neologismes lei, byvoorbeeld, tot nuwe denkwyses deur nuwe konseptuele kennis te vergestalt en kollektiewe denkprosesse te beïnvloed. Hulle speel 'n deurslaggewende rol in die karakterisering van opkomende sosiale praktyke deur vae konsepte te verhelder en die grondslag vir verdere leksikale innovasie, soos d.m.v. woordvorming, te lê. Binne krisisverwante diskoers word neologismes noodsaaklik in die vorming van 'n linguistiese werklikheid deur as sentrale punte binne 'n gestruktureerde netwerk op te tree. Neologismes beskik oor die potensiaal om ideologiese inhoud en kulturele waardes bloot te lê, aangesien hierdie terme belangrike kwessies in debatte kan uitlig. Terme soos "Cancel-Culture", "Wokeness", en "Greenflation" vang 'n reeks sienings en oortuigings in samelewings-

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

gesprekke oor ideologiese, kulturele en politieke spektrums vas. Wanneer neologismes egter leksikografies gedokumenteer word, word hul rolle in diskoers nie gespesifiseer nie. Hierdie artikel demonstreer hoe die sleutelaspekte van diskoers wat deur nuwe woordeskatitems beklemtoon word, sistematies in 'n onlangs ontwikkelde Duitse neologismewoordeboek gedokumenteer word. Dit is duidelik dat dié nuwe hulpbron poog om die beperkings van vorige leksikografiese metodes te oorkom deur spesifieke elemente van die diskoersleksikografie (Kämper 2006: 350) oor te neem. Hierdie artikel verken ook die moontlikheid dat nuwe metodes hiervoor aangewend kan word.

Sleutelwoorde: NEOLOGISMEHULPBRON, DISKOERSLEKSIKOGRAFIE, KRITIESE DISKOERSANALISE, DISKOERSPATRONE, DISKOERSFUNKSIES, WOORDINBEDDING

1. Introduction: Critical discourse and neologisms

In addition to the core semantic features that constitute the meaning of a neologism, the study of new words can, in certain cases, uncover argumentation strategies, ideological perspectives, evaluative dimensions, sentiment, discriminatory elements, and the dynamics of (dis)empowerment. Currently, public communication is increasingly shaped by socio-political discourses dominated by crises. Consequently, neologisms that have emerged since 2020 exhibit a strong correlation with crisis-related issues, often reflecting underlying concepts that are contested and revealing ambivalent attitudes.

With respect to German, newly borrowed terms such as *Cancel Culture*, *Wokeness*, *E-Fuel*, and *Greenflation* encapsulate a diverse array of views and ideologies prevalent in public debates. During the COVID-19 pandemic, numerous emotionally charged compounds emerged, including *Maskenmuffel* (referring to an individual who, despite official mandates, refuses to wear a face mask in public spaces), *Covidiot* (a derogatory term for someone exhibiting inappropriate behaviour), and *Querdenker* (describing a person who opposes widely accepted measures to contain the pandemic, as well as other related government regulations, often motivated by political and ideological convictions).

Emotion-laden terms also feature prominently in debates surrounding the climate movement, with labels such as *Klimaextremist* ('climate extremist') and *Klimaterrorist* ('climate terrorist') frequently used in a derogatory manner to describe individuals perceived as disruptive, harmful, or extreme in their environmental activism. These expressions reflect distinct ideological, cultural, and political positions. More broadly, a substantial number of new lexical items, meanings, and phrases have emerged, contributing significantly to the discourse surrounding key contemporary issues, including the climate crisis, COVID-19, the war in Ukraine, the supply chain crisis, the energy crisis, inflation, and immigration. These can be referred to as neo-discourses (Mattfeldt et al. 2025).

In this paper, critical discourse is understood as a social practice in which meaning, positions, and arguments are continuously negotiated through communicative acts. Neologisms can embody specific discourse properties, such as argumentation strategies, divergent perspectives, and sentiment, which can be

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

systematically identified. Critical Discourse Analysis employs an applied, usage-based approach grounded in corpus linguistics. As Baker et al. (2013: 20) note, this approach "examines how ideologies and power relations are expressed in language". Among others, prominent proponents of this method include Partington (2004), Baker (2006), Teubert (2006), Baker et al. (2013), and Collins and Baker (2023). These scholars share a focus on vocabulary in use, treating individual lexical items as discourse objects and analysing the lexical patterns that define the immediate context of a given term (Van Dijk 2008). Through the use of corpora, connections between text and lexis can be uncovered, exposing social structures by revealing their linguistic representation in lexical patterns.

In the field of Critical Discourse Analysis, neologisms have received relatively limited attention, with the focus typically placed on individual keywords. Only a few studies have addressed neologisms in detail, such as Teubert (2002), who explores the use of *globalisation*, and Koller et al. (2019), who examine the term *Brexit* from multiple perspectives. However, it is both individual neologisms and clusters of new lexical items that play a crucial role in constructing discourse patterns, representing arguments, and shaping communicative foci. These elements can be systematically unfolded and described both semantically and lexicographically.

This paper pursues three primary objectives. Firstly, it aims to demonstrate the close interconnection between neology and critical discourse by shifting the focus from isolated analyses of individual neologisms to a broader examination of discourse-related clusters of new words and their roles within public debates and discussions. Secondly, it explores complementary methodological approaches to investigating these phenomena. Finally, the paper proposes lexicographic strategies for documenting neo-discourses and their associated terms, with a particular focus on a newly developed German online dictionary designed to capture neologisms that have emerged since 2020, reflecting key concepts and different facets in contemporary discourse. This paper is therefore also an appeal for a more reflective approach as pursued by Critical Lexicography (Chen 2019).

2. The case of *Greenflation*: Depicting the complexity of meaning and discourse

Corpus-Assisted Discourse Studies (CADS) is a methodological approach that integrates techniques from corpus linguistics with discourse analysis. It involves the use of large corpora and computational tools to analyse linguistic patterns, frequencies, and structures (Partington et al. 2013). CADS enables researchers to uncover implicit meanings, ideologies, and social attitudes within discourse by examining language usage across diverse texts and contexts. This approach is particularly effective for identifying trends, comparing discourses, and providing empirical evidence of linguistic and social practices. By examining textual patterns and constructions, researchers can identify common representations of specific subjects. For instance, Baker et al. (2013) provide an in-depth analysis

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

of how the term *Muslim* is typically used in context to describe Muslims and how these representations recur in the British press through particular phrases and constructions, illustrating the advantages of corpus-based analysis. Similarly, the present study examines the term *Greenflation* as represented in DeReKo, the German reference corpus. Beyond analysing collocations and syntagmatic patterns, recent discourse studies have increasingly adopted a new vector-based method of word embedding (Bubenhofer 2022), which serves as a tool for identifying lexical-semantic relationships at the paradigmatic level. This paper argues that both methods should be employed complementarily to investigate discourse, neologisms, and their meanings, roles, and functions.

2.1 Collocation analysis: Syntagmatic level

The study of collocations represents a well-established methodology across various linguistic disciplines, providing a means to conduct systematic semantic analyses of common linguistic patterns. This approach is essential for exploring the contextual environments of terms and, consequently, for elucidating their characteristic features as demonstrated in actual usage.

> The meaning attributes of a node's collocates can provide a helpful sketch of the meaning/function of the node within the particular discourse. (Baker et al. 2008: 278)

Using the example of ***Greenflation***, which arose in the context of the energy crisis in Germany around 2021, the analysis of contextual patterns within its related discourse sheds light on communicative signals. The extraction and interpretation of collocates not only reveal important semantic and syntactic structures but also highlight essential discourse properties. In Example 1, numerous indicators of discursiveness can be observed in the immediate vicinity of the term. *Greenflation* is a key term (*Schlagwort*) that evokes concerns (*Befürchtungen*) about rising prices (*Inflation*) caused by the energy transition (*Energiewende*).

(1) Ein aktuelles Schlagwort ist "*Greenflation*", also die Befürchtung, dass die Energiewende die Inflation anheizt. Teilen Sie diese Sorge? (*Frankfurter Rundschau*, 11.02.2022)

(Translation: A current key term is "greenflation", referring to the concern that the energy transition is driving up inflation. Do you share this concern?)

Typical co-occurrences and keywords can be organised according to various criteria, based on their semantic preferences and/or syntactic features. Among others, this organisation (see Table 1) encompasses qualitative interpretations that facilitate the identification of properties and domains with overlapping themes, discourse markers, and evaluative terms and positions.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

Table 1: Collocates functioning as lexical indicators for discourse properties, with terms with asterisks being neologisms

Semantic-discursive category	**Collocate**
discourse markers	*Stichwort* ('headword'), *Phänomen* ('phenomenon'), *sogenannte* ('so-called'), *Buzzword* ('buzz word'), *Schlagwort* ('key term')
environmental contexts	*Klimakrise* ('climate crisis'), *Klimawandel* ('climate change'), *Umwelt* ('environment'), *entstehen* ('develop'), *CO2-Emissionen* ('CO2-emissions'), *erneuerbare Energien* ('renewable energies'), *Energiewende* ('energy transition'), *Klimawende* ('climate transition'), *Klimaschutz* ('climate protection'), *CO2-freier Strom* ('CO2-free energy')
events: war in Ukraine	*Putin, Krieg* ('war'), *Freiheitsenergie** ('freedom energy'), *Ukrainer* ('Ukrainians'), *Ukrainekrise* ('Ukraine crisis'), *Ukrainekrieg* ('war in Ukraine')
topic: energy sector	*Gaskrise* ('gas crisis'), *Energiekrise* ('energy crisis'), *Gasknappheit* ('gas shortage'), *Gasmangellage* ('gas shortage'), *Energieknappheit* ('energy shortage'), *Climateflation** ('climateflation'), *LNG-Terminal* ('LNG terminal')
economic contexts	*Kosten* ('costs'), *Teuerungswelle, Inflation* ('inflation'), *Preise* ('prices'), *Strompreis* ('electricty costs'), *grüne Inflation** ('green inflation'), *Energiepreis* ('energy price'), *Gaspreis* ('gas price'), *Markt* ('market'), *Wirtschaft* ('economy'), *Lieferkettenengpass* ('supply chain shortage'), *Preisexplosion* ('price explosion')
positions/attitudes	*Ökonomen* ('economists'), *Politiker* ('politicians'), *Fachleute* ('experts'), *Gewinner* ('winner'), *Verlierer* ('loser'), *Konsumenten* ('consumers'), *Debatte* ('debate'), *reden über* ('talk about'), *Heizungsstreit* ('heating debate'), *diskutieren* ('discuss'), *argumentieren* ('argue')
affected groups	*Unternehmen* ('businesses'), *Haushalte* ('households'), *Wirtschaft* ('economy'), *Deutschland* ('Germany'), *Europa* ('Europe')
expressions of concerns/responses	*bedrohen* ('threaten'), *Bedrohung* ('threat'), *sich wappnen* ('to arm oneself'), *Angst* ('fear'), *Albtraum* ('nightmare'), *Gefahr* ('danger'), *Sorge* ('worries'), *Probleme* ('problems')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

effects	*Folgen* ('results'), *Auswirkungen* ('effects'), *Effekt* ('effect')
political measures taken by government to fight costs	*Balkonkraftwerk** ('balcony power plant'), *Maßnahmen* ('measures'), *Energiepolitik* ('energy politics'), *Klimaschutzgesetze* ('climate protection laws'), *Gasumlage* ('gas levy'), *Energiepreisbremse** ('energy price cap'), *Gaspreisbremse** ('gas price cap'), *Gaspreisdeckel** ('gas price cap'), *Strompreisdeckel** ('power price cap'), *Strompreisbremse** ('power price cap')
political measures taken by government to save energy	*Energielockdown** ('energy lockdown'), *Gaspriorisierungsplan** ('gas priority plan'), *Gas-Triage** ('gas triage'), *Streckbetrieb** ('idle operation of power plant')

Semantic insights emerge from the interpretation of collocational and syntagmatic contexts, offering crucial information for defining the meaning[1] of *Greenflation* as follows:

> bezeichnet größere Preissteigerungen für Rohstoffe, die aufgrund des Übergangs von fossiler Energieerzeugung hin zu einer ökologischeren, möglichst CO2-neutralen Stromproduktion sowie der Förderung erneuerbarer Energien entstehen
>
> (Translation: refers to larger price increases for raw materials that arise due to the transition from fossil energy production to a more ecological, ideally CO2-neutral electricity generation, as well as the promotion of renewable energies)

Besides paraphrasing a new word, a closer examination of the categories derived from the collocational analysis reveals additional characteristics of its usage in public communication. The controversial nature of the term is inherent in its characterisation as a key term (*Schlagwort*). This brevity is illustrated by the verbs that frequently co-occur with the noun, such as talk about (*reden über*) and debate (*argumentieren*). Furthermore, the discussion surrounding *Greenflation* often frames it as a subject that threatens (*bedrohen*), suggesting that individuals must arm themselves (*sich wappnen*). Concerns and fears are also articulated through accompanying nouns such as *Angst* ('fear'), *Albtraum* ('nightmare'), *Gefahr* ('danger'), and *Sorge* ('worries'). Patterns such as *Die Angst vor der „Greenflation" geht um* ('The fear of greenflation is spreading') illustrate the prevalent perception of greenflation as a potential threat. This negative evaluation is further reflected in other co-occurring terms that can be classified into specific domains or topics and closely related discourses. The lexical item *Greenflation* is deeply embedded in the discourse surrounding the energy crisis (*Gaskrise*, *Energiekrise*).

However, discourse surrounding *Greenflation* exhibits significant intersections with other prevailing crisis-related narratives. These include the economic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

crises stemming from the COVID-19 pandemic, as well as inflation driven by increased consumer prices and rising energy costs. Furthermore, new terminologies are emerging in the context of the war in Ukraine (*Ukrainekrieg*), which is closely linked to escalating energy prices resulting from sanctions imposed on Russian gas and oil imports. Finally, the discourse surrounding the climate crisis (*Klimakrise*) is of great importance, particularly in the context of Germany's coalition government (12/2021–12/2024), which includes the Green Party. This coalition was actively working to implement new environmental legislation aimed at reducing emissions and promoting the use of alternative renewable energy sources to achieve independence from Russian oil and gas. All of the discourses identified in this context are crisis-related, and as such, they convey negative connotations regarding specific critical socio-political situations.

An examination of the public reports on *Greenflation* within the underlying corpus reveals the emergence of new concepts, particularly those related to political measures aimed at addressing rising energy prices or promoting energy conservation. These concepts are exemplified by several newly coined terms (indicated with an asterisk in Table 1), such as *Energiepreisbremse*, *Gaspreisbremse*, *Gaspreisdeckel*, *Strompreisdeckel*, *Strompreisbremse*, *Energielockdown* and *Streckbetrieb*. These highlight a notable lexical growth associated with the discourse surrounding the energy crisis, which contributes to the establishment of new discourse and knowledge structures. Generally, a large number of lexemes co-occur, priming sentiments and emotions when *Greenflation* is used contextually (see Example 2).

(2) Der Bremer Wirtschaftswissenschaftler Rudolf Hickel, ein Mann, der sowohl die Grünen als auch die Energiewende mit Sympathie betrachtet, ahnt Übles: „Die *Greenflation* kann noch erhebliche Probleme aufwerfen." Hickel verweist auf eine Addition von marktbedingten und politisch gewollten Preissteigerungen, die am Ende vor allem Geringverdiener und Transferbezieher belasten könnten. Zu den Kostensteigerungen bei Rohstoffen gesellten sich noch die bereits 2021 beschlossenen schrittweise wachsenden CO2-Bepreisungen. Die Gesamtwirkung sei in Berlin noch nicht hinreichend analysiert worden. „Ich habe die große Sorge." (*Döbelner Allgemeine Zeitung*, 08.01.2022)

(Translation: Bremen economist Rudolf Hickel, a man who views both the Greens and the energy transition with sympathy, senses trouble: "Greenflation could pose significant problems." Hickel points to a combination of market-driven and politically mandated price increases that could ultimately burden low-income earners and benefit recipients. In addition to rising raw material costs, the gradually increasing CO2 pricing decided in 2021 adds to the strain. The overall impact has not yet been sufficiently analysed in Berlin. "I have major concerns.")

Example 2 also highlights participants in public discourse. Nouns denoting individuals or groups, such as economists, politicians, experts, businesses,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

households, and consumers, all of which function as subjects or objects, express the concept of who is affected by or concerned with greenflation. Phrases such as *Gewinner und Verlierer* ('winner and loser') illustrate the controversy and disagreement surrounding the issue. Collocates not only enhance the meaning of *Greenflation* but also impart their connotations, which are shaped and contextualised through their shared environments. Nevertheless, this discursive markedness is typically not captured in dictionaries. The evaluation of new concepts and their corresponding terms by speakers cannot be readily found in lexicographic resources. However, this evaluation is evident in the numerous terms with negative connotations that convey effects, fears, concerns, and the various political actions being undertaken. Moreover, semantic categories, as illustrated in Table 1, which resemble frame-semantic slots and contribute to a word's topical and discursive appropriateness, are generally not documented in most reference works.

2.2 Word embedding: Paradigmatic level

By exploring corpora, researchers can systematically extract new words automatically (Klosa and Lüngen 2018) and categorise them into various thematic or semantic fields, domains, or discourses (still mostly manually). However, such lists do not provide insight into the relationships between words within a specific discourse or domain. Until recently, it has also been challenging to compare these words with semantically or conceptually related terms that exhibit varying degrees of similarity. Word embedding, a linguistic distributional method, visually represents lexical networks in a continuous vector space by capturing semantically similar words in the vicinity of a search term (Fankhauser and Kupietz 2019). This method addresses the aforementioned research gap by analysing structures not only at the paradigmatic level but also in relation to search terms based on their semantic proximity, as determined by the comparison of their collocation profiles. Its strength lies in its ability to identify commonalities in meaning, operating on the premise that words occurring in similar contexts tend to possess similar meanings (Harris 1954). In contrast to traditional collocation analysis, the advantage of word embedding is its capacity to detect and classify "unseen or out-of-context words" (Rudkowsky et al. 2018: 150).

This paper contends that syntagmatic collocation analysis should be integrated with a paradigmatic investigation of semantically related terms. Distributed word embedding facilitates the exploration of additional dimensions of discourse by identifying words with similar contextual patterns (e.g. synonym searching), recognising neighbouring terms with comparable connotations and extracting clusters of associated words or phrases. Such methods can uncover latent themes or topics within a discourse and identify potential layers of sentiment or media tone, as demonstrated in analyses of press texts (see Rudkowsky et al. 2018).

Recently, the application of word embedding techniques has gained traction in the field of discourse studies (e.g. Bubenhofer 2022). In this context, vis-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

ualised word vectors facilitate the analysis of meaning by comparing words or phrases based on their semantic similarity, aiding in the exploration of how meanings shift or cluster within various discourses. Word embedding also enables the identification of central themes or concepts within a text by analysing the proximity of word vectors. Additionally, the clustering of words or sentences based on their embedding allows for the identification of distinct segments or topics within a discourse. For contrastive analyses, word embedding can reveal how cultural or ideological perspectives shape word usage and meaning by comparing embeddings across different corpora. Thus, integrating word embedding into discourse studies offers researchers deeper insights into the structure, meaning, and evolution of discourse, establishing it as a valuable tool for linguistic and communicative analysis.

With regard to German, DeReKoVecs is a word embedding tool designed to analyse the underlying data within the DeReKo corpus. The objective is to ascertain whether a close interrelationship exists among the various discourses identified through the collocates of the term *Greenflation*. Specifically, the investigation focuses on the term *Energiekrise* ('energy crisis') as the superordinate and lexical representation of the discourse of which *Greenflation* is a part (Figure 1 below).

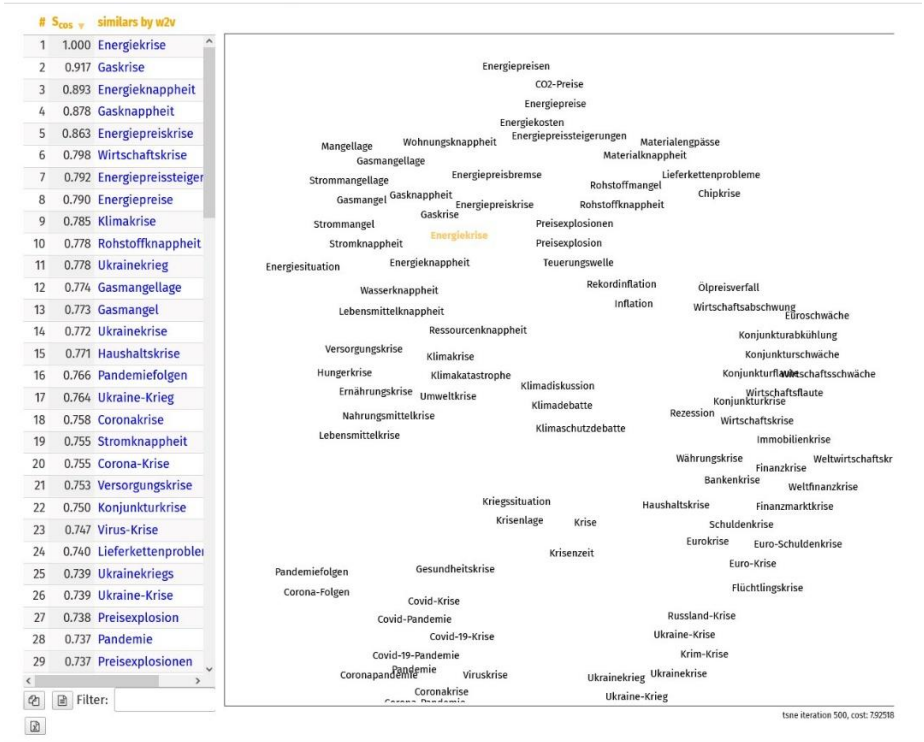


Figure 1: Word embedding of *Energiekrise* ('energy crisis')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

The term *Energiekrise* is closely associated with terms that specifically delineate the concept of the energy crisis itself. Among these are words indicating a critical shortage of gas, such as *Gasknappheit* ('gas shortage'), *Gaskrise* ('gas crisis'), *Gasmangel* ('gas shortage'), and *Gasmangellage* ('gas shortage situation'). The escalating costs resulting from this shortage are represented by terms such as *Preisexplosion* ('price explosion'), *Energiepreiskrise* ('energy price crisis'), *Energiekosten* ('energy costs'), *Energiepreise* ('energy prices') and *Teuerungswelle* ('wave of inflation'). Additionally, neighbouring terms such as *Klimadiskussion* ('climate discussion'), *Klimadebatte* ('climate debate'), *Klimaschutzdebatte* ('climate protection debate'), *Klimakrise* ('climate crisis'), and *Umweltkrise* ('environment crisis') suggest that the discourse surrounding the energy crisis parallels that of the climate crisis. Furthermore, the analysis uncovers a dimension not previously explored through collocation analysis: the supply chain crisis that emerged due to the reduction of gas and oil deliveries from Russia to Germany. This includes terms such as *Rohstoffmangel* ('shortage of raw materials'), *Lieferkettenprobleme* ('supply chain problems'), *and Materialknappheit* ('shortage of raw materials'). The vocabulary associated with the war in Ukraine is also prominently represented by terms such as *Ukrainekrieg* ('war in Ukraine'), *Ukraine-Krise* ('Ukraine crisis'), *Krim-Krise* ('Crimea crisis'), and *Russlandkrise* ('Russia crisis'). As pointed out before, *Greenflation* is framed as a consequence of the German energy crisis, arising from both green politics and the ambition to reform the energy sector, as well as the war in Ukraine, which has resulted in the cessation of gas and oil imports from Russia. In this context, semantic prosody tends to be evaluative and negative, often "extending over more than one unit in a linear string" (Stubbs 2001: 65). In this regard, also see Example 3:

(3) Das Jahr 2022 war für uns nicht ganz einfach: Wegen Klimakrise, Kriegskrise, Inflationskrise, Energiekrise und Coronakrise hatten wir nichts zu lachen. (*Neue Welt*, 28.12.2022)

(Translation: The year 2022 was not easy for us: Due to the climate crisis, the war crisis, the inflation crisis, the energy crisis, and the corona crisis, we had nothing to laugh about.)

Additional terms related to various crises include those linked to the COVID-19 pandemic (*Covidkrise*), the economic and financial crisis (*Wirtschaftskrise*), the food crisis (*Lebensmittelkrise*), the migration crisis (*Flüchtlingskrise*), and the Euro crisis (*Eurokrise*). Each of these terms denotes recent or ongoing challenges that demonstrate semantic overlaps with the concept of the energy crisis. Notably, crisis-related terminology is often emotionally charged.[2] The dense network of discourses surrounding *Greenflation*, as evidenced by collocation results, is further corroborated by an examination of the overarching energy crisis discourse and its interconnectedness with other crisis-related discourses as illustrated by Figure 2. Figure 2 is a sketch entailing some lexical indicators retrieved

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

from both corpus-linguistic methods together with some of their interpreted abstract discourse categories.

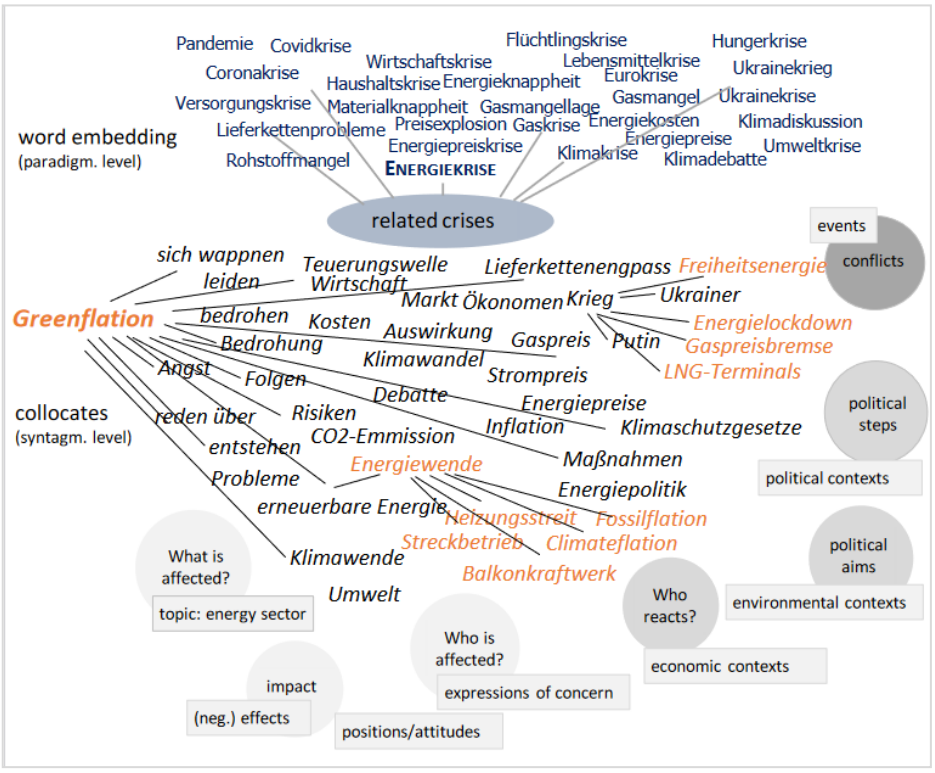


Figure 2: Illustrative sketch of retrieved corpus data and interpretation

In summary, both methods reveal patterns of negativity and emotionality. The neologism *Greenflation* frequently co-occurs with other evaluative collocates at a syntagmatic level. Several collocates that co-occur with *Greenflation* are newly introduced terms (marked in orange), which denote alternative forms of inflation (e.g. *Climateflation* and *Fossilflation*) or new political measures to fight rising prices. Collectively, these terms signify the emergence of new discourse strands centred on a shared topic of debate. As a central component of the discourse surrounding the energy crisis in Germany, it is situated within a broader framework of semantically related terms, primarily crisis-related words that carry negative connotations also at a paradigmatic level. Most of these terms denote distressing and problematic events with socio-political implications, both nationally and globally. The prevalence of terms that express sentiment and negative connotations is thus additionally substantiated by the findings from the word embedding analysis.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

3. (Some) functions in discourse

In Critical Discourse Studies, discourse functions pertain to the roles that specific linguistic features assume within a given text or discourse, particularly in connection with power, ideology, and social relations (cf. Fairclough 1992). Discourse functions refer to the various communicative purposes that language serves in interaction. These include, for example, informative, expressive, interactional, persuasive, textual, and structuring functions (Wodak and Meyer 2015). Wodak and Meyer (2015: 9) point out:

> Discourse serves various functions in society: it is not only a means of transmitting information, but also a way of establishing, maintaining, and challenging social relationships, power structures, and ideologies.

Discourse functions facilitate the transmission of information, and the expression of emotions and personal opinions. Furthermore, lexical elements serving these functions can influence the behaviour of others, or structure argumentative narratives to persuade the audience.

By employing corpora and identifying patterns in language use, insights can be gained into these discourse functions and their influence on the construction of social meanings (cf. Baker 2006). These patterns exemplify communicative strategies that construct identities or positions, legitimise policies and shape public opinion. Certain neologisms can express a specific speaker's stance or attitude, and encapsulate evaluative dimensions concerning particular circumstances. These terms may refer to individuals or to concrete or abstract issues, manifesting either derogatory or positive connotations. Notably, words that have emerged since 2020 exhibit a significant degree of sentiment, as many are related to crises. These sentiments become intrinsic semantic-pragmatic features that affect the subject, domain, or discourse to which they pertain. All such contributions to the construction of critical discourse are understood as discourse functions.

An examination of the term *Greenflation* reveals how a newly coined term can carry substantial discursive and communicative significance at a particular time or within a specific context. This term emerged during a period of energy transition around 2021, driven by two significant events. Firstly, Germany's reliance on Russian gas and oil was disrupted due to the cessation of deliveries resulting from the war in Ukraine. In some segments of the public, the end of gas and oil imports from Russia is often perceived as the primary cause of rising prices. Additionally, right-wing populism fosters pro-Russian sentiments and opposes European sanctions against Russia, resulting in resistance to German political decisions that support Ukraine. Secondly, German politics have been actively reforming the energy sector by increasing the utilisation of renewable sources and phasing out nuclear power and fossil fuels. This transforma-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

tion necessitated the acquisition of alternative energy resources within a rapidly inflating global market.

The term *Greenflation* can be employed to delegitimise certain actions, policies, or beliefs, often by invoking authority, morality, or rationality, particularly in relation to green political initiatives (see Example 4).

(4) Der Übergang in eine grünere, CO2-neutrale Weltwirtschaft treibt die Preise und die Inflation an. Fachleute sprechen von „*Greenflation*". Für den starken Inflationsschub mögen noch pandemiebedingte Unterbrechungen der Produktion und globale Logistikprobleme verantwortlich sein. Aber das erklärt nicht die stark gestiegenen Preise für Energie (Gas, Öl, Kraftstoff und Strom). Emissionshandel und CO2-Abgabe zeigen Wirkung. Teurere Energie ist einerseits politisch gewollt, um von den fossilen Energien wegzukommen, andererseits machen sich auch Angebotsverknappungen in den Förderstaaten bemerkbar. (*Gießener Anzeiger*, 02.12.2021)

(Translation: The transition to a greener, CO2-neutral global economy is driving up prices and inflation. Experts refer to this as "greenflation". While production disruptions due to the pandemic and global logistics problems may be responsible for the strong surge in inflation, this does not explain the sharp rise in prices for energy (gas, oil, fuel, and electricity). Emissions trading and CO2 taxes are having an impact. On the one hand, more expensive energy is politically desired in order to move away from fossil fuels; on the other hand, supply shortages in producing countries are also becoming noticeable.)

Consequently, the concept designated by *Greenflation* is used to challenge the authority of politicians, especially those in the coalition government (12/2021–12/2024) and the Green Party. As a result, *Greenflation* has acquired immediate significance in public discourse, contributing to the portrayal of essential knowledge structures associated with this transformation process. These structures encompass the expansion of renewable energy sources and the increased implementation of solar, wind, and geothermal power. Additionally, energy storage has emerged as a focal area, with discussions surrounding battery technologies and other solutions to address the intermittency of renewable energy. The gradual phase-out of fossil fuels, carbon pricing, political measures, awareness campaigns, and issues related to research and innovation all contribute to shaping these new knowledge structures. The contextualisation of such events or issues in terms of time and space can significantly influence their perception and understanding (e.g. as immediate versus long-term, or local versus global). As a loanword, *Greenflation* is framed both as a local issue that generates socio-economic challenges and as a global concern, given the international nature of the energy market. Within the context of *Greenflation* and the energy crisis, these topics are extensively reported in the news, featured in newspapers, and debated in tele-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

vision discussions. This underscores the potential of neologisms to introduce new modes of thinking by encapsulating novel conceptual knowledge. In the case of *Greenflation*, this implies that the transformation of the energy sector may lead to rising costs rather than decreasing prices, thereby influencing and shaping collective thought processes.

Greenflation may be regarded as an inevitable outcome of green politics. Consequently, the term plays a vital role in revealing ideological content, cultural values, and a spectrum of views and beliefs within societal discourse across the political landscape. There are both proponents and opponents of the political strategies that give rise to alternative energy productions and rising costs, as observed among politicians, media representatives, scientists, economists, and climate activists. Controversial positions intensify discussions, particularly regarding the socio-economic impact on low-income households and industries that are heavily dependent on energy for production. These effects contribute to the perception of the concept of greenflation as predominantly destructive, embedding negative connotations of the term as such associations are readily activated. Thus, its semantic prosody is evaluative, frequently reflecting the speaker's stance, which is typically negative. *Greenflation* is framed within a context of crisis and threat, mirroring ideological positions and shaping public opinion. Overall, *Greenflation* assigns responsibility for social problems to specific social groups and influences public perception. In this context, language evokes emotional responses that can be instrumental in shaping attitudes and behaviours.

Greenflation is a good example to show how a term is associated with a range of closely interconnected events related to current political decisions, inflation, the climate crisis, the war in Ukraine, sanctions, and the energy crisis. This illustrates how one discourse frequently draws upon and interlinks multiple discourses (e.g. economic, political, and environmental) to construct meaning and shape understanding. *Greenflation* functions as both an effect and a central node that connects the discourses of the climate crisis, the energy crisis, and the war in Ukraine. In this capacity, new terms can serve as pivotal nodes within a structured discourse network, bridging crucial public debates. As emerging nodes, they contribute to the structuring process of discourse. Once these neologisms become more established in the lexicon, they lay the groundwork for further lexical innovation, as evidenced by word formations such as *Greenflation-Warner* ('greenflation skeptics') and *Greenflation-Faktoren* ('greenflation factors'). These developments, in turn, initiate a new cycle of establishing new or specific discourse properties.

In CADS, discourse functions are analysed to reveal the underlying power dynamics, ideologies, and social structures that both shape and are shaped by language. By investigating these functions, researchers can enhance their understanding of how discourse contributes to the construction of social reality and the maintenance or contestation of power. While only some important discourse properties found in the contexts of *Greenflation* were sketched here, it is essential to recognise that not all neologisms inherently exhibit discourse functions

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

as described here. The roles of new words as individual items within specific contexts, as well as their relationships with other new terms and established words in the lexicon, warrant further examination within CADS. *Greenflation* and its associated discursive potential are integral to its semantics. Accordingly, this paper posits that lexicographic descriptions should take this into account.

4. Discourse neologisms and discourse lexicography

Within the lexicographic landscape of German reference works, discourse dictionaries ("Diskurswörterbücher") have established themselves as a significant component of specialised lexicography. This branch is dedicated to the systematic documentation and analysis of controversial keywords and phrases that play a pivotal role in particular public discourses. A primary objective of discourse lexicography is to shed light on the dynamic interplay between language and society. It illustrates how words and expressions not only convey meanings but also reflect and shape social attitudes and power structures. Kämper (2006: 350) highlights „die Gebundenheit des Wortschatzes in topikalischen, sprecherbezogenen, textlichen, zeitlichen und funktionalen Hinsichten" (translation: the relevance of vocabulary in topical, speaker-related, textual, temporal, and functional respects), necessitating the examination of relevant texts.

To date, discourse dictionaries have traditionally predominantly focused on the treatment of headwords in various reference works including *Brisante Wörter von Agitation bis Zeitgeist* (Strauß et al. 1989), *Zeitgeschichtliches Wörterbuch: Schlüsselwörter und Orientierungsvokabeln* (Stötzel and Eitz 2003), *Opfer — Täter — Nichttäter: Ein Wörterbuch zum Schulddiskurs 1945–1955* (Kämper 2007), *Diskurswörterbuch zum Protestdiskurs 1968* (Kämper 2012), and *Wörterbuch zum Demokratiediskurs 1967/68* (Kämper 2013). Through discourse analysis, discourse lexicography can enhance the understanding of the mechanisms by which language constructs and influences society. This is particularly relevant in times of heightened political conflict, during which language often serves as a tool for shaping public opinion. Thus, the discourse dictionaries available today (e.g. Kämper 2007, 2012, 2013) reflect critical keywords related to the language of the Third Reich and illustrate new cultural values expressed through the lexis of the student and social movements of 1968, which were integral to a broader global wave of protests and radical political activism during that era.

So far, German discourse lexicography has predominantly adopted a historical approach. Additionally, these reference works have not yet integrated methodologies from corpus linguistics and discourse analysis to examine the usage and significance of lexical units within diverse social, political, and cultural contexts. Concerning critical neologisms, the treatment provided by the aforementioned dictionaries remains unexplored. As previously demonstrated, a substantial number of newly coined words in German possess discursive properties that significantly contribute to their respective discourses. A portion of their mean-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

ing is intricately linked to their communicative impact, encompassing attitudes, emotions, beliefs, and personal opinions. Many new terms related to crises cannot be adequately described semantically without incorporating discourse features into their definitions. Analysing controversial terms enables users to comprehend how and why certain neologisms arise within specific social or political contexts, as well as their roles within discourse. Additionally, it allows for an exploration of how neologisms both reflect and influence societal changes or emerging trends. These objectives have not been addressed in earlier neologism dictionaries in German, such as the *Neologismenwörterbuch* (2006ff.).

The new German neologisms guide IDS *Neo*$^{2020+}$ includes two new aspects.[3] Firstly, the descriptions of headwords incorporate discourse features when pertinent. Lexicographically, it embraces key principles of discourse lexicography as articulated by Kämper (2006) in order to be able to shed light on various discourse positions instead of representing generalised stances on new concepts (Chen 2019). A particular challenge of lexicographic descriptions is the attempt to sensitise users about explanations of an item's use as not being prescriptive or stable. As Chen (2019: 377) points out "it is stated that interpretations and explanations are never finished and authoritative, but that they are dynamic and open to new contexts and new information".

A critical approach to discourse and lexicography reflects on the multitude of ideological perspectives and depicts them to illuminate beliefs, attitudes, and power structures in language and society. Although this resource is not explicitly designed as a discourse dictionary for neologisms, sections devoted to discursive aspects are inherently linked to the broader documentation of meaning and usage. Secondly, the new dictionary examines lexical items through complementary corpus methodologies, ensuring methodological consistency with other studies conducted by CADS. By integrating these discourse details, users can become attuned to the contentious applications of new words and recognise the significance of discourse properties as semantic elements of a headword.

In summary, as well as exploring meaning and contextual use, the documentation of neologisms that are integral to contemporary critical discourses should seek to elucidate the intricate interactions among language, power, and society. This approach offers valuable insights into the ways in which language use both influences and is influenced by discourses.

5. *Greenflation* in IDS *Neo*$^{2020+}$

5.1 Headword-related details

In addition to standard information regarding meaning, grammar, synonyms, and examples, each entry in the new neologism dictionary includes a section titled "Background Information". This section is dedicated to extra-linguistic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

information, emphasising encyclopaedic and discourse knowledge. It provides insights into the historical context, illustrating the relationship between significant events and the emergence of new words, accompanied by an interactive timeline. Users can also access information about the domains or preferential topics associated with the headword, as well as the relevant discourse(s) surrounding it. Furthermore, this section offers explanations of observed controversial viewpoints, cultural biases, evaluative arguments, and discernible power dynamics as revealed in media reports, all of which are supported by corpus examples (see Figure 3)[4].

More specifically, the entry informs users that the headword is used in contexts similar to those of *Fossilflation* and *Climateflation*. Additional information includes the representation of *Greenflation* across various text types, such as press articles, social media, and online communications, highlighting the differing thematic focuses within these formats. This section of the dictionary also delineates thematic domains, including POLITICS, ECONOMICS, ENVIRONMENT/ENERGY, and CLIMATE CHANGE. It explicitly identifies the relevant discourses, encompassing the ENERGY CRISIS, CLIMATE CRISIS, and WAR IN UKRAINE. Observations regarding the usage of *Greenflation* in public discourse note its contentious nature and the diversity of perspectives surrounding it. Two distinct stances along with sentiment descriptions, stylistic choices, evaluations, or argumentative frameworks are explicitly documented. On the one hand, the discussion surrounding the pricing and social impacts of *Greenflation* is frequently framed in the context of arguments put forth by opponents of energy transition initiatives and green policies. On the other hand, proponents advocate sustainability and emphasise the imperative of transforming the energy sector to address climate change. These two opposing perspectives are presented alongside one another to highlight the incompatible discourse positions. Corpus examples substantiate the descriptions.

Overall, the information presented in Figure 3 refers to a single headword, offering a focused depiction of single lexical items and their associated communicative contexts. To meet additional consultation needs, such as identifying all new terms associated with a specific discourse (e.g. ENERGY CRISIS), a second reference module has been developed: the IDS Dashboard.[5]

5.2 Details for lexical fields

In addition to offering detailed information on headwords, a key objective of the new resource is to present a range of content accessible through a user-friendly dashboard that supports browsing. Employing a masonry-style layout, the various sections integrate information about the project, the dictionary, and neologisms more broadly, making use of multiple formats such as interactive graphs, reports, podcasts, videos, and campaigns (see Storjohann 2024).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

Figure 3: Encyclopaedic information including details on discourse for entry *Greenflation*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

To provide an overview of discourses and their associated neologisms, rather than focusing solely on individual words, one section of the dashboard is dedicated to recent public debates characterised by communicative controversy or significant public response. This section highlights various new lexical bundles that contribute to shaping different discourses and frequently represent newly emerging concepts. Numerous approaches and options exist for describing discourses in their entirety. In IDS *Neo*$^{2020+}$, users can explore the discourses currently dominating public communication, with representative neologisms displayed for quick reference. Prior to engaging with a specific discourse section, the terms that lexically represent these topical fields are presented alongside their frequency trends since their emergence in a comparative format, such as *Energiekrise* ('energy crisis'), *Ukrainekrieg* ('War in Ukraine'), and *Klimakrise* ('climate crisis') in Figure 4.[6]

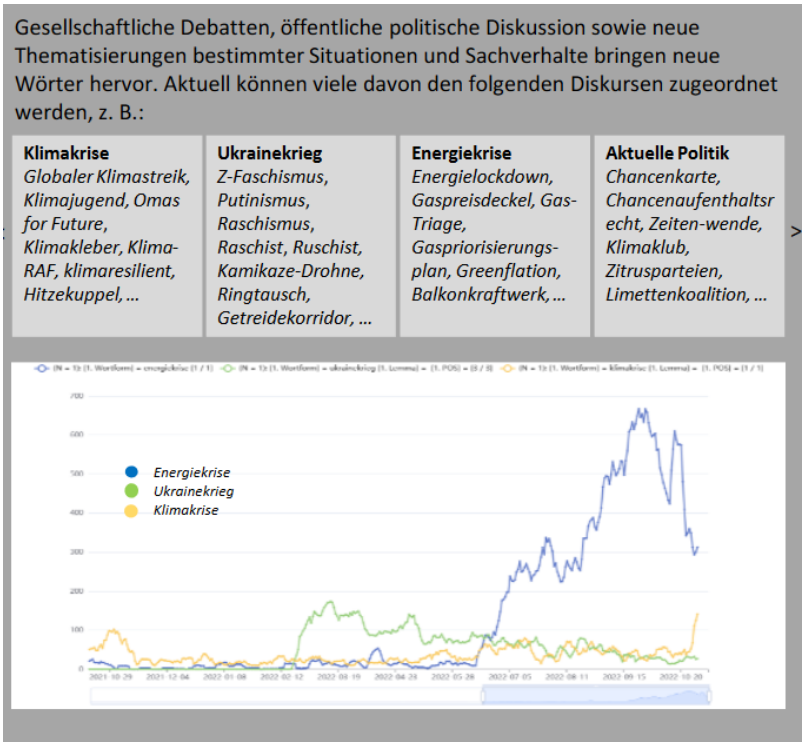


Figure 4: Overview of discourses and designating terms together with frequency graphs

Upon selecting the tile of a specific discourse, users are directed to a new interface that enables the exploration of contemporary discourses in conjunction with newly introduced lexical items. Through a flexible menu, users can customise

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

various display options, organising the information alphabetically (see Figure 5), chronologically, by frequency, by subdomain, or as a network.[7]

Figure 5 presents a synthesis of frequency graphs alongside alphabetical ordering, thereby providing a comparative visual depiction of the most frequently encountered lexical items associated with *Energiewende* ('energy transition'), *Gaspreisdeckel* ('gas price caps'), and *Strompreisdeckel* ('energy price caps'). This interface includes frequency graphs that compare some of the most prominent terms with their respective discourses.

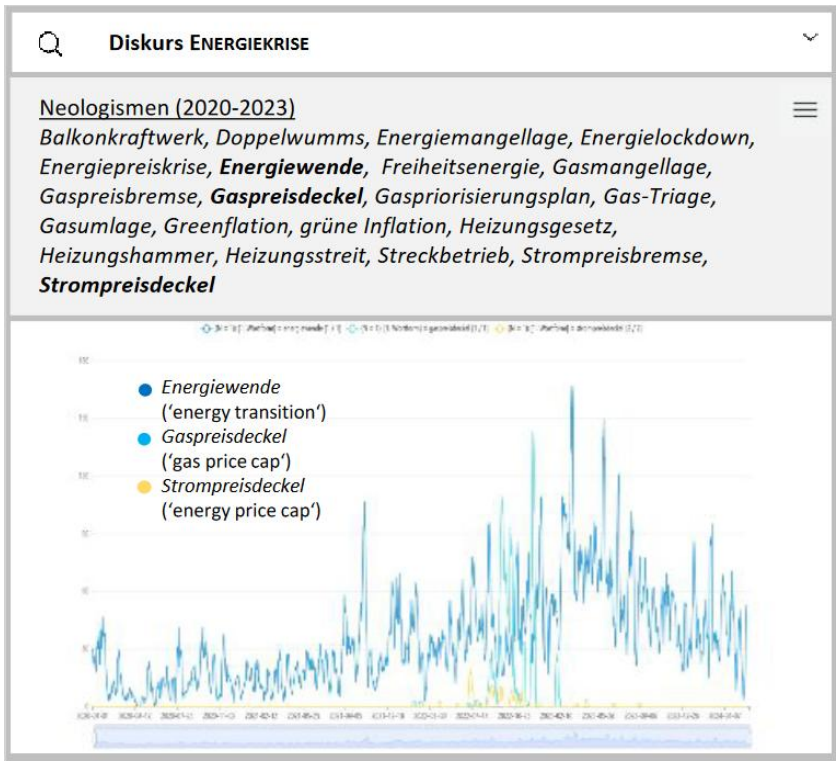


Figure 5: List of new terms belonging to the discourse of the energy crisis in alphabetical order

Moreover, users are afforded the opportunity to gain an understanding of the temporal evolution of the discourse via a timeline that highlights the emergence of neologisms in everyday language, ideally accompanied by corpus examples that validate their early usage (see Figure 6). Concise contextual descriptions of the overarching discourse — encompassing socio-political significance, diverse perspectives, arguments, and public reception — may precede the listing of headwords.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

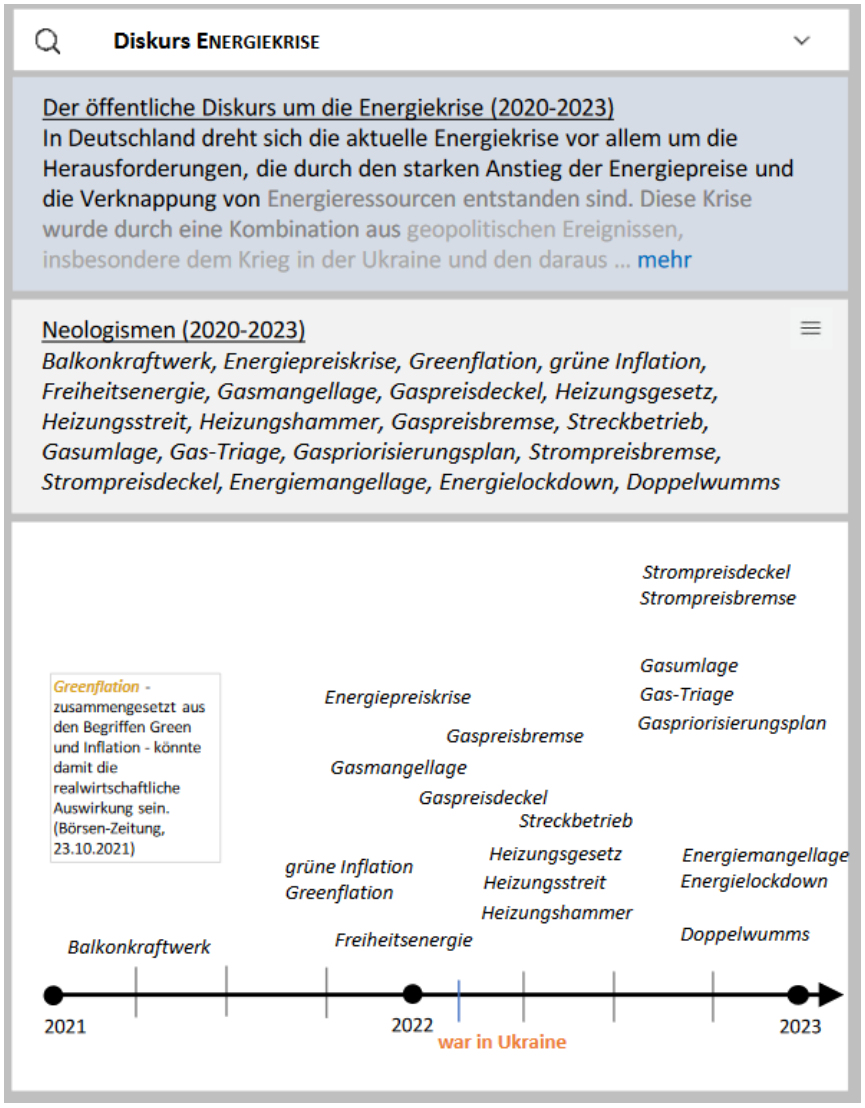


Figure 6: Chronological emergence of new words within a discourse together with short a description and corpus examples

The advantages of these descriptions enhanced by various visualisations are as follows. Firstly, they provide valuable insights into the broader thematic context. The presentation of clusters of neological discourse-related terms, along with various features and usage patterns, elucidates developments within specific segments of the vocabulary. Secondly, they heighten users' awareness of ongoing lexical changes and the domains in which these changes are currently

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

manifesting. Finally, through the concurrent display of numerous lexemes, users can identify areas where a proliferation of new concepts — and, consequently, new knowledge structures — has emerged within a particular domain or thematic subject. In this manner, some discourse functions of neologisms are indirectly illustrated. As previously noted, the documentation of entire discourses complements traditional entries that focus on individual headwords.

6. Limitations of discursive descriptions in IDS *Neo*$^{2020+}$

In the field of Critical Discourse Studies, discourse is commonly defined as a socially constructed system of communication that extends beyond linguistic elements and structures. It encompasses the ways in which language is employed in real-world contexts to construct, reflect, and challenge power dynamics, ideologies, and social practices (Wodak 2004). This communicative system can be analysed through a corpus, which consists of a collection of texts that share a common topic and exhibit intertextual references. According to Gardt (2007: 28), key features of discourse include its specific communicative framework and intentions, as well as the construction of reality through communicative actions. Discourse is a form of social practice through which knowledge and concepts are collectively negotiated and shared. It also plays a crucial role in the construction of identities and the formation of shared social knowledge, norms, values, attitudes, and ideologies.

In particular, crisis-related neo-discourses are highly complex. New words that are controversial or emotionally charged, and are integral to such critical neo-discourses, encapsulate multifaceted elements of this intricacy within their semantics. In section 3, the various discourse functions serve as a clear indication of this complexity.[8] The analysis of new words extends beyond lexical-semantic approaches as described here in 2.1 and 2.2. The study of contemporary neo-discourses requires an examination of both linguistic elements (e.g. vocabulary, grammar, and sentence structure) and social practices (e.g. speech acts and interactions). Therefore, discourse analysis integrates these aspects to provide a comprehensive understanding of how communication operates at a deeper level.

While corpus-extracted patterns and typical usage structures provide a systematic means of accessing key aspects of discourse such as speakers' perspectives, beliefs, principles, and sentiments, Critical Discourse Studies in fact employs a range of linguistic approaches to analyse how language constructs, sustains and challenges social power relations. For instance, at the grammatical level, studies examine features such as modality and passivisation. Rhetorical analyses focus on framing, argumentation, and presupposition, while pragmatic approaches investigate speech acts, exploring how functions such as promising, ordering, or apologising contribute to the reinforcement or contestation of power dynamics. These diverse analytical perspectives are not mutually exclu-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

sive but rather complementary, offering a more comprehensive understanding of discourse as a complex and highly dynamic phenomenon.

Lexicographic projects inherently face limitations in capturing a complete or comprehensive representation of all linguistic and contextual parameters involved. Nevertheless, the lexical approach to meaning reveals essential semantic features that can be documented lexicographically, contributing to definition-building and knowledge representation. This approach also provides insights into the socio-cultural and political contexts relevant to a given moment in a word's usage. Still, any attempt to document discourse within a dictionary remains fragmentary, just as any corpus represents only a portion of the broader discourse it reflects. However, crisis-related neologisms are embedded within discourse structures, where their meanings are shaped and negotiated. This relationship can be explicitly articulated and emphasised through lexicographic representation, offering valuable insights and heightening awareness of the intricate connection between language and society.

7. Conclusion: Neologisms and discourse functions

Crisis-related (critical) discourses are intricate and often intertwined, exhibiting a dynamic rather than static nature. Overall, complementary corpus approaches facilitate a deeper understanding of the semantic and discursive properties, structures, functions, and interrelationships within these discourses. Such methodologies aid lexicographers in representing the complexity of discourse, the knowledge and meanings negotiated by members of the discourse community, and the associations tied to new words. The emergence of new terms and their usage in discussing contemporary social realities shape discourses in a reciprocal manner, just as societal discussions influence the development of new words (Teubert 2010). While the discursive functions of new words may vary, their usage and roles within discourse significantly contribute to their meanings. Given their communicative immediacy, these new words and their meanings are of considerable interest to dictionary users and should be documented in neologism dictionaries in a manner that effectively captures the essential discourse properties reflected in each entry.

Until recently, experience in lexicographically representing the semantic complexity of new discourses and new words has been limited. To this end, the new IDS *Neo*$^{2020+}$ resource aims to integrate foundational approaches developed by CADS alongside ideas from discourse lexicography. This means that the new dictionary seeks to enhance this representation by incorporating essential new details regarding discourse, as observed in the underlying corpus. Although numerous methodological approaches to discourse analysis exist (see Jäger 2012, Wodak and Meyer 2015), the lexicographic approach adopted here emphasises a strong lexical focus, placing new lexical items in the centre of explorative discourse studies. These show that generally, neologisms are terms that denote novel concepts and fulfill various purposes and functions within critical discourses.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

While not all of these functions can be precisely documented in neologism dictionaries, it is imperative to include significant aspects of a term's meaning that involve discourse features and contextually defined elements within the lexicographic description. Such elements can be addressed to a certain extent, especially when their discursive functions affect meaning and usage. Specific entry rubrics and sections in the reference guide facilitate multi-level consultation combining lookup with browsing. Users can explore detailed information about headwords, access discourse-relevant information, and examine clusters of new words that represent emerging discourses and highlight various aspects of their characteristics. By providing a comprehensive and nuanced understanding of new words within their discursive contexts, the new resource aims to empower users to engage critically with evolving language practices and contribute to the ongoing dialogue surrounding contemporary social issues.

Endnotes

1. This definition is taken from the entry *Greenflation* of the new (as yet unpublished) dictionary IDS *Neo*$^{2020+}$.
2. In contrast to collocation analysis, this approach examines the paradigmatics of a word; consequently, it concentrates exclusively on nouns while excluding verbs and adjectives.
3. For a more detailed description of the new resource, see Storjohann (2024).
4. Figure 3 illustrates an entry taken from the lexicographic concept. The design of the entry in its final online format may differ upon publication.
5. The dictionary is scheduled to be launched in early 2025, while the dashboard is expected to be released in late 2025.
6. Figures 4 to 6 illustrate examples taken from the lexicographic concept. The design of the display in its final online format may differ upon publication.
7. In this paper, the discussion will be limited to the presentation of a select number of the aforementioned display options.
8. Discourse functions of neologisms are not explicitly marked as information in the dictionary.

References

Dictionaries

IDS *Neo*$^{2020+}$: *Neue Wörter im Gebrauch*. 2025. Mannheim: Leibniz-Institut für Deutsche Sprache. (to appear)

Kämper, H. 2007. *Opfer — Täter — Nichttäter. Ein Wörterbuch zum Schulddiskurs 1945–1955.* Berlin/New York: De Gruyter.

Kämper, H. 2012. *Diskurswörterbuch zum Protestdiskurs 1968.* Mannheim: Institut für Deutsche Sprache.

Kämper, H. 2013. *Wörterbuch zum Demokratiediskurs 1967/68.* Berlin: Akademie-Verlag.

Neologismenwörterbuch. 2006ff. *OWID — Online Wortschatz-Informationssystem Deutsch.* Mannheim: Leibniz-Institut für Deutsche Sprache.
http://www.owid.de/wb/neo/start.html [30 January 2025]

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

Stötzel, G. and T. Eitz (Eds.). 2003. *Zeitgeschichtliches Wörterbuch der deutschen Gegenwartssprache: Schlüsselwörter und Orientierungsvokabeln.* Hildesheim: Olms.

Strauß, G., U. Haß and G. Harras. 1989. *Brisante Wörter von Agitation bis Zeitgeist*. Berlin/New York: De Gruyter.

Digital tools

DeReKo (Deutsches Referenzkorpus).
https://www.ids-mannheim.de/digspra/kl/projekte/korpora/ [30 January 2025]

DeReKoVecs (word embedding tool).
https://corpora.ids-mannheim.de/derekovecs [30 September 2024]

Other references

Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

Baker, P., C. Gabrielatos and T. McEnery. 2013. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Baker, P., C. Gabrielatos, M. KhosraviNik, M. Krzyżanowsk, T. McEnery and R. Wodak. 2008. A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press. *Discourse & Society* 19(3): 273-306.

Bubenhofer, N. 2022. Exploration semantischer Räume im Corona-Diskurs. Kämper, H. and A. Plewnia (Eds.). 2022. *Sprache in Politik und Gesellschaft: Perspektiven und Zugänge*: 197-216. Berlin/Boston: De Gruyter.

Chen, W.G. 2019. Towards a Discourse Approach to Critical Lexicography. *International Journal of Lexicography* 32(3): 362-388.

Collins, L. and P. Baker. 2023. *Language, Discourse and Anxiety*. Cambridge: Cambridge University Press.

Gardt, A. 2007. Diskursanalyse. Aktueller theoretischer Ort und methodische Möglichkeiten. Warnke, I. (Ed.). 2007. *Diskurslinguistik nach Foucault: Theorie und Gegenstände:* 27-52. Berlin/Boston: De Gruyter.

Fairclough, N. 1992. *Discourse and Social Change.* Cambridge: Polity.

Fankhauser, P. and M. Kupietz. 2019. Analyzing Domain Specific Word Embeddings for a Large Corpus of Contemporary German. Knight, D. (Ed.). 2019. *Proceedings of the 10th International Corpus Linguistics Conference, CL2019, Cardiff, Wales, UK, 22–26 July 2019.*
https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/9117 [30 January 2025]

Harris, Z.S. 1954. Distributional Structure. *Word* 10(2–3): 146-162.

Jäger, S. 2012. *Kritische Diskursanalyse. Eine Einführung*. Sixth edition Münster: Unrast-Verlag.

Kämper, H. 2006. Diskurs und Diskurslexikographie. Zur Konzeption eines Wörterbuchs der Nachkriegsdiskurse. *Deutsche Sprache* 34(4): 334-353.

Klosa, A. and H. Lüngen. 2018. New German Words: Detection and Description. Čibej, J., V. Gorjanc, I. Kosem and S. Krek (Eds.). 2018. *Proceedings of the XVIII EURALEX International Congress, EURALEX 2018, Ljubljana, Slovenia, 17–21 July 2018*: 559-569. Ljubljana: Ljubljana University Press.

Koller, V., S. Kopf and M. Miglbauer (Eds.). 2019. *Discourses of Brexit*. London: Routledge.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2040 (Lexicofocus)

Mattfeldt, A., I. Warnke and L. Herford. 2025. Neodiskurse und ihre Morphologie — Bemerkungen zum Substantivprimat. Michel, S. (Ed.). 2025. *Diskursmorphologie: Ansätze und Fallstudien zur Schnittstelle zwischen Morphologie und Diskurslinguistik:* 121-152. Berlin/Boston: De Gruyter.

Partington, A. 2004. Corpora and Discourse: A Most Congruous Beast. Partington, A., J. Morley and L. Haarman (Eds.). 2004. *Corpora and Discourse*: 11-20. Bern: Peter Lang.

Partington, A., A. Duguid and C. Taylor. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS).* Amsterdam/Philadelphia: John Benjamins.

Rudkowsky, E., M. Haselmayer, M. Wastian, M. Jenny, Š. Emrich, and M. Sedlmair. 2018. More than Bags of Words: Sentiment Analysis with Word Embeddings. *Communication Methods and Measures* 12(2–3): 140-157.

Storjohann, P. 2024. IDS-*Neo*$^{2020+}$: A Novel Resource for New German Words in Use. *International Journal of Lexicography* 37(4): 389-403.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Teubert, W. 2002. Die Bedeutung von Globalisierung. Panagl, O. and H. Stürmer (Eds.). 2002. *Politische Konzepte und verbale Strategien. Brisante Wörter — Begriffsfelder — Sprachbilder*: 149-167. Frankfurt a. M.: Peter Lang.

Teubert, W. 2006. Korpuslinguistik, Hermeneutik und die soziale Konstruktion der Wirklichkeit. *Linguistik Online* 28(3): 41-60.

Teubert, W. 2010. *Meaning, Discourse and Society*. Cambridge: Cambridge University Press.

Van Dijk, T.A. 2008. *Discourse and Context: A Sociocognitive Approach*. Cambridge: Cambridge University Press.

Wodak, R. 2004. Critical Discourse Analysis. Seale, C., G. Gobo, J.F. Gubrium and D. Silverman (Eds.). 2004. *Qualitative Research Practice*: 197-213. London: Sage.

Wodak, R. and M. Meyer. 2015. *Methods of Critical Discourse Analysis.* Third edition. London: SAGE.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

The Role of Combining Forms in Creating New English Compounds: Data-Driven and Construction Approaches*

Jin-hong Huang, *College of Foreign Languages, Fujian Normal University, Fuzhou, China (jhhuang@fjnu.edu.cn) (https://orcid.org/0000-0003-0233-789X)*
and
Yongwei Gao, *College of Foreign Languages and Literature, Fudan University, Shanghai, China (Corresponding Author, ywgao@fudan.edu.cn) (https://orcid.org/0000-0001-7026-9273)*

Abstract: This paper explores the role of combining forms (CFs) in the formation of new compounds in contemporary English and examines their productivity and underlying mechanisms. This study applies the data-driven and construction approaches to investigate 11 neoclassical and native CFs, utilising both dictionaries and corpora when employing the data-driven approach. This paper culls new compounds from the *Oxford English Dictionary* and three neologism dictionaries. The three main findings are as follows: Firstly, the 11 CFs used in forming new compounds can be classified as highly productive, moderately productive, or low in productivity. Secondly, the construction morphology (CxM) can be used to analyse the formational mechanisms of new compounds. According to CxM, the CF compounds are abstracted as hierarchical schemas that are form–meaning pairs. The compounds exhibit syntactic and semantic constraints on their formation. Thirdly, the status of neologisms as compounds, blends or derivatives is scrutinised, along with their treatment in dictionaries. This study provides insights into the ongoing evolution of compounding in present-day English and discusses the role of CFs in lexical innovation.

Keywords: COMBINING FORMS, NEW ENGLISH COMPOUNDS, DATA-DRIVEN APPROACH, CONSTRUCTION MORPHOLOGY, PRODUCTIVITY, FORMATIONAL MECHANISM

Opsomming: Die rol van kombinasievorms in die skep van nuwe Engelse samestellings: Datagedrewe en konstruksiebenaderings. In hierdie artikel word die rol van kombinasievorms (KV's) in die vorming van nuwe samestellings in moderne Engels verken, asook hul produktiwiteit en onderliggende meganismes. Hierdie studie pas die datagedrewe en konstruksiebenadering toe om 11 neoklassieke en inheemse KV's te ondersoek, en sowel woordeboeke

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

as korpora word tydens die toepassing van die datagedrewe benadering gebruik. Nuwe samestellings uit die *Oxford English Dictionary* en drie neologismewoordeboeke word in hierdie artikel geselekteer. Die drie hoofbevindings is soos volg: Eerstens kan die 11 KV's wat gebruik word om nuwe samestellings te vorm, geklassifiseer word as hoogs produktief, matig produktief of laag in produktiwiteit. Tweedens kan die konstruksiemorfologie (KxM) gebruik word om die vormingsmeganismes van nuwe samestellings te ontleed. Volgens die KxM word die KV-samestellings geabstraheer as hiërargiese skemas wat uit vorm–betekenispare bestaan. Die samestellings toon sintaktiese en semantiese beperkings op hul vorming aan. Derdens word die status van neologismes as samestellings, reduksiesamestellings of afleidings ondersoek, sowel as hul hantering in woordeboeke. Hierdie studie verskaf insigte in die deurlopende evolusie van samestelling in hedendaagse Engels en bespreek die rol van KV's in leksikale innovasie.

Sleutelwoorde: KOMBINASIEVORMS, NUWE ENGELSE SAMESTELLINGS, DATAGEDREWE BENADERING, KONSTRUKSIEMORFOLOGIE, PRODUKTIWITEIT, VORMINGSMEGANISME

1. Introduction

Combining forms (CFs), a lexical concept that first appeared in the English language as early as 1884 and was defined by the *Oxford Dictionary of English* (ODE) as "a form of a word normally used in compounds in combination with another element to form a word", have traditionally played an important role in forming new words in the English language. Researchers have drawn a line between CFs and other morphological elements such as affixes, splinters, affixoids, and roots (Fradin 2000). Here it is necessary to distinguish CFs briefly from affixes and splinters. A CF can be used to attach to an independent word (e.g., autopilot), an affix (e.g., *autocyst*), or another CF (e.g., *autography*), while an affix can combine either a free word (e.g. self-made) or a CF (e.g., *contranym*)[1], but not another affix (e.g., **pre-* + *-ism*). In terms of semantics, a CF tends to convey more specific meanings than an affix. The number of CFs has always been on the rise whereas there are seldom new affixes emerging in English. Splinters, the clipped elements of blends, are originally meaningless and non-independent (Adam 1973). Some splinters evolve into CFs through frequent use (e.g. *Franken-* from *Frankenstein*), however only those splinters that show high productivity and widespread usage tend to acquire this status. In this study, a CF is defined as follows: As a bound morpheme with an independent morphological status and an integral form and meaning, a CF can appear in compounds and blends and can be categorised as a 'neoclassical CF' or 'native CF' (cf. Section 2.1).

English compounds are words formed by combining two or more bases to create new lexemes with a distinct meaning. They can be written as a single word (e.g., *microbreak*), with a hyphen (e.g., *micro-expression*), or as separate words (e.g., *micro scale*). In present-day English, compounds are continually evolving as new combinations to reflect technological advancements, social changes, and cultural trends.

Traditionally, the productivity of most CFs pales in comparison with that

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

of affixes. However, a number of CFs are fairly productive in English today. CFs like *immuno-* and *-plasty* have been used to spawn at least one hundred scientific terms respectively. *Cyber-*, a relatively new CF, has been very productive in forming new English compounds whose number may have amounted to dozens or even over a hundred, as are recorded by major English dictionaries such as the ODE and *Collins English Dictionary* (CED). Frequently used *cyber-* compounds include *cyberbullying*, *cybercafé* and *cybercrime*, to mention a few. Other newly emerged CFs like *e-* (e.g., *e-book*) and *-preneur* (e.g., *dadpreneur*) have so far formed scores of new words that have been widely used in English. Thanks to their vitality, the current number of CFs has apparently far surpassed that of affixes, as is shown in the *Oxford English Dictionary* (OED) in which there are over two thousand CFs and only several hundred affixes, including both prefixes and suffixes.

A significant number of new compounds are created with CFs and other word-forming elements. While scholars such as Fischer (1998) and Mattiello (2023) have examined the productivity of CFs, the productivity of CFs in forming new compounds remains unexplored. Although a profusion of studies into compounds has been conducted from different perspectives (e.g., Scalise and Vogel 2010; Bauer 2017), little is known about the formational mechanisms of new English compounds formed with CFs. Given the current situation, this paper aims to examine these aspects. Specifically, this paper deals with 11 CFs (see Table 1 in Section 4.1). The two research questions are as follows: (1) What is the productivity of the 11 CFs in forming new compounds? (2) What is the mechanism underlying the formation of new compounds formed with the CFs?

A data-driven approach in linguistic studies emphasises the analysis of real-world language use through large datasets, such as dictionaries and corpora, which facilitate the identification of empirical evidence for linguistic phenomena. To address these reasearch questions, this paper employs a data-driven approach involving dictionaries and corpora (see Section 4.3) to examine both the productivity of CFs in forming new compounds and the formational mechanisms of such compounds. The construction morphology proposed by Booij (2010) is applied to analyse the formational mechanisms of these CFs.

2. Previous studies on CFs and English compounds

2.1 Research on CFs

CFs are traditionally defined as neoclassical elements of Greek or Latin origin. Scholars have explored whether neoclassical CFs should be classified as instances of affixation or compounding (Bauer et al. 2013: 431-490). Some researchers extend the definition of CFs to include elements originating from native English words (i.e., 'native CFs' in Wiemeyer's (2019) terms), whereas Amiot and Dugas (2020) divide these word-forming elements into classical CFs and modern CFs. CFs also can be grouped by their positions in words, mainly including ini-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

tial combining forms (ICFs) and final combining forms (FCFs) (Plag 2003: 156). Approaches to classifying CFs therefore vary (see Mattiello 2023).

Apart from the classifications of CFs, prior studies focus on their semantics (e.g., Lehrer 1998; Lalić-Krstin et al. 2022) and productivity (e.g., Fischer 1998; Rita-Kasari 2013). While these studies have greatly enhanced the understanding of CF classification, semantics, and productivity, they devoted less attention to how CFs function semantically and productively when forming compounds. This study aims to address this research gap by investigating the formation, semantic properties, and productivity of new CF compounds.

2.2 Research on English compounds

Previous studies into English compounds have concentrated on their classifications, word forms or structures, semantics, word classes, as well as their formational mechanisms. Scholars have been rather inconsistent in their classification of compounds, not only in terms of the classes deemed important, but also in the nomenclature used for the classes (e.g., Fábregas and Scalise 2012). While noun compounds have been extensively studied, systematic empirical research on adjectival and verbal compounds remains scarce (e.g., Bauer et al. 2013). Structural analyses have shed light on the internal organisation of compounds, considering aspects such as headedness, constituency, stress, and orthography (e.g., Anderson 1992). In addition, prior studies have distinguished between idiomatic and compositional compounds and have applied various theoretical frameworks to analyse their meanings (e.g., Ten Hacken 2016). Research on word classes has provided detailed categorisations of compound nouns, compound verbs, compound adjectives, and even less commonly discussed categories such as compound adverbs and compound interjections (e.g., Adams 1973). Scholars have also examined the mechanisms underlying the formation of compounds, employing perspectives from schema theory, blending theory, or paradigmatic analysis (e.g., Fauconnier and Turner 2002; Bagasheva 2020).

Despite the valuable insights mentioned above, studies specifically examining CF compounds across these dimensions are limited (e.g., Díaz Negrillo 2014; Pulcini 2020). In particular, new compounds involving CFs are an underexplored area. Booij (2009: 201-216) applies construction morphology to analyse compounds in different languages, partly involving English compounds. However, only a few studies have adopted this approach to exploring neoclassical compounds (cf. Hayashi 2024). This paper therefore uses the approach of construction morphology to examine the compounds formed with CFs.

3. Theoretical considerations

3.1 Construction Morphology in word formation

Construction Morphology (CxM), a morphological theory developed by the Dutch

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

linguist Geert Booij, integrates insights from Construction Grammar and introduces the concept of 'construction' into morphological analysis. The notion of 'construction' in CxM is derived from Goldberg's (2006: 5) hypothesis, which characterises constructions as linguistic patterns whose form or function is not strictly predictable from their constituent parts. The basic ideas of CxM have been explicated and defended in Booij (2010), and in a number of introductory chapters on CxM (e.g., Booij 2018). In CxM, a word is conceptualised as a pairing of form and meaning, with words in the lexicon understood as systematic form–meaning correspondences (Booij 2010: 5).

3.1.1 Word-based morphology

A distinctive feature of CxM is its word-based approach to morphology, which contrasts with the morpheme-based approach (Kiparsky 1982). While the morpheme-based approach sees complex words as concatenations of morphemes, word-based morphology treats the word itself as the model for new words. In other words, CxM posits that "words are formed from words" (Aronoff 1976: 46). Within the framework of generative word formation, Aronoff (1976) introduces the theory of "word-based morphology", which hypothesises that "all regular word formation processes are word-based. A new word is formed by applying a regular rule to a single already existing word" (ibid: 21). CxM considers each word as a holistic entity and examines its internal structure by comparing it with paradigmatically related words. There is no strict separation of grammar and lexicon (Booij 2018: 3). For instance, instead of viewing *writer* as a simple combination of write and *-er*, CxM sees it as part of a broader family of words (e.g., *singer* and *teacher*) that English speakers recognise and use productively.

3.1.2 Schemas

Unified schemas that allow for the derivation of multiple complex word formation products are a central concept in CxM. The model of CxM uses constructional schemas to account for the systematic form–meaning relations between words. Booij (2015: 450) asserts that a "schema expresses the systematic and abstract form–meaning correspondence found in a set of word pairs". Schemas are used to represent the internal structures of complex words. The relationship between a constructional schema and its instantiations can be illustrated through an inheritance tree, with more abstract schemas occupying the higher nodes, and specific lexical items situated at the lower nodes. The lexical item inherits properties from the higher-level schema. An English suffixal schema (1), as outlined by Booij (2018: 4), can serve as an example to demonstrate this relationship:

(1) $[[x]_{Ni}\ y]_{Aj} \leftrightarrow$ [possessing characteristic properties of $SEM_i]_{SEMj}$

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

The double arrow in the schema (1) symbolises the correspondence between form and meaning. The variable *x* denotes the phonological form of the noun. Co-indexation indicates that the semantic representation (SEM) of the noun is embedded within the meaning of the corresponding adjective. This schema assumes an independently specified semantic representation of the noun, thus highlighting that the schema itself is grounded in paradigmatic relations among lexical items. The variables *i* and *j* stand for the lexical indexes on the phonological and syntactic properties of words.

Hierarchical constructional schemas represent generalisations about existing complex words while simultaneously providing a structured framework for the formation of new words (Booij 2010). The relationship between an abstract constructional schema and its individual instantiations can be represented in a tree structure, where the schema functions as the dominant node, and individual lexical items occupy lower-level nodes that inherit the properties of their superordinate categories. For instance, the derivative *baker* can be represented within such a hierarchical structure (Booij 2005: 124), as shown in schema (2):

(2) $[[x]_X\ y]_Y$
$|$
$[[x]_V\ er]_N$ 'one who V's'
$|$
$[[bak]_V\ er]_N$ 'one who bakes (professionally)'
$|$
$[bake]_V$

Booij (2005: 123) uses schema (3) to represent a compound. For compounds formed with ICFs and FCFs, they can be abstracted as $[[ICF]_X\ [y]_Y]_Y$ and $[[x]_X\ [FCF]_Y]_Y$ respectively, where the form of the variable *x* or *y* is not entirely predictable. Therefore, both $[[ICF]_X[y]_Y]_Y$ and $[[x]_X\ [FCF]_Y]_Y$ can be identified as constructions in the framework of CxM. CxM regards compounds as constructions, where form–meaning correspondences are not strictly predictable. By abstracting CF compounds into schema (3), CxM offers a flexible and systematic way to examine the formation of new compounds. Moreover, it provides a clear method for understanding the formational mechanisms of these compounds, as the schemas allow for the variation of components while preserving the underlying structural and semantic patterns. Through the inheritance of properties across different levels of abstraction, CxM can effectively account for both the consistency and innovation observed in the formation of new compounds.

(3) $[[x]_X\ [y]_Y]_Y$ 'Y with some relation to X'

In comparison with the schema theory (Bybee 1985) or the paradigmatic approach to compounds (Bauer 2017), words in CxM are treated as form–meaning pairings, and CxM deals with the semantic relationships between elements. The schema

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

theory or the paradigmatic approach, on the other hand, focuses on the abstract structures and patterns that produce new words within established paradigms, with less emphasis placed on meaning as a fundamental factor in word formation.

3.1.3 Semantic and syntactic constraints on compounds

CxM asserts that constructional schemas impose semantic constraints on the formation of compounds, so that the meaning of the resulting compound remains interpretable within systematic lexical patterns. For example, the CF *photo-* in *phototherapy* is subject to the constraint that the second element must be semantically associated with treatment or the medical field. The application of semantic constraints in CxM follows prototypical patterns, which account for the varying productivity of CFs in forming new compounds. A case in point is *tele-*, which exhibits high compatibility with communication-related nouns (e.g., *telephone*) but does not typically co-occur with physical object nouns like *telechair*. On the other hand, CxM maintains that syntactic constraints imposed by constructional schemas ensure the syntactic predictability of compounds. In this regard, the CF *geo-* in *geography* is subject to the constraint that the second element can be a noun or a CF, thereby producing a noun compound. The systematic nature of syntactic constraints in CxM shows that CFs combine only bases that conform to specific word class patterns.

3.2 Morphological productivity

Morphological productivity, a key concept in the word-formation research and morphological theories, is the likelihood that a construction will apply to a new item (Bybee 2010: 94). Each lexical slot in a construction has its own degree of productivity. If a process in morphology is fully regular and actively used in the creation of new words, this process is considered productive (Richards and Schmidt 2010: 463). Fischer (1998: 63) argues that the current productivity of CFs can be measured by the number of new words included in dictionaries. Compilers of dictionaries largely consider the established words for inclusion. Bauer (2001: 159) takes the view that using dictionaries for the measurement of productivity is "likely to prevent relatively rare established words being taken as neologisms, and it allows the genuinely productive use of morphological processes to be seen".

Both qualitative and quantitative methods can be used to assess productivity. This paper adopts the quantitative measure and counts the number of new compounds included in dictionaries over specific periods to analyse the productivity of CFs in forming new compounds. Counting the number of new words in a given phase is a measure of productivity (Plag 2021: 489). However, the use of dictionaries is not immune from problems. As Bauer (2001: 158-159) notes, con-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

sulting a dictionary provides a conservative idea about what is 'part of the language' by often excluding neologisms, rare words, and established words deemed invaluable to the target audience.

According to Booij (2010: 50), productivity is a type of holistic property of morphological constructions. Schemas in the context of morphological productivity apply to both productive and non-productive form–meaning correspondences and show that regularity can exist without productivity. However, a unified schema enhances the productivity of word-formation patterns by reinforcing morphological generalisations. Unified schemas provide a framework for understanding key differences in the evolution of individual subpatterns, particularly with regard to morphological productivity and the semantic dimensions of word-formation constructions (Kempf and Hartmann 2018: 442). In turn, productivity is linked to type frequency, which influences the degree of a schema's entrenchment (Barðdal 2008).

4. Methodology and procedures

4.1 Research objectives

The 11 CFs examined in this study and their origins are listed in Table 1. By referring to the OED, the dates of first occurrence of the said CFs in English formations were obtained. One exception of a CF not included in the OED but included in the list is *-ware*, derived from 'software'. Based on their etymological origins, the 11 CFs are classified into neoclassical CFs and native CFs, while their roles as ICFs or FCFs are also identified.

Table 1: CFs examined in this study

CF	Origin	Date of first occurrence in English formations	Type	
bio-	Latin	the early 19th century	ICF	neoclassical CF
macro-	Greek	the 1820s		
micro-	Greek	1603		
techno-	Greek	the 19th century		
tele-	Greek	the first half of the 19th century		
-phobia	Latin	the 18th century	FCF	
-plasty	Greek, combined with the suffix *-y*	the first half of the 19th century		
cyber-	cybernetics	the 1960s	ICF	native CF
e-	electronic	the late 1980s		
immuno-	immune + *-o-*	the beginning of the 20th century		
-ware	software	the 1960s	FCF	

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

As allomorphic variants of classical Latin or Greek words, the five ICFs *bio-*, *macro-*, *micro-*, *techno-*, and *tele-* that end with a vowel /o/ or /i/ are neoclassical CFs. Two other neoclassical CFs *-phobia* and *-plasty* are FCFs. This paper also examines four native ICFs, including *cyber-*, *e-*, *immuno-*, and *-ware*, of which *immuno-* derives from 'immune' and *-o-* while the other four are truncated from existing English words. *Cyber-*, *e-*, and *-ware* are new CFs whereas the other eight CFs have been in use longer. The 11 CFs were selected because they provide a representative sample including both traditional and emerging CFs, diverse origins, structural roles (ICFs and FCFs), and distinct historical stages of development. All of the said CFs have been productive in making new words to varying degrees in the past decades. These CFs are well-suited for analysis, as they have formed a sufficient number of new compounds for investigation. The time span from the 1960s to the 2020s was chosen because sufficient time has passed to enable a measurement of productivity of the 11 CFs in forming new compounds.

4.2 A data-driven approach

This paper employs a data-driven approach, which means "methods based on statistical models learnt from data" (Mitkov 2022: 1252). Mittelberg et al. (2007: 30) emphasise that 'data-driven' does not equal 'data-based', since in data-driven work the data are not just the material basis for the analysis, or just used to test or exemplify theories formulated prior to the advent of large corpora. Rather, the data actually guide the course of analytical processes and theory building. Corpora have been particularly useful in enabling the creation of data-driven models (Baker et al. 2006: 41). Dictionaries compiled using large corpora of real-world texts exemplify a data-driven approach because they derive word meanings and usage patterns from empirical evidence, rather than theoretical or prescriptive rules. Both dictionaries and corpora are utilised when employing the data-driven approach in the present study.

4.3 Sources of data

The data sources of this paper are dictionaries and corpora. Firstly, the new compounds formed with the 11 CFs were selected from major English dictionaries, such as the OED. The OED is used because it is a comprehensive and authoritative resource on the English language, tracing the meanings, history, and evolution of words. Other online dictionaries do not systematically provide the dates of emergence of entries, which is essential for measuring the productivity of CFs in forming new compounds during certain periods. Therefore, only the OED was selected among large-size online dictionaries.

Secondly, this study makes good use of three English–Chinese dictionaries of neologisms to collect new compounds. *A Dictionary of Contemporary English Neologisms and Trendy Words* (DCEN, Gao 2024) records nearly 17 000 entries of

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

present-day English neologisms and trendy terms, most of which have emerged or gained widespread usage since the 1970s (1970s–2020s). This dictionary particularly focuses on recording new words that have been frequently used in the media over the past decade. *An English–Chinese Dictionary of Neologisms in the New Era* (ECDN, Gao 2023) includes over 4 000 new English words and phrases that have appeared or been frequently used since the beginning of the 21st century (2000–2015). *A 21st Century English–Chinese Dictionary of Neologisms* (CECDN, Gao 2021) documents more than 2 500 new words and meanings that have been active in use since the early 2000s (2000s–2020s). For this paper, the three dictionaries were chosen because of their thorough and timely recording of English neologisms across several decades, which allows for a detailed study of new compounds formed with CFs. The use of these four dictionaries to collect new compounds also has limitations. The OED's recording of new words may not be timely, and some emerging compounds can be missing. Similarly, the three English–Chinese neologism dictionaries may not cover all new compounds, despite their provision of comprehensive and representative samples.

Thirdly, this study utilises three corpora: the Corpus of Historical American English (COHA; Davies 2010), the Corpus of Contemporary American English (COCA; Davies 2008), and the News on the Web (NOW; Davies 2016). These corpora were used to determine the first recorded dates of new compounds. COHA, comprising over 475 million words, provides historical American English data from 1820 to 2019 that includes texts from novels, magazines, newspapers, and nonfiction. However, it lacks spoken data where the new compounds in question may appear. To address this, COCA was included, as it contains over one billion words from various genres, including spoken texts. NOW, a monitor corpus with 16,4 billion words from online newspapers and magazines (2010–present), updates monthly with 200–220 million words. As new compounds frequently emerge in news and media, NOW is particularly useful for tracking their development.

The OED is particularly valued for its ability to pinpoint the first recorded use of words and thus the dates of first appearance of compounds can be attested to make sure whether they are neologisms. The appearance date of new compounds enables us to explore the productivity of CFs during a certain span of time. If a compound is not included in the OED, the authors refer to other online dictionaries that provide the etymological information where may contain the dates of first use of headwords. The online dictionaries in question involve ODE, CED, *Merriam-Webster's Collegiate Dictionary* (MWCD), and *Wiktionary* (WKD). If the dates of first appearance are all absent in the five dictionaries (i.e. OED, ODE, CED, MWCD, and WKD), reference is also made to the three corpora (COHA, COCA, and NOW) to attest their dates of first use in the dataset.

4.4 Data compilation

Each of the CFs was searched in the OED. The OED records all the said word-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

forming elements as CFs, except *-ware*. The OED provides the section of "compounds and derived words" for the CFs, and the types of formation are also indicated. In the case of *bio-*, its formations included in the OED are 357 compounds, 6 derivatives, and 4 borrowings. The numbers of compounds formed with the CFs covered in the OED are listed in Table 2. The numbers of compounds in the three neologism dictionaries are also investigated, as is shown below.

Table 2: Number of compounds in four dictionaries

	OED	**DCEN**	**ECDN**	**CECDN**
Time span of inclusion	1960–2024	1970–2024	2000–2015	2000–2021
CF	**Compound number in each dictionary**			
bio-	357	125	29	9
macro-	143	12	0	4
micro-	635	110	30	24
techno-	57	18	2	2
tele-	201	21	3	2
-phobia	93	30	11	1
-plasty	42	15	2	6
cyber-	68	104	23	6
e-	26	64	28	4
immuno-	83	16	0	0
-ware	13	34	17	1

5. Results and analysis

5.1 Dates of first use for new compounds

The new compounds formed with each CF and recorded in the OED, DCEN, ECDN, and CECDN were compiled together, with duplicates excluded. The dates of first use for these compounds are sourced from dictionaries such as the OED, ODE, CED, MWCD, and corpora like COHA, COCA, and NOW. Most of the CFs in question were used to form compounds before 1960 or have formed such words since the 1960s (except *e-*). This paper uses 1960 as the starting year to measure the productivity of the CFs in producing new compounds while the ending year is 2024. Compounds formed between 1960 and 2024 and recorded in the aforementioned four dictionaries are enumerated in Table 3. The '2020s' are used to refer to the period from 2000 to 2024, for consistency with others. With the data-driven periodisation in place, the analysis of these compounds are divided into ten-year intervals. Note that the absence of new CF compounds in certain periods in the four dictionaries does not imply that these CFs did not form any compounds during these periods. Instead, it indicates that the com-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

pounds they formed did not become established to warrant lexicographic inclusion.

The *bio-* formations within English were found from the early 19th century. The OED, DCEN, ECDN, and CECDN include 214 *bio-* compounds that appeared before the 1960s. A surge occurred in the 1960s and 1970s, with 44 and 48 new compounds, respectively. Despite fluctuations, 168 new *bio-* compounds that have emerged since the 1960s are included in the aforementioned dictionaries. *Macro-* first appeared in English formations in the 1820s, and a total of 108 *macro-* compounds that emerged before the 1960s are listed in the four dictionaries. The 1960s saw 22 new *macro-* compounds included, followed by fewer than five being added per decade from the 1980s onward, totaling 42 since the 1960s. *Micro-* was first seen in use in English words as early as 1603. The four dictionaries have included as many as 508 *micro-* compounds that appeared before the 1960s. The 1960s saw the peak of new *micro-* compounds with 57, followed by a gradual decline. 176 new *micro-* compounds that emerged from the 1960s to the 2020s are recorded in the four dictionaries. The other eight CFs showed varying degrees of productivity, as indicated in Table 3.

Table 3: Number of new compounds in four dictionaries by dates of first appearance

CF	Undetermined	Before 1960s	1960s	1970s	1980s	1990s	2000s	2010s	2020s	Sum of 1960–2024
bio-	1	214	44	48	29	18	8	20	1	168
macro-	0	108	22	7	4	2	2	4	1	42
micro-	2	508	57	33	22	16	12	35	1	176
techno-	0	22	9	10	13	4	1	5	0	42
tele-	0	163	10	14	9	2	0	3	0	38
-phobia	0	78	4	1	3	5	2	9	1	25
-plasty	0	37	3	1	0	2	1	6	0	13
cyber-	3	0	4	4	7	51	9	27	2	104
e-	0	0	0	0	5	24	18	21	0	68
immuno-	0	37	29	14	3	0	0	0	0	46
-ware	1	0	1	2	9	5	4	7	3	31

5.2 The formal properties of new compounds

This section examines the formal properties of new compounds formed with 11 CFs from the 1960s to the 2020s and explores the distribution of solid, hyphenated, and open compounds among these compounds (see Table 4). With the exception of *techno-* and *e-*, the other nine CFs show a strong tendency towards forming solid compounds. The number of solid compounds formed with *bio-*, *micro-*, as well as *cyber-* surpasses 100, demonstrating the established roles of these CFs in such compounds, as in *bioabsorption*, *microbeer*, and *cyberaggression*. CFs such as *tele-*, *-phobia*, *-plasty*, *immuno-*, and *-ware* exclusively form solid compounds. The

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

large number of solid compounds reflects their dominance in fields such as biology, medicine, science, and technology.

Table 4: Forms of new compounds recorded in the four dictionaries

CF	Solid compound	Hyphenated compound	Open compound
bio-	164	4	0
macro-	35	7	0
micro-	161	15	0
techno-	19	22	0
tele-	38	0	0
-phobia	25	0	0
-plasty	13	0	0
cyber-	101	3	0
e-	1	67	0
immuno-	46	0	0
-ware	31	0	0

Hyphenated compounds are less common but remain significant for certain CFs. For example, *techno-* exhibits a higher proportion of hyphenated compounds than solid ones. This suggests that *techno-* compounds often represent newly emerging or less standardised terms, such as *techno-junkie* and *techno-thriller*. In a similar vein, *e-* displays a striking preference for hyphenation, with 67 hyphenated compounds like *e-economy* and *e-payment*. None of the CFs in the dataset form open compounds. Although compounds such as *cyber cheating* and *cyber school* are included in the dictionaries, 'cyber' here functions as an independent word instead of a CF. They were not taken into account in this discussion.

5.3 The word classes of new compounds

The data in Table 5 demonstrate that the majority of new CF compounds are nouns. CFs such as *bio-*, *micro-*, and *cyber-* are particularly productive in forming noun compounds. Examples include *biocapacity*, *microangiography*, and *cybersecurity*. Adjective compounds (e.g., *immunoregulatory* and *microaerobic*), while less common, are still notable, with CFs like *bio-*, *micro-*, and *immuno-* forming over 15 adjectives. In contrast, verb compounds (e.g., *cyberstalk* and *micromanage*) are rare among CFs, with *bio-*, *micro-*, *tele-*, and *cyber-* contributing only a handful of examples. Only *immunoelectrophoretically*, an adverb compound of *immuno-*, is recorded. The new compounds of *macro-*, *techno-*, and *-plasty* recorded are nouns or adjectives while those of *-phobia*, *e-*, and *-ware* are nouns.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

Table 5: Word-classes of new compounds formed with CFs

CF	noun	adjective	verb	adverb
bio-	134	27	7	0
macro-	32	10	0	0
micro-	154	18	4	0
techno-	30	11	0	0
tele-	34	1	3	0
-phobia	25	0	0	0
-plasty	12	1	0	0
cyber-	95	4	5	0
e-	68	0	0	0
immuno-	26	18	1	1
-ware	31	0	0	0

Table 6 illustrates the compounding of CFs with other constituents, with the "CF + noun" construction dominating across all ICFs. *Bio-* forms 134 noun compounds, while *micro-* and *cyber-* contribute 154 and 91, respectively. Though "CF + adjective" and "CF + verb" constructions appear, they are less frequent, and the "CF + adverb" construction occurs only once in *immunoelectrophoretically*. Some ICFs, such as *techno-*, *cyber-*, and *immuno-*, also form "CF + CF" constructions. As for the FCFs, *-phobia* mainly forms "noun + CF" constructions and the only *-phobia* compound in the form of "CF + CF" is *coronophobia*. The dominant construction of *-plasty* is "CF + CF". Compounds formed with *-ware* exclusively use the "noun + CF" construction.

Table 6: Constituents of new compounds formed with CFs

ICF	CF + noun	CF + adjective	CF + verb	CF + adverb	CF + CF
bio-	134	27	7	0	0
macro-	32	10	0	0	0
micro-	154	18	4	0	0
techno-	30	9	0	0	2
tele-	34	1	3	0	0
cyber-	91	4	5	0	4
e-	68	0	0	0	0
immuno-	25	18	1	1	1
FCF	**CF + CF**	**adjective + CF**	**noun + CF**	/	/
-phobia	1	0	24	/	/
-plasty	9	1	3	/	/
-ware	0	0	31	/	/

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

5.4 The semantic properties of new compounds

5.4.1 The semantic preference of CF compounds

ICFs and FCFs combine other elements to form lexemes. Such elements are termed the 'first element' (e.g., *intersex* in *intersexphobia*) and the 'second element' (*aerosol* in *bioaerosol*) respectively. Exploring the semantic categories of the first or second elements of the CF compounds can reflect their semantic preference. For example, Fischer (1998: 146) presents nine categories of the second elements of *cyber-*, including persons, living beings, objects, body parts or sense organs, places or buildings, and so on. The new compounds formed with the CFs from the 1960s to the 2020s are analysed below. This study classifies the first or second element into five semantic categories: objects, abstractions, behaviours and activities, people, and animals. The "animal" here denotes any living creature other than a human being. These five categories can cover all the first or second elements of the 11 CFs.

Table 7: Semantic types of first or second elements of CF compounds

CF	Objects		Abstractions		Behaviors & Activity		People		Animals	
	number	proportion	number	proportion	number	proportion	number	proportion	number	proportion
bio-	46	27.3%	72	42.8%	41	24.4%	9	5.3%	0	0
macro-	14	33.3%	24	57.1%	2	4.7%	1	2.3%	1	2.3%
micro-	66	37.5%	68	38.6%	32	18.1%	6	3.4%	4	2.2%
techno-	3	7.1%	27	64.2%	3	7.1%	9	21.4%	0	0
tele-	5	13.2%	13	34.2%	16	42.1%	4	10.5%	0	0
-phobia	5	20.0%	7	28.0%	4	16.0%	8	32.0%	1	4.0%
-plasty	10	76.9%	2	15.3%	1	7.6%	0	0	0	0
cyber-	5	4.8%	34	32.7%	37	35.5%	27	25.9%	1	0.9%
e-	29	42.6%	14	20.5%	19	27.9%	6	8.8%	0	0
immuno-	5	10.8%	24	52.1%	17	36.9%	0	0	0	0
-ware	1	3.2%	15	48.3%	15	48.3%	0	0	0	0

Bio- primarily combines words expressing abstractions (*bioavailability*), or those denoting objects (*biochip*), and behaviours and activities (*biohacking*). Only a few *bio-* compounds involve people (*biohacker*), and none refers to animals. *Macro-* and *micro-* display clear semantic preferences for abstractions (*macromodelling*) and objects (*macrolens*). *Micro-* also combines the second elements referring to behaviours and activities (*microblogging*). *Techno-* strongly favors the combination with the second elements representing abstractions (*technocentrism*). *Tele-* mainly combines with the second elements representing behaviours and activities (*telebroking*). *Cyber-* primarily combines the second elements that belong to behaviours and activities (*cyberauction*), and abstractions (*cybercommunity*). Further data on other CFs can be found in Table 7.

5.4.2 The semantic change of the CFs in compounds

The meanings of 11 CFs in Table 8 illustrate their dynamic evolution, as these CFs have adapted to new contexts and domains when forming new compounds. By examining the origins of CFs and their current usage, insight is gained into

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

how these morphemes reflect cultural, scientific, and technological developments. If *cyber-* is considered as an example, it is evident that *cyber-* means "computers" in early compounds found in the OED, like *cyberculture*. It appears in nine new compounds with this meaning, as in *cyberathlete*, *cybergeek*, and *cyberphobe*. Terms like *cyberart*, *cybersphere*, and *cyberbabe* reflect its use to describe information technology, focusing on digital systems or products. In 63 new compounds, *cyber-* conveys this meaning. Meanwhile, compounds such as *cyberspace* and *cyberlife* extend the meaning of *cyber-* to virtual reality and online existence. *Cyber-* in compounds such as *cybercrime* and *cyberwarfare* signifies the application of digital technologies to illegal or hostile activities. Furthermore, *cyberaggression*, *cyberstalking*, and *cyberscam* reflect the influence of digital and virtual environments on social activities.

Table 8: Meanings of CFs and the number of new compounds with this meaning

CF	Meaning	No.	Meaning	No.	Meaning	No.	Meaning	No.	Meaning	No.
bio-	living beings	146	biology	18	life	2	biochemical substances as weapons	2		
macro-	large-scale	28	large	12	large compared with the norm	1	visible	1		
micro-	small	157	microscopy	10	abnormally small	5	microphotography or microprinting	2	one millionth	2
techno-	of technology	28	craft, art	7	technological	7				
tele-	to or at a distance	30	done by means of the telephone	6	relating to television	2				
-phobia	fear	14	hostility, prejudice	7	aversion	4				
-plasty	formation	10	plastic surgery	3						
cyber-	information technology	63	virtual reality	32	computers	9				
e-	sth. in an electronic format	30	computer-mediated or internet-based	27	electronic financial transactions	11				
immuno-	immune	22	immunity	20	immunology	4				
-ware	software	28	software (in a metaphorical sense)	3						

6. Discussion

6.1 The productivity of CFs in creating new compounds

As demonstrated in Table 3 and Section 5.1, the productivity of the 11 CFs in forming new compounds varies significantly between neoclassical and native CFs from the 1960s to the 2020s. *Bio-*, *micro-*, and *cyber-*, each with over 100 new compounds, are highly productive. *Macro-*, *techno-*, *tele-*, *e-*, *immuno-*, and *-ware*, with more than 30 new formations, show moderate productivity, while *-phobia* and *-plasty*, with fewer than 30 new formations, exhibit low productivity. The thresholds of 30 and 100 new compounds here provide practical reference points rather than absolute indicators.

The rate at which the 11 CFs form new compounds also notably differs across decades. CFs like *bio-*, *micro-*, and *cyber-* exhibit a rapid expansion, especially

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

during periods of significant scientific and technological development. For instance, new *bio-* compounds experienced a sharp increase in the 1960s and 1970s, a period of intense biological research and medical advancements. These CFs quickly adapted and expanded as new concepts emerged, which indicates that the pace of forming compounds is closely tied to societal changes and technological revolutions.

6.2 The formational mechanism of CF compounds

In CxM, word formation is analysed through hierarchical constructional schemas that reflect systematic form–meaning correspondences. Using *cyber-* as an example, the unified schema $[[\text{cyber}]_X[y]_Y]_Y$ governs its overall combinatory potential, where *cyber-* combines another element *y* to form a lexeme that belongs to the word class Y. In accordance with the analysis in Section 5.4.2, *cyber-* has acquired three meanings so far. Therefore, the schema of *cyber-* can express the meaning of "something related to information technology, virtual reality, and computers". At a more specific level, this general schema branches into $[\text{cyber-}[y]_{Ni}]_{Nj}$ (94 instances, e.g., *cyberart*), $[\text{cyber-}[y]_{Ai}]_{Aj}$ (4 instances, e.g., *cybersecure*), $[\text{cyber-}[y]_{Vi}]_{Vj}$ (5 instances, e.g., *cyberloaf*), and $[\text{cyber-}[y]_{Ci}]_{Nj}$ (4 instances, e.g., *cybernaut*). Each of these schemas represents a distinct form–meaning pairing, with the lower-level schemas inheriting properties from the more abstract schema. This hierarchical structure demonstrates how CxM accounts for both consistency and variation in the formation of new CF compounds.

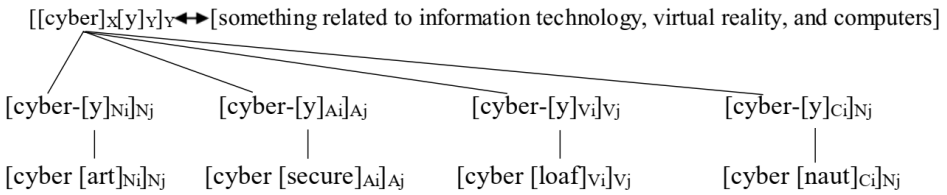


Figure 1: The constructional schema of *cyber-* compounds

Table 9 presents the constructional schemas of the other 10 CFs, and their schemas have varying degrees of productivity in the formation of new compounds. The CF compounds show syntactic constraints, since the first or the second elements which the CFs can combine are of different word classes. It can be noticed that *bio-*, *micro-*, and *tele-* all have three schemas. They can combine a noun, an adjective, or a verb to form new compounds. CFs such as *techno-* and *-plasty* also exhibit three schemas. Apart from compounding with a noun or an adjective, they can form compounds with CFs. *Macro-* can be abstracted into two schemas and this CF can form compounds with a noun or an adjective. These constraints reflect the interplay between form and meaning in CxM, where

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

some CFs serve highly specialised functions while others exhibit broader productivity across multiple schemas.

Table 9: The schematicity of CFs and the number of their new compounds

CF	Schema	No.	Schema	No.	Schema	No.	Schema	No.
cyber-	$[\text{cyber-[y]}_{Ni}]_{Nj}$	94	$[\text{cyber-[y]}_{Ai}]_{Aj}$	4	$[\text{cyber-[y]}_{Vi}]_{Vj}$	5	$[\text{cyber-[y]}_{Ci}]_{Nj}$	4
bio-	$[\text{bio-[y]}_{Ni}]_{Nj}$	149	$[\text{bio-[y]}_{Ai}]_{Aj}$	31	$[\text{bio-[y]}_{Vi}]_{Vj}$	7	/	/
immuno-	$[\text{immuno-[y]}_{Ni}]_{Nj}$	26	$[\text{immuno-[y]}_{Ai}]_{Aj}$	18	$[\text{immuno-[y]}_{Vi}]_{Vj}$	1	$[\text{immuno-[y]}_{ADi}]_{ADj}$	1
	$[\text{immuno-[y]}_{Ci}]_{Nj}$	1	/	/	/	/	/	/
macro-	$[\text{macro-[y]}_{Ni}]_{Nj}$	33	$[\text{macro-[y]}_{Ni}]_{Nj}$	10	/	/	/	/
micro-	$[\text{micro-[y]}_{Ni}]_{Nj}$	167	$[\text{micro-[y]}_{Ai}]_{Aj}$	20	$[\text{micro-[y]}_{Vi}]_{Vj}$	4	/	/
techno-	$[\text{techno-[y]}_{Ni}]_{Nj}$	31	$[\text{techno-[y]}_{Ai}]_{Aj}$	10	$[\text{techno-[y]}_{Ci}]_{Nj}$	2	/	/
tele-	$[\text{tele-[y]}_{Ni}]_{Nj}$	37	$[\text{tele-[y]}_{Ai}]_{Aj}$	1	$[\text{tele-[y]}_{Vi}]_{Vj}$	3	/	/
e-	$[\text{e-[y]}_{Ni}]_{Nj}$	68	/	/	/	/	/	/
-phobia	$[[\text{x}]_{Ci}\text{-phobia}]_{Nj}$	35	/	/	/	/	/	/
-plasty	$[[\text{x}]_{Ni}\text{-plasty}]_{Nj}$	3	$[[\text{x}]_{Ai}\text{-plasty}]_{Aj}$	1	$[[\text{x}]_{Ci}\text{-plasty}]_{Cj}$	14	/	/
-ware	$[[\text{x}]_{Ni}\text{-ware}]_{Nj}$	33	/	/	/	/	/	/

As analysed in Section 5.4.1, CF compounds exhibit semantic constraints based on the semantic types of elements that they combine. *Macro-*, *micro-*, *cyber-*, and *-phobia* can combine the elements representing objects, abstractions, behaviours and activities, people, and animals. In contrast, the second elements of *bio-*, *techno-*, *tele-*, and *e-* exclude animals, whereas those of *immuno-* exclude both people and animals. Similarly, the first elements of *-plasty* and *-ware* do not fall within the categories of people or animals.

Highly productive schemas such as $[\text{bio-[y]}_{Ni}]_{Nj}$ and $[\text{micro-[y]}_{Ni}]_{Nj}$ illustrate how frequently used constructions become entrenched in the lexicon, while less productive schemas like $[\text{immuno-[y]}_{Vi}]_{Vj}$, $[\text{tele-[y]}_{Ai}]_{Aj}$, and $[[\text{x}]_{Ai}\text{-plasty}]_{Aj}$ represent emerging or marginal patterns. The hierarchical organisation of schemas ensures that word-formation generalisations remain systematic, even as new compounds are created. The constraints observed in CF compounds highlight the importance of semantic and syntactic factors in determining a CF's combinatory potential.

6.3 Demarcation between compounds and others and their lexicographical treatment

In accordance with the data in the OED, DCEN, ECDN, as well as CECDN, CF formations involve compounds and blends. The following section examines the etymological treatment of CF formations in the OED, CED, ODE, MWCD, and WKD, as these five dictionaries provide such information for some entries.

The existing deficiencies are as follows: Firstly, the etymological information, especially the ways of formation provided for the same CF formation, may vary in the said dictionaries. Secondly, the etymological information offered for some CF formations is not sufficiently clear and accurate. For instance, *bioethicist* is a derivative. However, its way of formation is treated differently in the five abovementioned dictionaries. *Bioethicist* is a compound according to the OED, whereas it is indicated as being derived from *bioethics* in ODE, MWCD, and WKD. As is

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

attested in the OED, *bioethics* first appeared in 1970, while the date of first appearance of *bioethicist* is 1973. It can be reasonably asserted that *bioethicist* is a derivative rather than a compound. Table 10 displays the lexicographical treatment of *cyberphobic* (compound), *biochar* (blend) and other words as well.

This study on CF compounds shows that CxM transcends rigid distinctions among compounds, derivatives, and blends by representing them through unified schemas, which reduces ambiguity in morphological classification. However, as is displayed in Table 10, dictionaries adopt traditional classifications and present word-formation information based on structural analysis. Given that CF formations can arise from different morphological processes, it is essential for dictionary compilers to propose a standardised lexicographical annotation for CF compounds, distinguishing them clearly from derivatives and blends to ensure the authority and accuracy of dictionaries. For CF compounds like *cyberphobic*, which is formed by two CFs, the etymology section should explicitly label them as CF compounds, along with the analysis of word-forming elements (see Table 10). Entries such as *bioethicist*, which involve the addition of a conventional affix to a pre-existing compound or root, should be treated under derivation. Dictionary compilers should specify the base (*bioethics*) and the attached affix (*-ist*), clarifying that the process is affixal rather than compositional. Regarding blends such as *biochar*, they should be labeled as blends, with the source words clearly identified. Since word-formation mechanisms can be complex, it is desirable for dictionaries to avoid overgeneralisation and acknowledge uncertainties where necessary, while incorporating traditional classification alongside insights from CxM.

Table 10: The etymological treatment of CF formations in dictionaries

CF formation	Suggested treatment	OED	CED	ODE	MWCD	WKD
cyberphobic	CF compound, from *cyber-* (comb. form) + *-phobic* (comb. form)	by derivation <*cyberphobia* n.	derivative of *cyberphobia*	/	/	from *cyber-* + *-phobic*
bioethicist	derivative, from *bioethics* + *-ist*	compounding, *bio-* comb. form + *ethicist* n., after *bioethics*	/	derivative of *bioethics*	derivative of *bioethics*	from *bioethics* + *-ist*, analogous to *ethicist*
microlender	derivative, from *microlending* + *-er*	compounding, *micro-* comb. form + *lender* n. compare *microlending* n.	/	derivative of *microlending*	derivative of *microlending*	from *micro-*+ *lender*
biochar	blend, from *bio*mass and *char*coal	compounding, *bio-* (in *biomass* n.) + *char-* (in *charcoal* n.)	/	blend of *biomass* and *charcoal*	/	from *bio-* + *char*
cyberchondriac	blend, from *cyber-* (comb. form) + hypo*chondriac*	compounding, *cyber-* comb. form + *-chondriac* (in *hypochondriac* adj.)	derivative of *cyberchondria*	blend of *cyber-* and *hypochondriac*	/	blend of *cyber-* and *hypochondriac*
e-tailer	blend, from *e-* (comb. form) + re*tailer*	compounding, *e-* comb. form + *-tailer* (in *retailer* n.). compare *e-tailing* n.	/	blend of *e-* and *retailer*	derivative of *e-tail*	blend of *e-* + *retailer*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

7. Conclusion

Based on data from four dictionaries — the OED, DCEN, ECDN, and CECDN — the findings reveal the productivity of the 11 CFs in forming new compounds from the 1960s to the 2020s. The rate at which the CFs create new compounds notably differs across decades: *bio-*, *micro-*, and *cyber-* are highly productive; *macro-*, *techno-*, *tele-*, *e-*, *immuno-*, and *-ware* are moderately productive; and *-phobia* and *-plasty* are low in productivity.

It is found that the CF compounds, governed by both syntactic and semantic constraints, can be abstracted as hierarchical schemas to represent form–meaning correspondences. In terms of syntactic constraints, the combinatory patterns of CFs vary, with some able to attach to multiple word classes and others restricted to specific word classes. The semantic constraints of CFs differ, with some combining broadly across categories like objects, abstractions, behaviours and activities, people, and animals, while others show selective restrictions, excluding certain categories, such as people or animals. This study also finds that the 11 CFs have undergone semantic changes and have acquired established meanings.

This study contributes to the understanding of compounds formed with CFs. Unlike previous studies focusing on general compounds, newly coined CF compounds are examined, shedding light on ongoing lexical innovation. By integrating the data-driven and construction approaches, this research provides insights into the realised productivity of the CFs in forming new compounds and their underlying mechanisms.

Acknowledgements

We would like to express our sincere gratitude to the two anonymous reviewers and the journal editors for their valuable suggestions and insightful comments.

Endnote

1. *Contra-* is a prefix, and *-onym* is a CF.

References

Dictionaries and tools

***Collins English Dictionary Online* (CED).**
https://www.collinsdictionary.com/ [14 March 2025]

Davies, M. 2008. The Corpus of Contemporary American English **(COCA)**.
https://www.english-corpora.org/coca/ [14 March 2025]

Davies, M. 2010. The Corpus of Historical American English **(COHA)**.
https://www.english-corpora.org/coha/ [14 March 2025]

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

Davies, M. 2016. The News on the Web **(NOW)**.
https://www.english-corpora.org/now/ [14 March 2025]
Gao, Y.W. 2021. *A 21st Century English–Chinese Dictionary of Neologisms* **(CECDN)**. Shanghai: Fudan University Press.
Gao, Y.W. 2023. *An English–Chinese Dictionary of Neologisms in the New Era* **(ECDN)**. Beijing: Commercial Press.
Gao, Y.W. 2024. *A Dictionary of Contemporary English Neologisms and Trendy Words* **(DCEN)**. Beijing: Foreign Language Teaching and Research Press.
***Merriam-Webster's Collegiate Dictionary Online* (MWCD).**
https://www.merriam-webster.com/ [14 March 2025]
***Oxford Dictionary of English Online* (ODE).**
https://premium.oxforddictionaries.com/english/ [14 March 2025]
***Oxford English Dictionary Online* (OED).**
https://www.oed.com/ [14 March 2025]
***Wiktionary* (WKD).**
https://en.wiktionary.org/ [14 March 2025]

Other literature

Adams, V. 1973. *An Introduction to Modern English Word-formation*. London: Longman.
Amiot, D. and E. Dugas. 2020. Combining Forms and Affixoids in Morphology. Lieber, R. (Ed.). 2020. *The Oxford Encyclopedia of Morphology*: 855-873. Oxford: Oxford University Press.
Anderson, S.R. 1992. *A-morphous Morphology*. Cambridge: Cambridge University Press.
Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
Bagasheva, A. 2020. Paradigmaticity in Compounding. Fernández-Domínguez, J., A. Bagasheva and C. Lara-Clares (Eds.). 2020. *Paradigmatic Relations in Word Formation*: 21-48. Leiden/Boston: Brill.
Baker, P., A. Hardie and T. McEnery. 2006. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
Barðdal, J. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic.* Amsterdam: John Benjamins.
Bauer, L. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
Bauer, L. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.
Bauer, L., R. Lieber and I. Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
Booij, G. 2005. Compounding and Derivation: Evidence for Construction Morphology. Dressler, W.U., F. Rainer, D. Kastovsky, and O. Pfeiffer (Eds.). 2005. *Morphology and Its Demarcations*: 109-132. Amsterdam/Philadelphia: John Benjamins.
Booij, G. 2009. Compounding and Construction Morphology. Lieber, R. and P. Štekauer (Eds.). 2009. *The Oxford Handbook of Compounding*: 201-216. Oxford/New York: Oxford University Press.
Booij, G. 2010. *Construction Morphology*. Oxford: Oxford University Press.
Booij, G. 2015. Word Formation in Construction Grammar. Müller, P.O., I. Ohnheiser, S. Olsen and F. Rainer (Eds.). 2015. *Word Formation: An International Handbook of the Languages of Europe. Vol. 1*: 188-202. Berlin/New York: De Gruyter.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

Booij, G. (Ed.). 2018. *The Construction of Words: Advances in Construction Morphology*. Cham: Springer.

Bybee, J. 1985. *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam/Philadelphia: John Benjamins.

Bybee, J. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Díaz Negrillo, A. 2014. Neoclassical Compounds and Final Combining Forms in English. *Linguistik Online* 68(6): 3-20.

Fábregas, A. and S. Scalise. 2012. *Morphology: From Data to Theories*. Edinburgh: Edinburgh University Press.

Fauconnier, G. and M. Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fischer, R. 1998. *Lexical Change in Present-day English: A Corpus-based Study of the Motivation, Institutionalisation, and Productivity of Creative Neologisms*. Tübingen: Gunter Narr Verlag.

Fradin, B. 2000. Combining Forms, Blends and Related Phenomena. Doleschal, U. and A.M. Thornton (Eds.). 2000. *Extragrammatical and Marginal Morphology*: 11-59. München: Lincom Europa.

Goldberg, A.E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Hayashi, H. 2024. A Construction Morphology Approach to Neoclassical Compounds and the Function of the Linking Vowel. *Languages* 9(4):1-15.

Kempf, L. and S. Hartmann. 2018. Schema Unification and Morphological Productivity: A Diachronic Perspective. Booij, G. (Ed.). 2018. *The Construction of Words: Advances in Construction Morphology*: 441-474. Cham: Springer.

Kiparsky, P. 1982. From Cyclic Phonology to Lexical Phonology. Hulst, H.V.D. and N. Smith (Eds.). 1982. *The Structure of Phonological Representations. Part* 1: 131-175. Dordrecht: Foris.

Lalić-Krstin, G., N. Silaški and T. Đurović. 2022. Meanings of *-nomics* in English: From *Nixonomics* to *Coronanomics*, How *-nomics* Has Extended Its Original Meaning to Additional Senses. *English Today*: 1-8.

Lehrer, A. 1998. Scapes, Holics, and Thons: The Semantics of English Combining Forms. *American Speech* 73(1): 3-28.

Mattiello, E. 2023. *Transitional Morphology: Combining Forms in Modern English*. Cambridge: Cambridge University Press.

Mitkov, R. (Ed.). 2022. *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.

Mittelberg, I., T.A. Farmer, and L.R. Waugh. 2007. They Actually Said That? An Introduction to Working with Usage Data through Discourse and Corpus Analysis. Gonzalez-Marquez, M., I. Mittelberg, S. Coulson and M.J. Spivey (Eds.). 2007. *Methods in Cognitive Linguistics*: 19-52. Amsterdam/Philadelphia: John Benjamins.

Plag, I. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.

Plag, I. 2021. Productivity. Aarts, B., A. McMahon and L. Hinrichs (Eds.). 2021. *The Handbook of English Linguistics.* Second edition: 483-499. Hoboken, NJ, USA: Wiley-Blackwell.

Pulcini, V. 2020. The Combining Form *multi-* in English Compounds. Maci, S., M. Sala and C. Spinzi (Eds.). 2020. *Communicating English in Specialised Domains. Festschrift for Maurizio Gotti*: 301-315. Newcastle upon Tyne: Cambridge Scholars Publishing.

Richards, J.C. and R.W. Schmidt. 2010. *Longman Dictionary of Language Teaching and Applied Linguistics*. Fourth edition. London: Pearson Education.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2041 (Lexicofocus)

Rita-Kasari, E. 2013. *The Morphological Productivity of Selected Combining Forms in English*. Unpublished MA Thesis. Helsinki: University of Helsinki.

Scalise, S. and I. Vogel (Eds.). 2010. *Cross-Disciplinary Issues in Compounding*. Amsterdam: Benjamins.

Ten Hacken, P. (Ed.). 2016. *The Semantics of Compounding*. Cambridge: Cambridge University Press.

Wiemeyer, L. 2019. The Diachronic Productivity of Native Combining Forms in American English. Wiegand, V. and M. Mahlberg (Eds.). 2019. *Corpus Linguistics, Context and Culture*: 223-252. Berlin/Boston: Walter de Gruyter.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Semantic Verbal Neology in Spanish Dictionaries: Case Study of the Lexicographic Representation of *anestesiar* and *coreografiar**

Marina Berri, *Instituto del Desarrollo Humano, Universidad Nacional de General Sarmiento / CONICET, Argentina*
(mberri@campus.ungs.edu.ar)
(https://orcid.org/0000-0003-2073-8523)
and
Victoria de los Ángeles Boschiroli, *Instituto del Desarrollo Humano, Universidad Nacional de General Sarmiento, Argentina*
(vboschir@campus.ungs.edu.ar)
(https://orcid.org/0000-0003-4136-0451)

Abstract: The aim of this paper is to reflect on the lexicographic treatment of semantic verbal neologisms in Spanish general dictionaries. Although semantic neology is a highly productive mechanism for vocabulary production, semantic neologisms seem to pose particular challenges for their inclusion in dictionaries but also in their representation, given the difficulty in detecting them. These difficulties seem to be related to the fact that a semantic neologism can be the result of a variety of changes in their primary meaning, which entails establishing different relationships with the related primary sense in a general dictionary entry or compensating for the absence of that original sense in dictionaries of neologisms. In this study, the focus is on two verbs, *anestesiar* and *coreografiar*, that have been recorded as semantic neologisms in the Observatori de Neologia databank BOBNEO. In both cases, their lexicographic relevance is shown quantitatively (based on data from corpora such as CORDE, CORPES, and Spanish Web 2023) and qualitatively, using the Theory of Norms and Exploitations (Hanks 2013). The analysis then focuses on how they are dictionarised in two widely used general online dictionaries, namely the *Diccionario de la lengua española* and the *Diccionario del español actual*. The conclusion is that both verbs are under- or misrepresented and possible actions are suggested to correct this situation both in general dictionaries and dictionaries of neologisms.

Keywords: SEMANTIC NEOLOGY, VERBS, LEXICOGRAPHIC REPRESENTATION, SPANISH LEXICOGRAPHY, GENERAL DICTIONARIES, DICTIONARIES OF NEOLOGISMS, FIGURATIVE MEANING, POLYSEMY

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Opsomming: Semantiese verbale neologie in Spaanse woordeboeke: 'n Gevallestudie van die leksikografiese voorstelling van *anestesiar* en *coreografiar*. Hierdie artikel het besinning oor die leksikografiese hantering van semantiese verbale neologismes in algemene Spaanse woordeboeke ten doel. Alhoewel semantiese neologie 'n hoogs produktiewe werktuig vir woordeskatproduksie is, wil dit voorkom of semantiese neologismes, gegewe die problematiek om hulle op te spoor, besondere uitdagings oplewer in hul insluiting en voorstelling in woordeboeke. Hierdie uitdagings blyk verband te hou daarmee dat semantiese neologismes die resultaat kan wees van verskeie veranderings in hul primêre betekenis, wat kan meebring dat verskillende verhoudings met die verwante primêre betekenis in 'n algemene woordeboek gevestig moet word, of dat daar vir die afwesigheid van daardie oorspronklike betekenis in neologismewoordeboeke gekompenseer moet word. In hierdie studie is die fokus op twee werkwoorde, *anestesiar* en *coreografiar*, wat as semantiese neologismes in die Observatori de Neologia-databasis BOBNEO opgeteken is. In albei gevalle word hul leksikografiese relevansie kwantitatief (gebaseer op data uit korpusse soos CORDE, CORPES en Spanish Web 2023), asook kwalitatief met behulp van die Teorie van Norme en Eksploitasies (Hanks 2013) aangedui. Die analise fokus vervolgens op hoe hierdie werkwoorde in twee wyd gebruikte algemene aanlyn woordeboeke, naamlik die *Diccionario de la lengua española* en die *Diccionario del español actual,* opgeneem is. Daar word tot die gevolgtrekking gekom dat beide werkwoorde onderverteenwoordig of verkeerd voorgestel word, en moontlikhede word aan die hand gedoen om hierdie situasie in sowel algemene woordeboeke as neologismewoordeboeke reg te stel.

Sleutelwoorde: SEMANTIESE NEOLOGIE, WERKWOORDE, LEKSIKOGRAFIESE VOORSTELLING, SPAANSE LEKSIKOGRAFIE, ALGEMENE WOORDEBOEKE, NEOLOGISMEWOORDEBOEKE, FIGUURLIKE BETEKENIS, POLISEMIE

1. Introduction

Semantic neology is a productive mechanism through which vocabulary is extended by metaphor, metonymy, expansion, or restriction. The inclusion in dictionaries of semantic neologisms poses specific challenges, since they are harder to detect (automatically or otherwise) than formal neologisms. In Spanish, this problem is particularly salient in relation to the category of verbs. Moreover, their lexicographical treatment in general dictionaries gives rise to further difficulties, since, as will be discussed in this paper, they require the modification of an existing article, either by adding a new sense, in the case of the emergence of a new meaning, or by modifying an existing sense, which may be necessary in the case of extensions, restrictions or other types of displacements. Furthermore, in dictionaries of neologisms, defining a semantic neologism implies referring in some way to a previous meaning that this dictionary does not include.

The aim of this paper is to examine the lexicographic treatment of the semantic verbal neologisms *anestesiar* ('to anesthetise') and *coreografiar* ('to choreograph') in two major Spanish dictionaries. For this purpose, the problems involved in the detection and inclusion in dictionaries of semantic neologisms, focusing on the specificities of Spanish lexicography, are first analysed. Then the central tenets of the Theory of Norms and Exploitations (TNE), by Patrick Hanks (2013), will be

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

explained briefly and the main findings from a previous analysis based on this theory will be summarised (Berri and Boschiroli 2023, 2024), where the modifications of meaning and changes in the combinatory patterns of neologisms are accounted for based on the notion of norm. The Theory of Norms and Exploitations provides theoretical tools to analyse modifications in the meaning of a lexical unit that are associated with changes in patterns of co-occurrences, which are typical of semantic neologisms (Renouf 2013, 2020). The distinction between norm and exploitation also provides elements to analyse the lexicographic representation (or *dictionarisation*, Bernal et al. 2020) of neologisms. The passage from exploitation to norm makes it possible to account for the *neological process* (Freixa 2022) undergone by semantic neologisms (Berri and Boschiroli 2023, 2024), suggesting it is a useful tool to measure the neologicity of a semantic neologism. Based on these findings, the representation of the semantic neologisms *anestesiar* and *coreografiar* in the Real Academia Española's *Diccionario de la lengua española* (DLE) and the *Diccionario del español actual* (DEA) will be analysed. Finally, the conclusions will be outlined, focusing on the need to understand the different types of semantic shift and their possible effects both in existing dictionary entries and in dictionaries of neologisms.[1]

2. The problem of semantic verbal neologisms in Spanish

This work focuses on verbs that are classified as semantic (sense) neologisms (VSN), that is, existing forms for which "a new meaning" emerges and spreads to different degrees (cf. Renouf 2013, 2020). In previous works (Berri and Boschiroli 2023, 2024), we have explored VSN within the framework of Hanks' TNE. It has been argued that TNE can be used as a criterion to measure their different degrees of *neologicity* — that is, the gradual property of newness in a lexical item (cf. Sablayrolles 2006, Bouzidi 2010, Cañete and Freixa 2014, Adelstein 2019). This paper focuses on two of the verbs previously analysed elsewhere to illustrate the kind of problems that arise when they are dictionarised: *anestesiar* and *coreografiar*.

VSN are interesting to study for many reasons. To begin with, when compared to other neologisms, they tend to go under- or undetected. While other neologisms can be spotted automatically because they are formally different, semantic neologisms have traditionally been identified and extracted manually, since they lack internal formal features that helps in recognising them. Given this lack of formal features, the main way to recognise and prove their novelty is by observing their surrounding syntagmatic structures: Their immediate context tends to be different from the one of their original meanings. This is a phenomenon for which, as will become evident, TNE provides an explanation. Finally, verbs have been less researched in the Spanish-speaking tradition (with the remarkable exception of Martínez Linares 2015), despite nominal semantic neologisms being widely studied in connection to their referential function and metonymic nature, especially in studies of polysemy (Adelstein 2022, Straccia 2024,

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Gerding Salas and Kotz Grabole 2016). This makes this a particularly rich and relevant field to explore.

As well as being underdetected (or, perhaps, as a result of this), VSN seem to be quantitatively underrepresented in both general language and neologism dictionaries. As the figures in Table 1 show, verbs have been at most 10 percent of all the semantic neologisms included in the Real Academia Española's *Diccionario de la lengua española* (DLE) yearly update — and usually even less. Addition of new senses to existing verbs or amends to existing definitions appear to be a peripheral, albeit slowly growing, phenomenon.

Table 1: Yearly updates — *Diccionario de la lengua española,* 23rd edition

Year	**New senses**	**Amends to senses**
2020	55 (only 2 verbs, 4 %)	32 (only 1 verb, 3%)
2021	55 (only 5 verbs, 9%)	22 (only 3 verbs, 14 %)
2022	75 (only 7 verbs, 9%)	47 (only 1 verb, 2%)
2023	58 (only 7 verbs, 12%)	—

This is consistent with how VSN are found to be represented in dictionaries of neologisms. The *Antenario. Diccionario en línea de neologismos de las variedades del español* (Antenario), which is updated monthly, had published 1 500 entries by October 2024: 32 entries are semantic neologisms, with only 2 of these being verbs. Out of 2 400 entries in the *Diccionario de neologismos del español actual* (NEOMA) (2016, no updates), 103 are semantic neologisms, with only 7 verbs.

Neither underdetection nor underrepresentation is entirely surprising. As Adelstein (2022: 330) points out, a large part of semantic neologisms are not usually perceived as neologic, in other words, they do not pass the "neologic feeling" (that is, speakers' perception that a lexical unit is neologic, Bernal et al. 2020) test, because speakers tend not to think of them as new.

3. The role of dictionaries of neologisms in Spanish-language lexicography

Unlike what happens around other national languages, there is a tradition in Spanish lexicography to publish dictionaries of neologisms, such as the *Diccionario de neologismos del español actual*, *Diccionario de neologismos online*, *Antenario*, *Diccionario de voces de uso actual*, *Nuevo diccionario de voces de uso actual*, and *1300 neologismos en la prensa argentina*. This tradition is a result of the role the Real Academia's dictionaries has had in Spanish lexicography. Despite being highly questioned on different grounds (cf. Lara 2009, Rodríguez Barcia and Moskowitz 2019, Lauría 2022), it is clearly the most prestigious and consulted dictionary for Spanish speakers. But traditionally it has adopted prescriptive, Spain-oriented policies that result in excessive precautions when it comes to including new

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

words and senses to the dictionary.

As a result, dictionaries of neologisms have assumed the role of "filling the gap" between lexical reality and its representation — they fulfil a complementary, transitional, and remedial function (Adelstein and Boschiroli 2020). This work points out that this situation is particularly complex for semantic neologisms because, unlike most formal neologisms, it involves isolating them from the entry they should eventually be added to, that is, independently from the meaning in which they originate.

What exactly semantic neologisms are is a contested issue. There is a tension between their theoretical definition and the guidelines for their detection. Semantic neologisms, as suggested above, have been defined as the creation or emergence of a new meaning for an existing unit (Sablayrolles 2003). However, when it comes to their identification, an item is classified as a semantic neologism when there is a modification (of any kind) in the meaning of the lexical unit as included and represented in dictionaries, according to IULA's Observatori de Neologia (2024: 10). This approach is clearly more practical than trying to determine the nature or status of that change — in other words, whether it is a new meaning, a new sense, or a variation of a meaning. In fact, making this distinction is impossible if you are confronted with the single use you are extracting.

These different perspectives on semantic neology have diverse impacts on lexicographic representation. If what is detected is a change of meaning, rather than a new meaning, it will necessarily affect existing senses and definitions in an entry. On the other hand, if it is a new meaning, it should probably be treated in a new sense.

The results of identifying and recording as a semantic neologism any modification of the meaning of an existing unit will vary widely. The examples collected in the databank of the Observatori de Neologia of the IULA, BOBNEO, illustrate this variety. If the relationship between the "new meaning" and the existing meaning(s) are considered, the following is found:

(a) metaphorical displacements, such as in *habitar*, whose locative meaning (*habitar un lugar*, 'to inhabit a place') becomes 'to embrace' (as in *habitar la identidad* 'to inhabit an identity');

(b) metonymic displacements, as in *abrazar* 'to hug' ('to protest by symbolically hugging collectively, e.g. by linking arms, something you wish to protect'), where the hugging becomes the form of protest;

(c) meaning extensions, as the direct objects of the neological use of *coreografiar* are not just musical pieces or dances, but also scenes from films (see Section 5.2);

(d) meaning restrictions, as in *delinear* 'to outline', whose primary meaning 'to draw the outer shape of a figure' becomes 'to apply eyeliner'; and

(e) changes in the semantic prosody, as in *amenazar* 'to threaten', in uses where the threat is something usually regarded as positive.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

This suggests both prototypical and neological meanings should be considered when dictionarised. Furthermore, there are different scenarios regarding how widespread or frequent the new sense is, ranging from occasional uses, to expanding uses (*habitar*) and very frequent uses in different domains (*anestesiar*, *dilapidar*). This cannot be determined when the use is first recorded and requires further analysis.

Finally, many VSN are derived from nouns, and semantic change may occur in the base (*mapear* 'to map', *coreografiar*) (see Section 5.2). To make matters more difficult, some are also semantic calques (*mapear*). The phenomena that end up being categorised as VSN are so different that it is necessary to adopt clear criteria to decide on their dictionarisation.

4. Theory of Norms and Exploitations (TNE) and semantic neology

The process described in Section 3 can be approached through TNE when analysing VSN (Berri and Boschiroli 2023, 2024). One of TNE's central claims or findings is summarised in Hanks (2013: 171): "When a word develops a new meaning, it develops new phraseological patterns and collocational preferences at the same time". This suggests that new senses or meanings can be discriminated by identifying phraseological patterns — some of which will be highly recurrent, others used just once, and yet others somehow frequently.

Those uses that represent the same pattern are grouped together by carrying out corpus analysis (typically of samples of lines). By sameness of pattern, both syntagmatic and semantic regularity should be understood. These conventional, recurrent uses are called *norms*. According to Hanks (2013: 92), norms are patterns of ordinary usage in everyday language, with which a particular meaning is associated. Characteristically, norms are presented in dictionaries as separate senses whose meaning is explained through definitions.[2]

To illustrate this, the norm for the most prototypical, dictionarised use of *catapultar* ('to catapult') can be represented as follows (Berri and Boschiroli 2023):

Table 2: Norm of *catapultar*

Pattern: [[Humano]] *catapulta* [[Objeto Físico]] [[Human]] *catapults* [[Physical Object]]
Sense: Arrojar un objeto con una catapulta ('To throw an object with a catapult').
Example: Hacían todo lo posible por detener el ataque lanzando flechas y **catapultando** piedras. 'They did everything they could to stop the attack by throwing arrows and **catapulting** stones.'

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Just as in English, the norm includes a human agent, the goal of the action is a physical object, and the pattern is associated with the meaning of throwing something concrete. However, Hanks (2013) explains that norms can be creatively exploited — a phenomenon he calls *exploitation*. Exploitations are a deliberate departure from an established pattern of normal word use (that is, from a norm), either in order to talk about new or unusual things, or to say old things in a new, interesting, or unusual way. Example (1) illustrates one such use of *catupultar,* in the passive voice:

(1) Se intensifican (…) vínculos que ya existían y, a la vez, se **catapultan** otros hacia el olvido.
('Existing relationships become stronger, and, at the same time, others are catapulted into oblivion.')

In this case, the verb is undoubtedly used figuratively: the object is an abstract entity (a relationship) whereas the agent is unclear. The combinatory pattern does not match the one of the norm.

According to Hanks, an exploitation can be an isolated event (analogous to a nonce word or a hapax), or may end up expanding in use. If it becomes more frequent, a new norm can emerge (with a new meaning and a different pattern).

This passage from exploitation to norm is usually gradual. Hanks (2013: 92) characterises the space between one stage and the other as a grey area. This can be assimilated to what has been called a *neological process,* defined as the "set of phenomena that occur around a word from the moment it first appears to its full establishment in use or, in other cases, its disappearance" (Freixa 2022: 51, our translation). This process involves changes in three levels: linguistic, cognitive, and social. At the linguistic or structural level, *lexicalisation* takes place. This is defined as the formal process that spans from the creation of the word until it becomes fixed and includes both semantic and formal aspects. Parallelly, at the cognitive level, *hypostatisation* takes place, that is, the lexicalised unit is incorporated into the mental lexicon of the speakers. Finally, at the social level, *institutionalisation* takes place, that is the widespread use of the neological item in a community, which often leads to its dictionarisation.

Given this diachronic dimension, it can be concluded that TNE is a relevant framework for studying neologicity, especially in understanding how semantic neology should be addressed and how it is dictionarised.

5. Case studies

In this work, the focus is on two verbs whose neologicity have been described in terms of norms and exploitations (cf. Berri and Boschiroli 2023, 2024): *anestesiar* and *coreografiar*.

These non-core neological uses will be dated based on information from *Corpus*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

diacrónico del español (CORDE) and *Corpus de referencia del español actual* (CREA), the Real Academia's historical corpora. Their frequency will be determined by looking at their most frequent collocations using the Word Sketch and Concordances functions in Sketch Engine as well as their corpus Spanish Web 2023 (esTenTen23). Entries from two recently updated general Spanish dictionaries will be analysed: (1) the Real Academia's *Diccionario de la lengua española* (DLE 23.7), and (2) *Diccionario del español actual* (DEA), a corpus-based general language dictionary first published in the 1990s, whose third edition was published online in 2023.

5.1 Case 1: *anestesiar*

Like its equivalent in English, the conventional meaning of *anestesiar* is in the medical sense. Below are typical examples and the norm associated with this pattern:

(2) No podían **anestesiarme** completamente porque mi presión sanguínea era muy baja.
('I couldn't be fully **anesthetised** because my blood pressure was too low.')

(3) **anestesiar** el ojo de un paciente afectado por glaucoma
('**to anesthetise** the eye of a glaucoma patient')

Table 3: Primary norm of *anestesiar*

Pattern: [Humano = Médico]] *anestesia* [[Humano | Animal | Parte del cuerpo]] (con [[Droga]]) [Human= Doctor] *anesthetizes* [Human | Animal | Body Part] (with [Drug])
Sense: "Quitar la sensibilidad física por medio de la aplicación de una droga". ('To render physically insensible by applying a drug').

A full description of the pattern could include a specification of lexical sets associated with the semantic types. The description could also include the fact that the purpose is usually surgery, and that it is implied that the human in the second slot is a patient.

However, there are other uses that have been recorded extensively and move away from the norm, as can be seen at syntagmatic level:

(4) Más que confundido, el consumidor **está anestesiado**, abrumado ante el exceso de estímulos y ahora prácticamente solo responde al precio.
('The consumer is not so much confused as it **is numbed**, overwhelmed by the excess of stimuli, and now they almost only respond to price.')

(5) Berlusconi ha conseguido **anestesiar** a buena parte de la población.
('Berlusconi has managed to **numb** a large part of the population.')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

In (4), the human is not a patient, but a consumer, and instead of drugs, there are stimuli. In (5), Berlusconi is a politician, instead of a doctor, and the goal is the population.

As a matter of fact, these variations to the norm are not new. A search of the form "anestesiad*" in CORDE shows that, in total, there are 10 variations out of 28 examples between 1884 and 1969. The first recorded literary example (6) is from 1905 "anestesiada para el sentimiento", 'numb for feelings'), but already in 1929 there is an example from the press ("anestesiado por el miedo", 'numbed by fear').

(6) La veo **anestesiada** para el sentimiento. (1905)
('She seems to be **numb** for feelings.')

(7) En el fondo de la caverna … **anestesiado** por el miedo, estaba un hombre — yo — ... Todo mi ser temblaba, sudando … (1929)
('**numbed** by fear')

These are not occasional exploitations either. Sketch Engine's Word Sketch of the lemma *anestesiar* reveals that there are some regular collocations, such as the nouns *sociedad* 'society', *conciencia/consciencia* 'consciousness', *ciudadanía* 'citizens' or *sensibilidad* 'sensitivity', or the adverbs *emocionalmente* 'emotionally' and *moralmente* 'morally', that co-occur consistently and repeatedly with the verb. Moreover, whereas in the Word Sketch there are 677 cases of concordances with *paciente* 'patient' and 300 with *dolor* 'pain' — which, according to CORPES, are the two most frequent noun co-occurrences of *anestesiar* — there are 509 of *sociedad* and 401 of *conciencia/consciencia*. This suggests they are frequent patterns.

Based on this data, these new uses can be described as a secondary pattern:

Table 4: Secondary norm of *anestesiar*

Pattern: [[Institución | Humano]] anestesia [[Grupo humano | Humano]] (con [[Entidad = negativa]]) (por lo general sin agente) [[Institution | Human | Technology]] anesthetizes [[Human Group | Human]] (with [[Entity = negative]]) (usually agentless)
Sense: "Quitar la sensibilidad emocional y/o la capacidad de reacción" ('To deprive someone of their emotional sensitivity and/or their capacity for reaction')
Lexical set: *sociedad* ('society'), *público* ('public'), *los jóvenes* ('youth'), *la población mundial* ('world population') …
Example: Crear fake news para un público **anestesiado** frente a la violencia. ('Creating fake news for an audience that is **numb** to violence.')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

The analysis has shown regular lexical sets as second argument, and also an overall negative prosody in connection to the meaning, which has moved from the physical to the emotional.

How is this secondary norm accounted for in general dictionaries? Despite its frequency and age, there is no record or trace of this pattern in the DLE. *Anestesiar* is represented as a monosemous verb in its medical meaning. The noun base *anestesia* is used in the definition, but there are no further explanations in the noun either.

anestesiar
1. tr. Privar total o parcialmente de la sensibilidad por medio de la anestesia.
Sin.: **dormir**, **adormecer**, **narcotizar**, **insensibilizar**.

'To deprive totally or partially of sensitivity through anesthesia'

Figure 1: *Anestesiar* in DLE

The DEA, on the other hand, looks more promising. Unlike the DLE, it does offer two senses and two explanations:

anestesiar
(*conjug* 1a)
tr
1 Privar total o parcialmente de la sensibilidad [a alguien o a una parte del cuerpo], esp. por medio de un fármaco y con fines quirúrgicos. Tb fig.
2 Apagar o adormecer [algo].
'1. To deprive [sb or sb's part of the body] totally or partially of sensitivity esp. through a drug and for surgical purposes Also fig. 2. To dull or dumb [sth].'

Figure 2: *Anestesiar* in DEA

Sense 1 is, unsurprisingly, the medical sense. The style of the definition is noticeably similar to the pattern description of the first norm (Table 3), suggesting it is based on corpus data. However, after the definition there is a label "tb fig." ('also figurative'), suggesting there is a non-conventional use which in fact is not explained.[3]

Contrasting with the analytic style of the definition of the first sense, in the second sense two near synonyms (*apagar* 'to deaden' and *adormecer* 'to benumb') are given as explanations. These are highly polysemous verbs which could easily replace the first meaning, so the result is confusing.

The dictionary offers an advanced view display, where a variety of real-life examples are shown:

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

anestesiar
(conjug 1a)
tr
1 Privar total o parcialmente de la sensibilidad [a alguien o a una parte del cuerpo (cd)], esp. por medio de un fármaco y con fines quirúrgicos. Tb fig.
Ya 28.10.70, 5: Secuestro del jefe de las fuerzas aéreas ecuatorianas ... El chófer del vehículo fue anestesiado.
DPa 31.8.89, 22: Se le somete [al animal] a un choque eléctrico para anestesiarle.
SElM 16.3.03C: ¿La televisión nos anestesia o es, como dice Susan Sontag, la impotencia frente al horror?
2 Apagar o adormecer [algo].
SAbc 16.3.69, 54: El hielo ... anestesia las sensaciones nauseosas.
Merino Río 104: Recuperas la conciencia urbana que otra vez el espacio silvestre te había anestesiado.
Millás Algo 53: Nevenka iba metiéndose orfidales de manera metódica, para anestesiar el dolor moral.
MMolina Jinete 449: No era serenidad, sino una lucidez anestesiada, el pensamiento obsesivo de que yo no estaba verdaderamente allí.

Figure 3: *Anestesiar* in DEA: Advanced view

In sense 1, the first two examples illustrate the conventional meaning, whereas the third one ('Does TV numb us or is it, as Susan Sontag says, impotence in the face of horror?') matches the second norm given in Table 4. However, the examples in the second sense mix the medical meaning ('ice... anesthetises nauseous feelings', in the first sentence) with exploitations (such as 'wilderness had anesthetised/numbed you', in the third example, or 'numb lucidity', in the last one).

As a result, while the prototypical norm is properly represented, the new one, albeit hinted at, is not. Beyond the first norm, what is shown reads, indeed, like a grey area. Neither of the strategies adopted (the "figurative" label in sense 1, the synonymic definition in sense 2, the examples) account for the semantic and phraseological specificity of the secondary norm.

5.2 Case 2: *coreografiar*

The second case is *coreografiar* ('to choreograph'). The conventional primary norm, which is the same as in English (associated with dance and music), can be described as follows:

Table 5: Primary norm of *coreografiar*

Pattern: [[Humano]] *coreografía* [[Pieza musical | Actividad = Danza]] [[Human]] *choreographs* [[Music piece | Activity = Dance]]
Sense: "Planificar los movimientos de una danza al compás de una música". ('Plan/arrange to music the movements of a piece of dance.')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Set léxico: *pieza, ópera, baile, musical …*
Lexical set: *espectáculo* ('show'), *pieza* ('piece'), *ópera* ('opera'), *danza* ('dance'), *musical* ('musical') …

Example:
Se propuso **coreografiar** la Novena Sinfonía de Beethoven.
('to **choreograph** Beethoven's Ninth Symphony')

In this case, exploitations such as example (8) were found, where "his own ending" means "his death". This was probably not in relation to music, a show, or dancing. Instead, the verb is used creatively in an original pattern.

(8) ¿Fue capaz de **coreografiar** su propio final? ¿Convirtió su muerte en su actuación definitiva?
('Did he manage to **choreograph** his own ending? Did he turn his life into his final performance?')

Apart from exploitations, other uses that appear regularly and frequently in the corpus were also found. These include *coreografiar una escena de sexo* ('to choreograph a sex scene') and *coreografiar una escena de acción* ('to choreograph an action scene'), both associated to cinema:

(9) ¿Qué porcentaje de esa mirada intimista permanecerá con vida en un relato repleto de escenas de acción **coreografiadas** hasta el más mínimo detalle?
('What percentage of that intimist look will survive in a narrative that is full of action scenes that **were choreographed** to the last detail?')
(10) La británica se encarga de **coreografiar** las escenas sexuales de las películas.
('The Briton is in charge of **choreographing** sex scenes in films.')

These secondary uses seem to be quite more recent than the ones in the previous case. A search of "coreografi*" in CREA reveals that out of 44 results, four deviate from the primary norm, starting in 2002. Two of these are exploitations, the other two recurrent secondary uses, such as the ones identified in examples (9) and (10):

(11) Los empleados de traje y corbata entraban, comían, tomaban café, salían, y el amplio espacio libre entre las mesas y el discreto paso rápido de los camareros y los trajes de los comensales contribuían a crear un ambiente **coreografiado**. (2001)
('The employees in suits and ties would come in, eat, have coffee, leave, and the ample room between the tables and the discreet, quick step of the waiters and the diners' outfits contributed to create a **choreographed** setting.')
(12) violencia …. cuidadosamente **coreografiada** [en un film] (2003)
('carefully **choreographed** violence' [in a film]')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Moreover, the Word Sketch of the verb shows it is a very frequent phenomenon. There are fewer numbers of primary sense collocations (69 altogether, including *ballet, baile* 'dance', *danza* 'dance', and *rutina* 'routine') than of secondary sense collocations (121, including *secuencia* 'sequence', *escena* 'scene', and *pelea* 'fight'). This shows that recurrent new uses cannot be regarded as mere exploitations and need to be described in terms of a new norm.

However, in this case, if the focus is on the sense and the phraseological pattern, rather than a new norm, there seems to be an *extension* of the existing norm. The prosody has not changed, and although the music element has become optional, the basic meaning of 'planned sequence of movements' remains. As a result, the extended norm shown in Table 6 is proposed, rather than a separate new one.

Table 6: Extended (primary) norm of *coreografiar*

Pattern: [[Humano]] *coreografía* [[Espectáculo | Parte de un espectáculo = Danza, Obra musical, Película]] [[Human]] *choreographs* [[Show | Part of a show = Dance, Music Piece, Film]]
Sense: "Planificar en detalle la secuencia de movimientos de un espectáculo o de una parte de un espectáculo". 'Plan in detail a sequence of movements in a performance or part of a performance'.
Lexical set: *espectáculo* 'show', *escena* 'scene', *batalla* 'battle', *pieza* 'piece', *ópera* 'opera', *baile* 'dance', *musical* 'musical' …

As in the case of *anestesiar*, the DLE only records the primary norm in connection to dance. Also, the definition cross-refers to the noun, focused entirely on the dance meaning.

coreografiar
Conjug. c. *enviar.*
1. tr. Hacer la coreografía de un espectáculo de danza o baile.
'Make the choreography of a dance show'

Figure 4: *Coreografiar* in DLE

And once again, the DEA offers two senses and definitions. The first one explains the prototypical sense; the definition is analytical and, like that of the Real Academia, cross-refers to the noun ('to create a choreography [for a musical piece]'). The second one would be expected to explain the recurrent variations described

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

before. However, this sense refers to literary uses (labelled as such with the abbreviation "lit."), which are rather cryptically explained as 'Give the character of choreography', without any specification of the object.

coreografiar
(conjug 1c)
tr
1 Crear una coreografía [2] [para una obra musical]. Tb abs.
2 (lit) Dar carácter de coreografía [2] [a algo].

Figure 5: *Coreografiar* in DEA

The examples in the advanced view of sense are not entirely helpful in terms of specifying the second definition.

coreografiar
(conjug 1c)

tr
1 Crear una coreografía [2] [para una obra musical (cd)]. Tb abs.
País 14.1.85, 31: El interesante empeño de Medea .., coreografiado por José Granero, alcanzó su más alta temperatura dramática cada vez que aparecía Manuel Vargas.
Nor 7.5.08In: La compañía tiene sede en Burgos, donde coreografían, ensayan y reciben clases.
2 (lit) Dar carácter de coreografía [2] [a algo (cd)].
ElM 23.2.96In: En vez de ofrecer una carroñera sesión de patadas, golpes y navajazos, opta por casi coregrafiar las peleas entre los diestros y habilidosos practicantes de las artes marciales.
Trueba Perder 202: Nadie debía andar a solas, todo se coreografiaba para evitar encuentros indeseados.

Figure 6: C*oreografiar* in DEA: Advanced view

Although the first example does show one of the collocations we identified as typical of the extended use of the verb ("… opta por coreografiar las peleas entre diestros y habilidosos practicantes de las artes marciales" 'chooses to almost choreograph the fights among the skilled, dexterous martial arts practitioners'), the use of *casi* 'almost' seems to suggest at this historical point it might have still been an exploitation, like the one found in the second example, where the verb's object is *todo* 'everything', too vague to complement the lack of precision of the explanation. Adding to the confusion, the first example is not literary, as the label suggests. The whole of the second sense, then, fails to account for a regular use, and the extended uses remain mis- or unrepresented.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

6. Discussion

To assess the adequacy of the lexicographic representation of *anestesiar* and *coreografiar*, it is necessary to take a step back and clarify the role of monolingual general dictionaries aimed at human users is in terms of the kind of lexical information they should dictionarise. Of the many answers to this question, the focus here is on two.

The first is Hanks (2013: 90-91) who, considering the content of dictionaries, argues: "The facts recorded in dictionaries are, or should be, the very facts about the shared linguistic conventions on which participants in conversation rely." The second is Kilgarriff's (1992) claim that a usage-type deserves to be listed in a dictionary when it is "sufficiently frequent and insufficiently predictable".

In the present analysis of semantic neological verbs, it was found that their structural and semantic reality differs significantly from their representation in general dictionaries. Looking at the dictionary entries analysed from Hanks' and Kilgarriff's perspectives, a number of observations can be made.

Firstly, although it makes sense to record explicitly new uses whose stabilisation is not entirely predictable — that is, whose neological process is not complete — the analysis shows that major general dictionaries tend to ignore altogether documented variations from primary norms, in spite of their low degree of neologicity.[4]

On the other hand, when these uses are recorded, as in DEA, there are several inconsistencies. First, examples mix exploitations (that are not frequent) with stabilised secondary uses (that are frequent, hence socially shared and not predictable). Second, the use of the label "fig." for figurative uses is unpredictable: Sometimes it is included (as in *anestesiar*) whereas sometimes not (as in *coreografiar*). It also seems to cover unlikely cases, that is, unusual uses, rather than conventionalised ones. It thus loses any explanatory power it may have — hence, it is unsuitable as a replacement for definitions.

Moreover, in these cases, both the label "fig." and those inaccurately broad definitions that cover new meanings, extended meanings *and* exploitations do not specify collocations. As a result, they fail to grasp both the structural and semantic stability that is characteristic of a secondary use. Overall, there seems to be a confusion between exploitations and those uses that fall into the grey area described by Hanks and that are closer to a brand-new norm, as in the case of *anestesiar*, or a modification or extension of a previous norm, as in *coreografiar*.

7. Concluding remarks

To conclude, some considerations need to be taken into account when dictionarising VSN. Regarding general dictionaries, especially pertaining updating them, it is important to adopt a systematic approach to the inclusion of VSN that anticipates how to deal with the grey area between exploitations and secondary or new norms. This should specify the criteria to be applied to distinguish these

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

cases, as well as how extended and "new" meanings will be described — whether and how they will be incorporated into existing definitions, or as subsenses, and what kind of explanations there should be for them, ideally accounting for the most frequent collocations. The use of the label "fig." could perhaps be reserved for (very) grey area examples, where the collocations are increasingly vague and the meaning increasingly despecified, such as the ones in the examples below:

(13) Es preciso **coreografiar** una experiencia de cliente memorable y positiva, donde todos los empleados cuentan.
('It is necessary to **choreograph** an experience that invites and motivates employees to contribute willingly and happily to their potential at work.')

(14) Un buen tramo del evento [una conferencia de prensa] de 4 horas de duración, en gran parte **coreografiado**, se centró en …
(A large part of the 4-hour-long event [a press conference], largely **choreographed**, focused on …)

Regarding dictionaries of neologisms, given their remedial role in Spanish lexicography, they seem to be the best place to account for the grey area. The systematic inclusion of VSN needs to be planned, and there is a need for ways to account for different "semantic modifications" of existing meanings explicitly, such as the extension of *coreografiar* (from dance to movies). On the one hand, there should thus be room for grey area cases, while on the other, primary senses should be made explicit somehow (for example, in a note) to account for their status.

Finally, regarding the detection and recording of VSN, existing protocols need to be refined to implement more systematic ways of detecting them, possibly including new technology.

Endnotes

1. This paper is the result of our concerns as members of the Red de neología del español Antenas Neológicas network (https://www.upf.edu/web/antenas/presentacion), with nodes from six Spanish speaking countries (Argentina, Chile, Colombia, Mexico, Peru, and Spain). One of our main tasks is extracting press neologisms, both automatically and manually, and recording them in the Observatori de neologismos databank BOBNEO (Institut de Lingüística Aplicada, Universitat Pompeu Fabra https://bobneo.upf.edu/inicio.html). We also work actively as writers and editors in a recent development of Antenas, the *Antenario*, an online dictionary of neologisms of six different varieties of Spanish.
2. Hanks developed the *Pattern Dictionary of English Verbs* (PDEV), available at https://pdev.org.uk/. For Spanish, the project *Verbario* (https://www.tecling.com/verbario/) describes the patterns of verbs in Spanish (Renau and Nazar, in process).
3. The label "fig." has been widely discussed and debated in the literature. For a critical review of its use in Spanish lexicography, see Martínez Gajardo (2011), and for general lexicography, see Battaner Arias (2010).

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

4. This is not only the case in the dictionaries we analyzed, but also in other general language dictionaries that were not included in this study, such as the *Larousse* and the *Diccionario de uso del español*.

References

A. Dictionaries

Adelstein, A., I. Kuguel and G. Resnik. 2008. *1300 neologismos en la prensa argentina.* Los Polvorines: Universidad Nacional de General Sarmiento.

Alvar Ezquerra, M. 1994. *Diccionario de voces de uso actual.* Madrid: Arco/Libros.

Alvar Ezquerra, M. 2003. *Nuevo diccionario de voces de uso actual.* Madrid: Arco/Libros.

Antenario. Diccionario de neologismos de variedades del español.
https://antenario.wordpress.com/ [4 December 2024]

Diccionario de la lengua española **(DLE).** Twenty-third edition**.**
https://dle.rae.es/ [4 December 2024]

Diccionario del español actual **(DEA).** Third edition.
https://www.fbbva.es/diccionario/ [5 December 2024]

Diccionario de neologismos del español actual **(NEOMA).**
https://www.um.es/neologismos/ [4 December 2024]

Diccionario de neologismos online.
https://repositori.upf.edu/handle/10230/5993?locale-attribute=es [2 September 2024]

Larousse. https://red-larousse.com.mx/Dictionary [4 December 2024]

Moliner, M. 1967. *Diccionario de uso del español*. Madrid: Gredos.

Verbario. http://www.verbario.com/ [4 December 2024]

B. Digital tools

BOBNEO. Observatori de Neologia.
https://bobneo.upf.edu/ [4 December 2024]

Corpus diacrónico del español **(CORDE).**
https://www.rae.es/banco-de-datos/corde [4 December 2024]

Corpus de referencia del español actual **(CREA).**
https://www.rae.es/banco-de-datos/crea [4 December 2024]

Sketch Engine. https://www.sketchengine.eu/ [4 December 2024]

C. Other literature

Adelstein, A. 2019. Hacia una definición de neologicidad en semántica. Perassi, M.L and M. Tapia (Eds.). 2019. *Palabras como puentes. Estudios lexicológicos, lexicográficos y terminológicos desde el Cono Sur*: 209-230. Córdoba: Editorial Buena Vista.

Adelstein, A. 2022. Neología y semántica: grados de neologicidad en el ámbito nominal. Bernal, E., J. Freixa and S. Torner (Eds.). 2022. *La neología del español. Del uso al diccionario*: 327-345. Madrid/ Frankfurt am Main: Vervuert.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2042 (Lexicofocus)

Adelstein, A. and V. Boschiroli. 2020. Dictionarization and Lexical Variation in Dictionaries of Spanish Neologisms. *Word* 66(4): 292-316.

Berri, M. and V. Boschiroli. 2023. Neologicidad metafórica y grados de inestabilidad. Hernández, P.C. and M.G. Galvani Gelusini (Eds.). 2023. *Estudios SAEL 2023*: 75-88. San Miguel de Tucumán: Editorial Humanitas.

Berri, M. and V. Boschiroli. 2024. La neología semántica de los verbos: criterios estructurales para determinar grados de neologicidad. *Annali — Sezione romanza* 66(1): 31-48.

Battaner Arias, P. 2010. La marca figurado como marca de uso. Bernal, E., S. Torner and J. DeCesaris (Eds.). 2010. *Estudis de lexicografía 2003–2005*: 39-60. Barcelona: Institut Universitaria de Lingüística Aplicada.

Bernal, E., J. Freixa and S. Torner. 2020. Criterios para la diccionarización de neologismos: De la teoría a la práctica. *Revista Signos* 53(104): 592-618.

Bouzidi, B. 2010. Néologicité et temporalité dans le processus néologique. *Synergies Algérie* 9: 27-36.

Cañete, P. and J. Freixa. 2014. Filtros de neologicidad e inclusión lexicográfica. Camus, B. (Ed.). 2014. *Morfología y diccionarios*: 9-18. A Coruña: Universidade da Coruña.

Freixa, J. 2022. El proceso neológico. Bernal, E., J. Freixa and S. Torner (Eds.). 2022. *La neología del español. Del uso al diccionario*: 51-72. Madrid/Frankfurt am Main: Vervuert.

Gerding Salas, C. and G. Kotz Grabole. 2016. Distancia semántica y neológica léxica. *Literatura y lingüística* 33: 289-312.

Hanks, P. 2013. *Lexical Analysis. Norms and Exploitations.* Cambridge, MA: MIT Press.

Kilgarriff, A. 1992. *Polysemy*. Unpublished PhD dissertation. Brighton: University of Sussex at Brighton.

Lara, L.F. 2009. *Lengua histórica y normatividad.* Mexico: El Colegio de México.

Lauría, D. 2022. *Lengua y política. Historia crítica de los diccionarios del español de la Argentina.* Buenos Aires: Eudeba.

Martínez Gajardo, C. 2011. Revisión crítico-bibliográfica en torno a la marca fig. en diccionarios de uso del español. *Boletín de filología* 46(2): 219-234.

Martínez Linares, A. 2015. Cuestiones en torno al tratamiento de los neologismos semánticos: entre la intuición, el análisis lingüístico y el diccionario. Cuniţă, A. and C. Lupu (Eds.). 2015. *Neologie, neologism. Concepte, analize*: 69-100. Bucarest: Universidad de Bucarest.

Renouf, A. 2013. A Finer Definition of Neology in English: The Life-cycle of a Word. Hasselgård, H., J. Ebeling and S. Oksefjell Ebeling (Eds.). 2013. *Corpus Perspectives on Patterns of Lexis*: 177-208. Amsterdam: John Benjamins.

Renouf, A. 2020. Semantic Neology: Challenges in Matching Corpus-based Semantic Change to Real-world Change. Rautionaho, P., A. Nurmi. and J. Klemola (Eds.). 2020. *Corpora and the Changing Society: Studies in the Evolution of English*: 79-111. Amsterdam: John Benjamins.

Rodríguez Barcia, S. and A. Moskowitz. 2019. An Authentic Pan-Hispanic Language Policy? Spain as the Point of Reference in the Spanish Royal Academy's *Diccionario de la lengua española*. *International Journal of Lexicography* 32(4): 498-527.

Sablayrolles, J.F. 2003. Le sentiment néologique. Sablayrolles, J.F. (Ed.). 2003. *L'innovation lexicale*: 156-167. Paris: Honoré Champion.

Sablayrolles, J.F. 2006. La neologie aujourd'hui. Gruaz, C. (Ed.). 2006. *À la recherche du mot: de la langue au discours*: 141-160. Limoges: Éditions Lambert-Lucas.

Straccia, J. 2024. Aspecto léxico y resemantización. *RASAL Lingüística* 2: 69-89.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

Digging for *-ings*: A Survey of Selected (Pseudo)-Anglicisms in Dictionaries of Modern Greek*

Anna Vacalopoulou, *Institute for Language and Speech Processing, Athena Research Center, Athens, Greece*
(*avacalop@athenarc.gr*) (*https://orcid.org/0000-0002-1509-2951*)

Abstract: Anglicisms, a type of lexical borrowing, occur when a language incorporates English lexical items into its vocabulary, either verbatim or transliterated. Pseudo-Anglicisms are also notable; these appear English due to their form but are not genuinely English. This paper examines the inclusion and treatment of Anglicisms and pseudo-Anglicisms in Modern Greek dictionaries, focusing on nouns that end in *-ing*, a suffix typical of English morphology. The methodology involves a macro-structural comparison of five Greek dictionaries to assess their inclusiveness in addressing this phenomenon. The microstructural analysis investigates the presence of Greek equivalents within the dictionaries. Findings are cross-checked against a corpus to determine which alternatives Greek native speakers are more familiar with. Additionally, the research explores elements such as the domains, and styles of the identified lexical items to reveal trends in their coinage and usage.

Keywords: NEOLOGY, NEOLOGISMS, ANGLICISMS, PSEUDO-ANGLICISMS, LEXICAL BORROWING, GREEK LEXICOGRAPHY, GREEK DICTIONARIES, CRITICAL LEXICOGRAPHY

Opsomming: Op soek na *-ing*-gevalle: 'n Opname van uitgesoekte (pseudo-) anglisismes in woordeboeke van Moderne Grieks. Anglisismes, 'n tipe leksikale ontlening, kom voor wanneer 'n taal Engelse leksikale items in sy woordeskat inkorporeer, hetsy woordeliks of getranslitereerd. Pseudo-anglisismes is ook opmerklik; hulle lyk vanweë hul vorm Engels, maar is nie werklik Engels nie. Hierdie artikel ondersoek die insluiting en hantering van anglisismes en pseudo-anglisismes in Moderne Griekse woordeboeke, en fokus op selfstandige naamwoorde wat eindig op *-ing*, 'n agtervoegsel kenmerkend van die Engelse morfologie. Die metodologie behels 'n makrostrukturele vergelyking van vyf Griekse woordeboeke om hul inklusiwiteit in die hantering van hierdie verskynsel te bepaal. In die mikrostruktuuranalise word die teenwoordigheid van Griekse ekwivalente in die woordeboeke ondersoek. Bevindinge word dubbel gekontroleer teen 'n korpus om te bepaal met watter alternatiewe Griekse moedertaalsprekers meer vertroud is. Daarbenewens word elemente soos die domeine en style van die geïdentifiseerde leksikale items ook ondersoek om tendense in hul skepping en gebruik te onthul.

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

Sleutelwoorde: NEOLOGIE, NEOLOGISMES, ANGLISISMES, PSEUDO-ANGLISISMES, LEKSIKALE ONTLENING, GRIEKSE LEKSIKOGRAFIE, GRIEKSE WOORDEBOEKE, KRITIESE LEKSIKOGRAFIE

1. Introduction

The emergence of Anglicisms and pseudo-Anglicisms in modern languages reflects the dynamic interplay between cultures and languages in a globalised world. The linguistic influence of English on other languages is also known as Anglicisation, and it is a worldwide phenomenon through which English has gradually become more influential over other languages. As a result, English is spoken either as a native or a second language by approximately 17,7% of the world's population (Figure 1). Additionally, proficiency in English is considered an asset in the global market context, leading to an increasing number of people who can use the language to some extent. Although the specific number is hard to calculate, it has been estimated at approximately two billion people globally, approximately one out of four people on the planet (Prior 2024). This percentage is particularly high in Greece, as an impressive 64,5% of the population reports having some knowledge of the English language (Hellenic Statistical Authority 2022).

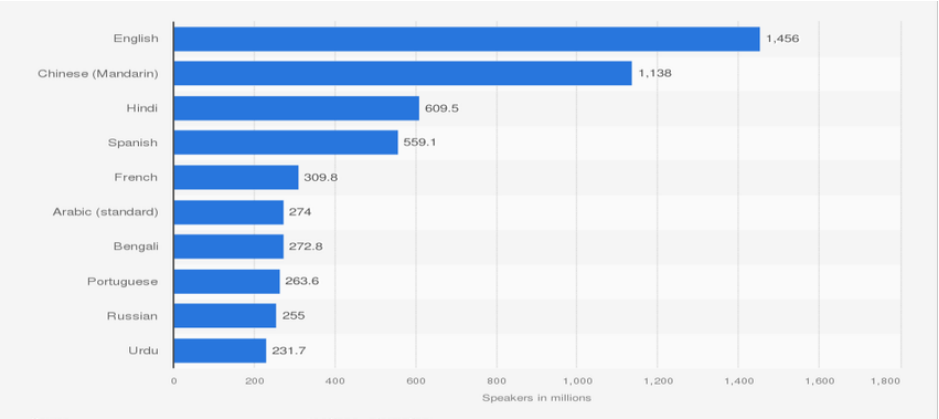


Figure 1: The most spoken languages worldwide in 2023 by number of speakers in millions (Statista 2023)

As lexical borrowing reflects the influence of one language system over another, these phenomena echo the predominance of English. Well-established as the lingua franca across numerous countries, peoples, and cultures, English has become a leading exporter of neologisms and a regular starting point for lexical innovation. This paper examines how selected dictionaries of Modern Greek have treated

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

Anglicisms and pseudo-Anglicisms, specifically focusing on nouns formed with the *-ing* suffix, a typical English formation. Rather than discussing the processes behind the incorporation of these linguistic borrowings by editorial teams, this survey adopts a critical lexicography perspective, focusing on how these borrowings are reflected in existing lexicographic resources.

2. Background

2.1 Collocation analysis: Syntagmatic level

Anglicisms are lexical items borrowed from English and integrated into the lexicon of another language. These borrowings can manifest across various linguistic levels, such as phonology, grammar, morphology, and vocabulary. The global prominence of English, rooted in historical, social, and technological developments, has led to an influx of Anglicisms in numerous languages, including Greek. The rise of Anglicisms on an international level has been particularly notable in key domains that have established common, globally used vocabularies. Such domains are predominantly the following: international business, with terms like *μίτινγκ* ('meeting') or *πρότζεκτ* ('project'); technology, with lexical items related to computers and the internet, such as the word *κομπιούτερ* ('computer') itself, or *σέλφι* ('selfie'); and the media landscape, which often propagates Anglicisms through popular culture and trends. Furthermore, recent studies in neology reveal a tendency towards the incorporation of new terms in the lexicon of some languages in the context of crises-related discourse (cf. Michalopoulou 2022; Storjohann 2024). Notable examples include the continuing climate crisis, the Ukraine war crisis, and the COVID-19 crisis. Indeed, the pandemic stands out as an example of the rapid introduction of novel lexical items into everyday language around the world. In the case of Greek, this has resulted in the incorporation of Anglicisms such as *λοκντάουν* ('lockdown') and *ράπιντ τεστ* ('rapid test'), as well as pseudo-Anglicisms such as 'click away' (not usually transliterated) to denote a newly-introduced method of contactless purchasing.

It is not uncommon for Anglicisms to carry significant social and political implications, influencing language policies in various nations. For instance, the Académie Française, in keeping with its role as the linguistic guardian of France, has regularly warned the nation against the increasing use of English in both the public and private sectors, arguing that it may lead to communication problems and challenge social cohesion (AFP in Paris 2022). Moving from institutional warnings to legal measures, the Italian government proposed a law in 2023 against "the growing Anglomania", anticipating fines of up to €100 000 for the use of foreign vocabulary in official communications of the Italian state (Carbonaro 2023). Along similar lines, the Russian government declared a ban on the use of foreign words by officials when performing their duties, allowing for the use of some "foreign words which do not have widely used corresponding equivalents in Russian" (Reuters 2023). Although no similar official measures have been taken

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

in the Greek context, lexical borrowing is often stigmatised as a danger to the integrity of the Greek language (cf. Anastassiades-Symeonides 2001).

2.2 Incorporation patterns in written Greek

This study focuses on a particular type of Anglicism, namely the incorporation of nouns ending in *-ing* into Greek vocabulary, which is a common suffix of English morphology but does not traditionally appear in Greek. English loanwords appear in written Greek with varying orthographic representations. They may either be adopted verbatim, spelled out in the Latin alphabet, or may be transliterated into the Greek alphabet (see Figure 2). It is also often the case that both forms appear in language corpora over the same period of time, possibly reflecting an uncertainty about how to best integrate these terms into Greek. These inconsistencies remain in practice, resulting in a hybrid approach to incorporating new terms that reflects ongoing debates about language purity and adaptation.

Figure 2: Verbatim incorporation, transliteration, and dual presentation of three Anglicisms in Modern Greek: *woke* (left), *sandwich* (center), and *CD* (right). [Picture left: https://metabook.gr/books/ardin-teukhos-128-fevroyarios-martios-2024-542477 (added graphics author); Picture middle: own photo and added graphics author; Picture right: https://vendora.gr/items/de1o00/cd-sinti-programata-afthentika.html?srsltid=AfmBOoroubcT9RjaDA8XQKSfyXcTnCoJIA_Ot8aRBUAUfTWOC5MGg3pK (added graphics author)]

When lexical borrowings are incorporated into Greek, the general rule is that their spelling should be kept in the simplest possible form, according to official school grammar textbooks. Following these guidelines, the *ng* consonant cluster in loanwords is to be transliterated as *γκ*, resulting in the loss of the *n* sound in Greek (Triandaphyllidis et al. 1941: 428). However, there is no consensus regarding this rule's application either in general language corpora or in dictionaries.

2.3 Pseudo-Anglicisms

In addition to genuine Anglicisms, the Greek language also features pseudo-Anglicisms, or false Anglicisms. These are words that seem English but do not

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

exist in the same form or meaning in English. As creations of Greek vocabulary, these terms either have no equivalent in English or carry different meanings from similar-looking English words. Although these are not really loanwords, they still reflect the impact of English on Greek. Well-known examples include *ερκοντίσιον* ('air-condition' meaning 'air-conditioning'), or *μπάσκετ* ('basket' meaning 'basketball') and other shortened forms of sports terms. The earlier-mentioned *click away* (not usually transliterated), is a different type of pseudo-Anglicism coined during the COVID-19 pandemic to describe a specific contactless purchasing method. This latter example particularly highlights the creative linguistic processes at play within Greek, as it demonstrates how new terms can be created with shifted meanings and specialised usage.

3. The study

3.1 Object of study

The present study focuses on *-ing* nouns, which are formed from the present participle and gerund forms of verbs, and how these are represented in Greek lexicography. This morphological feature is characteristic of English and has been observed in the vocabulary of other languages where such formations are not typically native. For instance, Harris (2010: 19) and Šabec (2018: 70) identify loanwords ending in *-ing* as one of the most frequently reported Anglicisms in French and Slovenian respectively. Similarly, *-ing* formations have been pinpointed as characteristic forms of Anglicisms in Greek by authors such as Xydopoulos (2019: 433). The suffix is also known to produce several pseudo-Anglicisms such as the ones reported by Stamenov (2015: 169-171) in Bulgarian by adding *-ing* to native words like *manastiring* ('monastery visit', derived from *manastir* 'monastery') and *balkoning* ('jumping off a balcony', from *balkon* 'balcony'). The variant *balconing* also exists in Spanish, as do several more *-ing* pseudo-Anglicisms: *edredoning*, *kitchening*, *puenting*, and *sofing* (Sánchez Fajardo 2018: 336).

3.2 Methodology

In order to explore the inclusion and presentation of *-ing* formations in current Greek lexicography, five dictionaries were selected for analysis. These encompass three professionally edited works and two crowd-sourced lexicons. As expected, the professional dictionaries provide a structured approach to lexicography, while the crowd-sourced entries reflect a more spontaneous, though variable, inclusion of neologisms. The lexicographic works that were used are the following:

— *Practical Dictionary of Modern Greek*[1] (HLNG 2014), an institutional work by the Academy of Athens published in print and available online.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

— *Modern Greek Dictionary*[2] (LKN 1998), another institutional dictionary by the Centre for the Greek Language available in both print and online formats.
— *Big Electronic Dictionary of Modern Greek*[3] (MILNEG), an online dictionary by a commercial publisher that is regularly updated.
— *Slang.gr*[4], an online, crowdsourced dictionary of Greek slang.
— *Wiktionary*[5], the Greek version of the online, crowdsourced lexicographic project run by the Wikimedia Foundation.

The first step of this study involved the identification of those *-ing* formations that appear as transliterated entries in the above five dictionaries by searching their electronic versions using wildcards. From the resulting list, proper nouns were excluded unless they are also used in a metaphorical sense (e.g., *Browning*, a type of machine gun designed by the company of the same name) as were words that are not participle or gerund forms (such as *darling*, *ring*, or *string*). These findings were then categorised as either Anglicisms or pseudo-Anglicisms (by checking their existence and use in English) and analysed according to their orthographic representation. Next, the five dictionaries were compared regarding their treatment of this particular type of formation to determine whether any patterns emerge based on dictionary typology. Finally, the findings were cross-checked against the Greek Web 2019 (elTenTen19) corpus[6] to determine which alternatives Greek native speakers are more familiar with. The searches were based on the *-ing* affix in the Sketch Engine text analysis software in both possible transliterations, i.e., either as *-ιγκ* /ɪg/ or as *-ινγκ* /ɪŋ/. Again, proper nouns as well as words that are not participle or gerund forms were located and excluded from the results.

3.3 Findings and discussion

The analysis of the selected dictionaries revealed varied approaches to the inclusion and treatment of *-ing* nouns. All the items included in the five dictionaries, along with their variant forms, are listed in the Appendix. From a macrostructural perspective, LKN contains 20 lexical items, HLNG includes 51, MILNEG 60, *Wiktionary* 33, and *Slang.gr* 28. Table 1 provides an overview of these contents, showing the total number of lexical items ending in *-ing* included in each dictionary, and highlights which of these are Anglicisms and which are pseudo-Anglicisms.

While some dictionaries are more inclusive, others capture only a small subset of this type of vocabulary, such as LKN. This can be attributed to the fact that LKN was originally a paper dictionary, which faced severe space limitations. In contrast, ongoing projects tend to be more comprehensive in documenting these items — whether through professional lexicography (e.g., MILNEG) or voluntary crowdsourcing (e.g., *Wiktionary* and *Slang.gr*). The HLNG emphasises the incorporation of loanwords and other neologisms in the dictionary's front matter.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

As expected, the majority of *-ing* entries across most dictionaries are Anglicisms, while pseudo-Anglicisms are less common, though their presence is significant in *Slang.gr*, where they dominate. Indeed, it is not unusual for a slang dictionary to include lemmas not typically found in general language dictionaries, often with a tendency toward more humorous and imaginative constructions, as pseudo-Anglicisms often are. As Furiassi and Gottlieb (2015: 24) mention, based on an earlier study by Gottlieb (2009), "jocular derivation" is a distinct type of pseudo-Anglicisms.

Table 1: Findings in the dictionaries' macrostructure

Title/Findings	LKN	HLNG	MILNEG	*Wiktionary*	*Slang.gr*
Total of *-ing* items	20	51	60	33	28
Anglicisms	19	50	57	30	7
Pseudo-Anglicisms	1	1	3	3	21

The *-ing* entries vary across dictionaries in terms of their orthographic representation. While many words conform to the simplification rules of the official Greek grammar, numerous alternative spellings influenced by the original English lexical items are also noted. Therefore, the suffix was either transliterated as *-ιγκ* /ɪg/ (62 instances overall) or as *-ινγκ* /ɪŋ/ (190 instances), depending on the editorial policy of each dictionary. Orthographic variants were added sporadically rather than exhaustively. For example, windsurfing only has one variant in each of the four dictionaries in which it appears as an entry (either *γουίντ σέρφιγκ* or *γουίντ σέρφινγκ*), whereas parking has multiple variants across different dictionaries (these being *πάρκιγκ*, *πάρκιν*, or *πάρκινγκ*). It could be argued that the existence of multiple spellings of an entry may be attributed to the high frequency of this lemma, although this assumption is not supported by the available literature. Interestingly, none of the lexical items ending in *-ing* has entry status in all five dictionaries.

Table 2: The ten most frequent orthographic alternatives in the corpus

English equivalent	-ινγκ /ɪŋ/	Frequency	-ιγκ /ɪg/	Frequency
marketing	μάρκετινγκ	34 853	μάρκετιγκ	508
parking	πάρκινγκ	20 416	πάρκιγκ	773
camping	κάμπινγκ	15 091	κάμπιγκ	588
doping	ντόπινγκ	5 686	ντόπιγκ	132

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

lifting	λίφτινγκ	2 777	λίφτιγκ	89
meeting	μίτινγκ	2 143	μίτιγκ	36
modeling	μόντελινγκ	2 125	μόντελιγκ	96
casting	κάστινγκ	1 346	κάστιγκ	17
jogging	τζόκινγκ	1 303	τζόκιγκ	32
scouting	σκάουτινγκ	1 262	σκάουτιγκ	35

The process of cross-checking the findings against the Greek Web 2019 (elTenTen19) corpus produced some revealing results. The overwhelming majority of instances ending in *-ing* are spelled according to their English equivalents despite the orthographic rules of the official grammar schoolbooks. Although not all 975 results produced by searching the corpus can be listed here, a small sample containing the ten most frequent items is shown in Table 2. Although some of the dictionaries, most notably LKN,[7] adhere to these rules as part of their editorial policy and exclusively transliterate the *-ing* suffix as *-ιγκ* /ɪg/ rather than *-ινγκ* /ɪŋ/, corpus findings prove that most Greek speakers are more familiar with the latter form, which coincides with the English spelling and pronunciation. This is indeed attested in the majority of the entries across the other four dictionaries, which seem to have followed a more descriptive approach in dealing with orthography.

As mentioned, the vast majority of pseudo-Anglicisms appear in *Slang.gr*, following this type of dictionary's tendency toward more jocular content. The analysis indicates that military life, particularly maintenance duties, is one of the most frequent domains for pseudo-Anglicisms in the dictionaries. Such entries include *τσάπινγκ* "the act of removing weeds with a hoe" (derived from *τσάπα* 'hoe') and *γόπινγκ* "the act of removing discarded cigarette butts to clean up an area" (from *γόπα* 'cigarette butt'). This trend highlights a particular sociolinguistic environment, as it is very common for a new army recruit to be assigned tedious chores like cleaning up military areas by removing unwanted items. Another prominent domain for pseudo-Anglicisms is sexual relationships, with entries such as *γύπινγκ* and *μπήχτινγκ* both denoting "predatory behaviour by a man".

4. Conclusion

This study has examined the treatment of *-ing* formations in five Greek dictionaries, revealing significant variations in their documentation and representation of Anglicisms and pseudo-Anglicisms. The findings highlight several key patterns in Modern Greek lexicography. Firstly, there is considerable variation

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

in the inclusiveness of dictionaries regarding *-ing* entries, ranging from LKN's conservative approach (20 entries) to MILNEG's more comprehensive coverage (60 entries). Secondly, while Anglicisms dominate in most dictionaries, *Slang.gr* stands out for its high proportion of pseudo-Anglicisms, reflecting its specialised role in documenting more informal and creative language use.

The orthographic representation of these terms presents a notable disconnect between prescriptive rules and actual usage. While official grammar guidelines recommend simplifying the *-ing* suffix to -ιγκ /ιg/, corpus evidence reveals that Greek speakers predominantly prefer -ινγκ /iŋ/, maintaining closer alignment with English spelling and pronunciation. This preference is reflected in most of the examined dictionaries, except LKN, which strictly adheres to official guidelines.

The analysis of pseudo-Anglicisms reveals particular sociolinguistic patterns, especially in military contexts and informal registers. These formations demonstrate the creative capacity of Greek speakers to adapt English morphological patterns for their own expressive needs. This study contributes to the understanding of how Modern Greek lexicography approaches the documentation of Anglicisms and pseudo-Anglicisms, while also highlighting the dynamic relationship between prescriptive norms and actual language use in the context of English influence on Modern Greek.

Endnotes

1. https://christikolexiko.academyofathens.gr/
2. https://www.greek-language.gr/greekLang/modern_greek/tools/lexica/
3. http://lexicon-neohel.patakis.gr/el/lexikon
4. https://www.slang.gr/
5. https://en.wiktionary.org/wiki/Wiktionary:Main_Page
6. elTenTen19, offered by Sketch Engine (https://www.sketchengine.eu/), contains 2,782,299,354 tokens downloaded in December 2019, and January 2020.
7. It is explicitly stated in the Introduction to the dictionary (p. ιγ) that the official rules have been followed.

References

Dictionaries

HLNG. *Practical Dictionary of Modern Greek.* 2014.
https://christikolexiko.academyofathens.gr/ [20 April 2024].

LKN. *Modern Greek Dictionary.* 1998.
https://www.greek-language.gr/greekLang/modern_greek/tools/lexica/ [20 April 2024].

MILNEG. *Big Electronic Dictionary of Modern Greek.*
http://lexicon-neohel.patakis.gr/el/lexikon [20 April 2024].

Slang.gr. https://www.slang.gr/ [20 April 2024].

Wiktionary. https://en.wiktionary.org/wiki/Wiktionary:Main_Page [20 April 2024].

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

Other literature

AFP in Paris. 2022. Académie Française Denounces Rise of English Words in Public Life. *The Guardian*, 16 February 2022.

Anastassiades-Symeonides, A. 2001. Ιδεολογήματα και δανεισμός [Ideology and Borrowing]. Haris, Y.I. (Ed.). 2001. *Δέκα μύθοι για την ελληνική γλώσσα* [Ten Myths about the Greek Language]: 63-71. Athens: Patakis.

Carbonaro, G. 2023. Ciao, Hello, No! Italy's Right-wing Government Wants to Ban English Words with €100,000 fines. *Euronews*, 3 April 2023.

elTenTen19. 2020. *Corpus of the Greek Web.* https://app.sketchengine.eu/ [20 April 2024].

Furiassi, C. and H. Gottlieb (Eds.). 2015. *Pseudo-English. Studies on False Anglicisms in Europe*. Berlin: De Gruyter Mouton.

Harris, J. 2010. *The Frequency and Distribution of Written and Spoken Anglicisms in Two Varieties of French*. Unpublished MA Thesis. Montreal, Quebec: Concordia University.

Hellenic Statistical Authority. 2022. Adult Education Survey. https://www.statistics.gr/en/statistics/-/publication/SJO18/ [1 December 2024].

Michalopoulou, E. 2022. *Κατασκευή ηλεκτρονικού σωματοκειμενικού λεξικού νεολογισμών των κρίσεων (οικονομικών, κοινωνικοπολιτικών, υγειονομικών) της περιόδου 2015–2020* [Construction of an Electronic Corpus-based Dictionary of Neologisms of the Crises (Economic, Socio-political, Healthcare) of the Period 2015–2020]. Unpublished MA Thesis. Kalamata: University of the Peloponnese, School of Humanities and Cultural Studies.

Prior, J. 2024. English Language Statistics: How Many People Learn English? *DoTEFL*, 9 Dec. 2024. https://www.dotefl.com/english-language-statistics/ [8 December 2024].

Reuters. 2023. Russia Bans Government Officials from Using Most Foreign Words. Reuters, 28 Feb. 2023. https://www.reuters.com/world/europe/russia-bans-government-officials-using-most-foreign-words-2023-02-28/ [20 April 2024].

Sánchez Fajardo, J.A. 2018. Towards a Typological Account of Pseudo-Anglicisms. *Pragmalinguistica* 26: 330-348.

Stamenov, C. 2015. Borrowing Word-Formation: *-ing* Suffixation and Blending in Bulgarian. *Sŭpostavitelno Ezikoznanie / Contrastive Linguistics* 40(3): 163-197.

Statista. 2023. The Most Spoken Languages Worldwide in 2023 (by Speakers in Millions). https://www.researchgate.net/figure/Figure-N01-The-most-spoken-languages-worldwide-2023-Statista_fig1_388066718 [1 Dec. 2024].

Storjohann, P. 2024. Neologisms and their Functions in Critical Discourse. *Book of Abstracts of 6th Globalex Workshop on Lexicography and Neology, GWLN-6, Pretoria, South Africa, July 3, 2024.*

Triandaphyllidis, M. et al. 1941. *Νεοελληνική γραμματική (της δημοτικής).* [Modern Greek Grammar (of Demotic)]. Athens: ΟΕΣΒ [in Greek].

Xydopoulos, G.J. 2019. Εξαγγλισμός και νέα ελληνική: η εισαγωγή νέων λεξικών σχημάτων με μεταφραστικό δανεισμό. [Anglicisation in Modern Greek: The Import of New Lexical Patterns through Calquing.] Archakis, A., N. Koutsoukos, D. Papazachariou and G.J. Xydopoulos (Eds.). 2019. *Language Variety: Festschrift for Angela Ralli*: 423-435. Athens: Kapa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

Appendix

The following table includes all the *-ing* entries that were found in the five dictionaries under analysis, except for entries that consist of proper nouns without a metaphorical meaning or that do not fall under the present participle or gerund forms (such as *ring* or *string*). Items in *italics* are pseudo-Anglicisms, while those in **bold** have alternative (jocular) senses to the ones of their English equivalents.

LKN	**HLNG**	**MILNEG**	***Wiktionary***	***Slang.gr***	**Meaning**
		αντιγόπινγκ		*αντιγόπινγκ κοντρόλ*	removing discarded cigarette butts from an area
	αντιντάμπινγκ	αντιντάμπιγκ αντιντάμπινγκ			anti-dumping
	αντιντόπινγκ	αντιντόπιγκ αντιντόπινγκ			antidoping
				άουτινγκ	outing
	γιότινγκ	γιότινγκ	γιότινγκ γιώτινγκ		yachting
			γκλάμπινγκ		glamping
				γλείφτινγκ	coaxing
			γόπινγκ	*γόπινγκ*	removing discarded cigarette butts from an area
γουίντ σέρφιγκ	γουίντ σέρφινγκ	γουίντ σέρφινγκ	γουίντ σέρφινγκ		windsurfing
				γύπινγκ	predatory behavior by a man
				επιτάφιινγκ	moving from one Epitaphios liturgy to another
				έρπινγκ	crawling
ζάπιγκ	ζάπιγκ ζάπινγκ	ζάπιγκ ζάπινγκ	ζάπινγκ		channel zapping
				κάλινγκ	collecting empty bullet cases after shooting practice
κάμπιγκ	κάμπιγκ κάμπινγκ	κάμπιγκ κάμπινγκ			camping
		κάρτινγκ			karting
	κάστινγκ	κάστινγκ	κάστινγκ		casting
		κέρλινγκ	κέρλινγκ		curling
κέτεριγκ	κέτερινγκ	κέτεριγκ κέτερινγκ	κέιτερινγκ		catering

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

	κικ μπόξινγκ	κικ μπόξινγκ			kick boxing
	κλάμπινγκ	κλάμπιγκ κλάμπινγκ			clubbing
κλίριγκ	κλίρινγκ	κλήριγκ κλίριγκ κλίρινγκ	κλήρινγκ κλίρινγκ		clearing
	κοκούνινγκ	κοκούνιγκ κοκούνινγκ			cocooning
	κουίσλινγκ κουίσλιγκ κουίσλινγκ	Κουίσλιγκ Κουίσλινγκ			Quisling
	λήζινγκ λίζινγκ	λίζιγκ λίζινγκ	λήζινγκ λίζινγκ		leasing
λίφτιγκ	*λίφτιγκ* *λίφτινγκ*	*λίφτιγκ* *λίφτινγκ*	*λίφτινγκ*		face-lift
				λούφινγκ	loafing
μάρκετιγκ	μάρκετιγκ μάρκετινγκ	μάρκετιγκ μάρκετινγκ	μάρκετινγκ		marketing
μίτιγκ	μίτινγκ	μίτιγκ μίτινγκ	μήτινγκ μίτινγκ		meeting
μόντελιγκ	μόντελινγκ	μόντελιγκ μόντελινγκ			modelling
				μουνγουόκινγκ	moon-walking
				μπήχτινγκ	predatory behavior by a man
	μπλόγκιν μπλόγκινγκ	μπλόγκιν μπλόγκινγκ			blogging
μπόντι μπίλντιγκ	μπόντι μπίλντινγκ	μπόντι μπίλντιγκ μπόντι μπίλντινγκ			body-building
μπόουλιγκ		μπόουλιγκ μπόουλινγκ		**μπόουλινγκ**	bowling (also bullying)
	μπούλινγκ	μπούλινγκ	μπούλινγκ		bullying
Μπράουνιγκ		Μπράουνιγκ			Browning
			μπρέικ ντάνσινγκ		break-dancing
		μπρεϊνστόρμιγκ			brain-storming
	μπρίφινγκ	μπρίφιν μπρίφινγκ			briefing
				μπύρινγκ	beer drinking
ντάμπιγκ		ντάμπινγκ	ντάμπινγκ		dumping (also dubbing)
	ντάνσινγκ		ντάνσινγκ		dancing
ντόπιγκ	ντόπιγκ ντόπινγκ	ντόπινγκ	ντόπινγκ		doping
	ντρέσιγκ ντρέσινγκ	ντρέσινγκ	ντρέσινγκ		dressing
πάρκιγκ πάρκιν	πάρκιγκ πάρκιν πάρκινγκ	πάρκιγκ πάρκιν πάρκινγκ	πάρκινγκ		parking

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

				πευκοβελόνινγκ	removing fallen pine needles
				πέφτινγκ	flirting
πίλιγκ	πίλιγκ πίλινγκ	πίλινγκ			peeling
	πίρσινγκ	πίρσινγκ			piercing
				πούτσινγκ	lazing about
	πρέσιγκ πρέσινγκ	πρέσιγκ πρέσινγκ	πρέσινγκ		pressing
				πύλινγκ	guard duty at the gate
	ράνκινγκ				ranking
	ράφτινγκ	ράφτινγκ	ράφτινγκ	**ράφτινγκ**	rafting (also sewing)
	ρόμινγκ ρόουμινγκ				roaming
				σάπινγκ	brain rotting
	σέξτινγκ		σέξτινγκ		sexting
σέρφιγκ	σέρφινγκ	σέρφινγκ	σέρφινγκ		surfing
		σκαϊντάιβινγκ			skydiving
	σκάουτινγκ	σκάουτινγκ			scouting
				σκόπινγκ	guard duty
	σόπινγκ	σόπινγκ			shopping
		σορτ σέλινγκ			short-selling
				σούρσινγκ	lazing around
	σπάμινγκ	σπάμινγκ			spamming
			σπάρινγκ		sparring
		σπόνσορινγκ			sponsoring
	στάιλινγκ	στάιλινγκ			styling
	στρέτσινγκ	στρέτσινγκ			stretching
			στρίκινγκ		streaking
		στρίμινγκ			streaming
	τάιμινγκ	τάιμιγκ τάιμινγκ			timing
τάκλιν	τάκλιν τάκλινγκ	τάκλιν τάκλινγκ			tackling
		τελεμάρκετινγκ τελεμάρκετιγκ τηλεμάρκετινγκ τηλεμάρκετιγκ	τηλεμάρκετινγκ		telemarketing
τζόκιγκ	τζόκιγκ τζόκιν τζόκινγκ	τζόγκιν τζόγκινγκ τζόκιν τζόκινγκ	τζόγκινγκ		jogging
	τράβελινγκ	τράβελινγκ	τράβελινγκ		travelling
	τράφικινγκ	τράφικιγκ τράφικινγκ	τράφικινγκ		trafficking
	τρέκινγκ	τρέκινγκ			trekking
		τσάπιγκ *τσάπινγκ*	*τσάπινγκ*	*τσάπινγκ*	removing weeds with a hoe
	τσάτινγκ	τσάτινγκ			chatting
	τσιρλίντινγκ	τσιρλίντινγκ			cheerleading
	φάκτορινγκ				factoring

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2043 (Lexicofocus)

				φάνγκερλινγκ	fangirling
			φέρινγκ		fairing
				φήλινγκ	feeling
				φίδινγκ	lazing around
				φίστινγκ	fisting
	φραντσάιζινγκ	φραντσάιζιγκ φραντσάιζινγκ			franchising
				φύλλινγκ	removing fallen leaves
	χάκινγκ	χάκιγκ			hacking
		χάντλινγκ			handling
χάπενιγκ	χάπενινγκ	χάπενινγκ	χάπενινγκ		happening
				ψώλινγκ	brain rotting

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

Semi-Automatic Detection of New Words in Modern Georgian*

Tamar Laluashvili, *School of Arts and Sciences, Ilia State University, Tbilisi, Georgia*
(*tamar.laluashvili.1@iliauni.edu.ge*)
(*https://orcid.org/0009-0008-3978-2333*)
and
Tinatin Margalitadze, *Centre for Lexicography and Language Technologies, Ilia State University, Tbilisi, Georgia*
(*tinatin.margalitadze@iliauni.edu.ge*)
(*https://orcid.org/0000-0001-9485-1698*)

Abstract: The study of neologisms in the Georgian language has gained significance due to the rapid socio-political changes in the country after the collapse of the Soviet Union and the country regaining independence. Technological advancements of the 21st century have also played a role. These developments have led to the introduction of numerous new terms and concepts into the language. However, there has been no established methodology for identifying neologisms in modern Georgian. To address this issue, a methodology was worked out at Ilia State University based on the study of existing methods applied to other languages. A corpus of the Georgian language was developed from textual materials retrieved from online platforms such as online newspapers and magazines, online media websites, websites of non-governmental organisations, and governmental agencies. Two lemmatisation tools were then applied to it to identify potential neologisms. This paper presents the methodology for the semi-automatic detection of new words in modern Georgian.

Keywords: NEOLOGISM, GEORGIAN LANGUAGE CORPUS, LEMMATISER, OUT-OF-VOCABULARY LEXIS, NEOLOGISM DETECTION METHODOLOGY

Opsomming: Die semi-outomatiese opsporing van nuwe woorde in moderne Georgies. Die studie van neologismes in die Georgiese taal het belangwekkend geraak as gevolg van die vinnige sosio-politieke veranderinge in die land nadat die Sowjetunie ineengestort het en die land onafhanklikheid herwin het. Die tegnologiese vooruitgang van die 21ste eeu het ook 'n rol gespeel. Hierdie ontwikkelings het gelei tot die ontstaan van talle nuwe terme en konsepte in die taal. Tot dusver was daar egter geen gevestigde metodologie om neologismes in moderne Georgies te identifiseer nie. Om hierdie kwessie te ondersoek, is 'n metodologie aan die Staatsuniversiteit van Ilia ontwikkel, wat gebaseer is op die bestudering van bestaande metodes wat op ander tale toegepas word. 'n Korpus van die Georgiese taal is uit tekstuele materiaal wat van aanlyn platforms soos

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

koerante, tydskrifte, aanlynmediawebtuistes, webtuistes van nieregeringsorganisasies, en webtuistes van regeringsagentskappe verkry is, ontwikkel. Daarna is twee lemmatiseringshulpmiddels daarop toegepas om potensiële neologismes te identifiseer. In hierdie artikel word die metodologie vir die semi-outomatiese opsporing van nuwe woorde in moderne Georgies bespreek.

Sleutelwoorde: NEOLOGISME, GEORGIESE TAALKORPUS, LEMMATISEERDER, BUITEWOORDESKATLEKSIS, NEOLOGISMEOPSPORINGSMETODOLOGIE

1. Introduction: Historical background

The present study is part of the three-year project, dedicated to the comprehensive research of neologisms in the modern Georgian language. The project is supported by the Shota Rustaveli National Science Foundation of Georgia (FR-23-4304) and its main goals are to develop a methodology for the semi-automatic detection of neologisms in modern Georgian, to create and publish an online dictionary of neologisms, and to set up a special website for the monitoring of neologisms in Georgian in the future. In this paper the methodology developed at Ilia State University for the semi-automatic detection of new words in modern Georgian is presented.

Changes taking place in a language are usually very slow and difficult to notice, and they occur over long periods of time before becoming perceptible on a synchronic level. But there are exceptions from this general tendency and contemporary Georgian is a good example of this. Currently, Georgia and the Georgian language are in an extremely interesting era from a historical point of view. After the collapse of the former Soviet Union, the country saw the emergence and rapid development of a free market economy, multiparty political system, private banking sector, the national armed forces, and many other such structures, which neither existed nor were even imaginable under the Soviet empire (Margalitadze 2020).

In 2005, Georgia joined the Bologna Process, and in 2014, the country signed an Association Agreement with the European Union. All this brought about intense relations with foreign countries on diplomatic or political, as well as on educational, cultural and economic levels, and led to the introduction of concepts, words and terms reflecting the new realities of life into the Georgian language. These include *შეზღუდული პასუხისმგებლობის კომპანია / shezghuduli p'asukhismgeblobis k'omp'ania* ('Limited Liability Company'), *ინდმეწარმე / indmets'arme* ('individual entrepreneur'), *ასპირანტი ქვეყნები / asp'iranti kveq'nebi* ('aspirant countries'), *ევრო-ატლანტიკური მისწრაფება / evroatlantik'uri mists'rapeba* ('Euro-Atlantic aspiration'), *კოლონელი / k'oloneli* ('colonel'), *ხარისხის უზრუნველყოფა / khariskhis uzrunvelq'opa* ('quality assurance'), and *სილაბუსი / silabusi* ('syllabus').

Covid-19 and the Russian-Ukrainian war also gave rise to new words in Georgian, for example *რაშიზმი / rashizmi* ('Rushism', formed by blending Russia + Fascism), *რაშისტი / rashisti* ('Rushist', a blend of Russia + Fascist), *ორკი /*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

orki ('Orc'),[1] *დრონი / droni* ('drone'), and *ატაკამსი / atak'amsi* ('ATACMS'). The recent developments in Georgia, the enactment of the so-called "Russian Law", massive demonstrations of the local population, violent attacks of the police special forces on the protest rallies, and the country's democratic backsliding "enriched" the vocabulary of Georgian with some neologisms of contemptuous meanings, for example *რობოკოპი / robok'op'i* ('Robocop'),[2] *ტიტუშკა / titushk'a* ('titushka'),[3] *ზონდერი / zonderi*,[4] and *ზონდერმოსამართლე / zondermosamartle* ('a corrupted, bought off judge').

These processes were further intensified by the increasing abundance of computer, telecommunications and mobile technologies. Consequently, all prerequisites were in place, which could cause substantial changes in the lexis and even morphology. This makes the detection and analysis of latent diachronic processes occurring in the Georgian language especially interesting at this linguistic-historical moment.

The study of neologisms is particularly relevant for lexicography, which, in addition to studying the theoretical aspects of the issue, also serves purely practical purposes. It involves updating existing dictionaries, adding new words, and assigning new meanings to existing ones. Contemporary users evaluate the quality of dictionaries by their ability to keep pace with the latest vocabulary and meanings. Dictionaries that fail to capture modern vocabulary tend to lose their appeal and popularity. Despite this, the present study reveals that many words that have entered the vocabulary of the Georgian language in the 21st century are not attested in either the explanatory or orthographic dictionaries of the language, nor in recently published Georgian bilingual dictionaries. For example, words connected to online media, such as *ონლაინმედია / onlainmedia* ('online media'), *ონლაინგაზეთი / onlaingazeti* ('online newspaper'), and *ონლაინგამოცემა / onlaingamotsema* ('online publication'), are not included even in the *Orthographic-Stylistic Dictionary of a Journalist* published in 2010 by the Institute of Linguistics of Georgia.

Additionally, increased interest in neologism research on an international level is reflected in the publication of studies examining neologisms across various languages and the development of a more systematic approach to the subject (cf. Trap-Jensen 2020; Klosa-Kückelhaus and Kernerman 2023). Conferences dedicated solely to neologisms, along with the establishment of an ongoing conference series such as GLOBALEX, also demonstrate the growing interest and commitment to this field (cf. Klosa-Kückelhaus and Kernerman 2022). Furthermore, the European Network on Lexical Innovation project, funded by the European Cooperation in Science and Technology (COST), serves as a valuable initiative that connects multiple universities around the world, including Ilia State University in Georgia. This further exemplifies the importance and relevance of this study and the project from which it stems.

2. An overview of neologisms

Neologisms refer to new words or meanings that appear in a language. The term

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

neologism entered the English language in 1772 from French and means 'new word or expression', derived from Greek νέος (neos) 'new' + λόγος (logos) 'word'. In many cases, the reason for their emergence is new things, phenomena, or concepts that a language community learns and needs to name. They therefore enrich and add dynamism to language. A language has different registers, such as literary language, scientific language, colloquial language, slang, and dialectal language. Neologisms may appear in any register of a language, both written and oral. Georgian is no exception, and the appearance of neologisms can be observed in all registers of modern Georgian (for examples, see Section 5).

Neologisms arise in many ways. New words may be coined in a language or borrowed from other languages. Coinage of new words is interesting as their study reveals the lexical creativity of a language at a certain stage of its development. While coinage of completely new words is rare in Georgian, some cases can still be found, for example the Georgian equivalent of the English adjective *vulnerable* is a neologism *მოწყვლადი / mots'q'vladi.* It has developed from an old Georgian verb *წყვლა / ts'q'vla* ('to slay') by adding the suffix *-ადი / -adi*. Borrowing may reflect something new for speakers of the borrowing language. This is the case with the examples given in Section 1. However, borrowings may also be introduced under the influence of the prestige of another language (Trask 1996: 19). In this case, synonyms in a language consisting of foreign and native word pairs arise, such as *კონტრიბუცია / k'ontributsia* ('contribution') and *წვლილი / ts'vlili* ('contribution'); *კოლაბორაცია / k'olaboratsia* ('collaboration') and *თანამშრომლობა / t'anamshromloba* ('collaboration'), and *სტაფი / stapi* ('staff') and *თანამშრომლები / tanamshromlebi* ('staff').

Another type of neologisms is the development of new senses of existing words. Sense development is often the result of semantic borrowing and is influenced by a foreign tongue and polysemy of an equivalent word in a foreign language. There are many examples of sense development in modern Georgian due to the influence of English, with Georgian *ქიმია / kimia* ('chemistry') being a typical example. Originally, *ქიმია / kimia* ('chemistry') in Georgian meant only 'the scientific study of the structure of substances …'. The influence of the English *chemistry* and its polysemous meaning 'the relationship between two people, usually a strong sexual attraction' (Oxford Learner's Dictionary 2024) caused Georgian *ქიმია / kimia* to be used in this sense too.[5]

There are also morphological neologisms, connected to the emergence of new grammatical constructions. This is despite Georgian having a complex morphology and verbs being particularly rich in grammatical categories (see a more detailed discussion about lemmatisation issues of Georgian verbs in Section 4). Nevertheless, the development of some analytical constructions in modern Georgian verbs can be observed, for example განცხადება გაკეთდა / *gantskhadeba gak'et'da* ('a statement was made'). Such cases can be regarded as morphological neologisms.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

3. Methods and tools developed for semi-automatic detection of neologisms

With the advancement of corpus linguistics, new opportunities have emerged for studying various aspects of language. Corpora have significantly contributed to the investigation of neologisms as well. Before the advent of corpus linguistics, researchers had to analyse large volumes of text manually to assess word novelty. Apart from being time-consuming, this method was subjective as it relies on the judgment of individual researchers. The advancements in modern technology have, however, expanded research possibilities. Modern electronic corpora enable the rapid and efficient processing of vast amounts of text, facilitating the study and evaluation of different linguistic phenomena, including neologisms. Although this process is largely automatic or semi-automatic, human intervention remains essential for the analysis of the material (Grochocka 2011: 62).

Different methods for the detection of neologisms have been developed in different languages, of which the following methods are commonly used: research based on the exclusion principle, application of lexical and punctuation discriminants, and statistical analysis (Janssen 2009: 69).

The exclusion principle is a well-established method for identifying neologisms, comprising two important components: the building of a study corpus for the extraction of potential neologism candidates and the creation of an exclusion list that is based on the macrostructures of a given language's dictionaries or reference corpora. Each word within the study corpus is systematically compared to the exclusion list, and words absent from the list are considered neologism candidates (Janssen 2009: 69-70; Grochocka 2011: 63). Some neologism detection tools that have been developed based on the exclusion principle are *Wortwarte* (Lemnitzer 2000) for German, *BuscaNeo* in Observatori de Neologia (OBNEO) (Cabré Castellví and Estopà Bagot 2009) for Spanish and Catalan, and *Logoscope* (Falk, Bernhard and Gérard 2014) for French.

This approach proves effective in identifying formal (or orthographic) neologisms and lexical borrowings. However, it has certain limitations, such as being incapable of detecting semantic neologisms. Moreover, the mere absence of a word from a dictionary does not necessarily indicate that it is a neologism. Online texts often contain typographical errors that should not be considered neologisms, and proper names are also excluded from the candidate list. Furthermore, there are established words that may not be documented in dictionaries but do not qualify as neologisms. For example, a well-established Georgian word *ბრძენკაცი / brdzenk'atsi* ('a wise man') is not documented in the *Explanatory Dictionary of the Georgian Language* (EDGL).

The second commonly-used method is applying lexical and punctuation discriminants for neology detection. Lexical discriminants are phrases that precede a neologism candidate and emphasise the word's novelty or unfamiliarity. Such phrases include *termed*, *so-called*, *known as*, and *defined as*. Punctuation marks such as quotation marks (single and double), italics, or parentheses can also serve as

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

discriminants (Paryzek 2008: 165). This method suggests that the lexical unit is unfamiliar to the reader, and it is therefore presented in a different format. However, it is important to note that not all words formatted in this manner are neologisms; non-neologisms and other types of textual noise may also be marked this way (Janssen 2009: 70).

One more method for identifying neologisms is based on statistical analysis, which usually involves counting words in the study corpus and comparing the count to that of a reference corpus. Janssen (2009) identifies four approaches, of which the first relies on hapax legomena where words that occur only once in the study corpus are considered neologism candidates. This assumption formed the basis of the tool *NeoloSearch*, developed by Janicijevic and Walker (1997), which automates the process of the retrieval and analysis of neologisms. According to the second approach, a neologism is a word of the study corpus that has zero frequency in the reference corpus. The third approach is to compare the frequency of words in two corpora. If the frequency of a word in the study corpus is higher than in the reference corpus, there is a good chance that the word is a neologism, and an increase in the frequency of a word in the study corpus indicates that it may be a semantic neologism. The fourth approach counts the frequency of a word in contexts, treating a change in context as an indication that the meaning or usage of the word has changed.

Studies (for example, Cook and Stevenson 2010) show that statistical analysis can detect such complex linguistic processes as semantic change. The meanings of words change in different ways, new meanings appear, new connotative meanings develop, and a metaphorical transfer, amelioration or pejoration occur. Even though modern technologies make it easier to detect such processes, it is still semi-automatic and it is inevitable for professional linguists to intervene and exclude false candidates and validate true neologisms.

For a successful application of a combination of neologism detection methods see *Ordtrawler* ('Word Trawler'), an automatic neologism detection prototype developed by Halskov and Jarvad (2010) at the Danish Language Council. The study by Halskov and Jarvad (2010) indicates that a combination of these techniques achieves the highest level of precision, approximately 40%.

4. Methodology for the semi-automatic detection of new words in modern Georgian

4.1 Studies on neologisms in Georgia

Georgian scholars have been interested in the issue of neology and have studied it from different perspectives (cf. Beliashvili 2015; Goshkheteliani and Kikvadze 2017; Mtchedlishvili 2019; Margalitadze 2020; Rayfield 2023). While these studies are not corpus-based, they mostly investigate the topic from a lexicological point of view. Research on neology is primarily preoccupied with borrowings, especially anglicisms, and the influence of English on Georgian.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

Kirvalidze (2017) and Davitishvili (2018) study types of and reasons for borrowing in Georgian, identifying direct borrowings, for example *ლეპტოპი / lep'top'i* ('laptop'), translation loans such as *სწრაფი კვება / sts'rapi k'veba* ('fast food'), and semantic borrowings, like *მეხსიერება / mekhsiereba* ('memory' in information technology). Some authors also analyse political neologisms in English and methods of their adoption into Georgian (cf. Mtchedlishvili 2019).

The findings point to different reasons for borrowing English words. According to Goshkheteliani and Kikvadze (2017), the reason in the majority of cases is a lack of a Georgian equivalent (e.g., laptop). Preference might also be given to a borrowed word instead of a descriptive Georgian equivalent, such as *ვერშოფი / verkshopi* ('workshop') over the descriptive equivalent *სამუშაო მეხვედრა / samushao shekhvedra* ('working meeting') (Goshkheteliani and Kikvadze 2017). Rayfield (2023) identifies the domains in which many English words are borrowed as being key fields developed in the 21st century, namely human resources, public relations, information technologies, and social media. He also highlights inconsistencies in the spelling of borrowings, noticing that one and the same word often has several spellings in Georgian, e.g. *ჩათი, ჩატი, ჩეთი, ჩეტი / chat'i, chati, chet'i, cheti* ('chat' in social media).

4.2 Methodology selected for the Georgian language

Despite this interest in neologisms, there was no established methodology for identifying new words in modern Georgian, which determined the decision to work out such a methodology in this project. For this purpose, existing methods of detecting new words in other languages (cf. Lemnitzer 2000; Cabré Castellví and Estopà Bagot 2009; Falk et al. 2018) were studied, after which the current approach was formulated. The present study is focused on formal (or orthographic) neologisms, as the methodology discussed below helps identify them. The detection of the development of new word senses requires different approaches, which is not addressed in this research.

For this study, the exclusion method was employed. At first, a study corpus was composed and two lemmatisers were applied to it. The first served as an exclusion source and the second was used for reducing the number of neologism candidates. This is discussed in more detail below. The corpus includes textual material from Georgian online newspapers and magazines, news and media sites, and websites of governmental and non-governmental organisations (NGOs), amounting to forty-three sites covering the last 20 years. Thirteen of these are of online magazines and newspapers, eleven of news and online media sites, nine of NGOs, and the rest of different governmental agencies. Textual material from the websites of governmental agencies was retrieved from the English–Georgian Parallel Corpus (2024).[6] The final study corpus is available online (https://neologism.iliauni.edu.ge/) and contains over 100 million tokens.

At the initial stage of the research, Georgian Wikipedia was considered for

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

inclusion in the study corpus. Wikipedia is interesting from several points of view. It has many authors and features articles on diverse topics such as music, cinema, history, art, politics (Barbaresi 2015: 72). Wikipedia's sentences are much better structured and grammatically correct when compared to many other websites that contain grammatically incorrect texts (Yano and Kang 2008: 2). On the other hand, Wikipedia articles include many proper names, geographical names, and terms from different domains, and Georgian Wikipedia contains many typographical errors, which made it difficult to reduce the number of potential neologisms. It was thus excluded from the study.

There were five main stages involved in the data retrieval, including (1) the creation of "web crawlers", (2) retrieving the material from websites, (3) processing the material in LancsBox,[7] (4) cleaning the material from unnecessary data, and (5) uploading the material to the lemmatiser. The "web crawlers" were developed at UniLab, Cyberlaboratory of Ilia State University, and were used for downloading material from the websites.

Websites of different genres were analysed individually to identify the most productive ones concerning the creation of neologisms. The information obtained is important for the next stage of the project, in which a concept will be developed and a website will be set up for monitoring neologisms in Georgian. As mentioned above, after the data collection, unwanted items such as numbers, non-Georgian characters, symbols, and duplicates were removed, texts were tokenised, and the final lists were uploaded into the lemmatiser.

4.3 Issues of lemmatising Georgian words

The Georgian language belongs to the agglutinative language type. Nouns, adjectives, pronouns, and numerals are inflected and there are seven cases. Basic means for the expression of morphological categories in Georgian are prefixes and suffixes. The categorial system of verbs is especially complex. On the one hand, verbs have inflectional categories such as person, number, mood, tense, iteration, sequence of action; on the other hand, there are derivational categories — aspect, voice, version, causative, location, and direction-orientation (Shanidze 1973). Categories have their morphological markers (cf. Margalitadze 2022). For example, in terms of person, the verb can be inflected as — *v-ts'er* ('I write'), *ts'er* ('you write'), or *ts'er*-s ('he/she writes'). For number, there is — *v-ts'er* ('I write') or *v-ts'er-t* ('we write'), and for time there is — *v-ts'er* ('I write') or *da-vt'ser* ('I shall write'). Iteration is expressed through forms such as *ts'er-a* ('he/she wrote') or *ts'er-d-a* ('he/she used to write'), whereas version can be indicated by *ts'er-s* ('he/she writes' — neutral version), *i-ts'ers* ('he/she writes something for oneself' — subjective version), or *u-ts'ers* ('he/she writes something for somebody' — objective version). Location is shown through *ts'ers* ('he/she writes' — neutral) or *a-ts'ers* ('he/she writes something on something' — superessive), and voice by *ts'er-s* ('he/she writes' — active voice) or *i-ts'er-eba* ('smth. is being written' — passive voice). Finally, the causative form is represented by *a-ts'er-*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

ineb-s ('he/she has somebody write something'). The complex morphology of Georgian necessitates the lemmatisation of words in the study of neologisms.

4.4 Semi-automatic detection of neologisms in modern Georgian

As mentioned above, two lemmatisers were used in this research. The first is the lemmatisation tool for the Georgian language developed under the direction of Irina Lobzhanidze, a professor at Ilia State University. This lemmatiser was created based on the Finite-State Lexicon compiler (Lexc), which is a tool for creating lexical transducers. The lexicon of the morphological analyser created with this tool consists of two main components: (a) lemmas, which are presented in the form of lexicons corresponding to specific parts of speech, and (b) additional classes. In this case, a lemma is presented as an unmarked form of a word, that is the root or headword as it appears in printed or electronic dictionaries or word lists. The lemmas are based on the eight-volume EDGL and Melikishvili's Verb Index (2014).

Initially, the lexicon contained 78 000 units for nouns and 85 000 units for verbs. This lexicon has since been expanded with data obtained during the corpus testing process (cf. Lobzhanidze 2021: 64-65). Accordingly, the lemmatisation tool can only process lexical units, which are included in these resources. The lemmatiser is available on the Ilia State University website.[8] As already pointed out, a considerable segment of the vocabulary of contemporary Georgian is not represented in explanatory and orthographic dictionaries and, as a result, the said lemmatiser cannot identify them. This allowed the lemmatiser to be used as an exclusion source. The list of neology candidates was constituted by words not recognised by this tool, which are unlemmatised, out-of-vocabulary (OOV) lexis. After the tokenisation and lemmatisation of the corpus material, the OOV lexical units were subjected to analysis and the potential neologisms were sampled.

As a result of this process, the lemmatiser was able to analyse and lemmatise 278 137 words from the online media corpus data, leaving 167 070 words unlemmatised. From the NGOs' data, 61 777 words were lemmatised and 21 576 were OOV lexis. As for the data collected from websites of governmental agencies, the lemmatisation tool processed 32 821 words, and 5 923 units comprise OOV lexis (see Table 1).

While analysing OOV lexis, it was detected that they contained not only neology candidates but also undocumented lexis. These cannot be considered neologisms but are not documented in dictionaries. For instance, the example mentioned earlier, *ბრძენკაცი / brdzenkatsi* ('a wise man') is an obvious case of an undocumented word. Many undocumented derived words were encountered, such as *ბრძოლისუუნარო / brdzolisuunaro* ('combat-incapable') and *ხელნაკეთობა / khelnak'et'oba* ('a handmade item'). One of the reasons for the existence of undocumented words may be the approach of editorial boards of Georgian monolingual dictionaries, with this approach not being corpus-based. The study of

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

this segment of vocabulary is important but requires a different methodology that could not be included within the framework of this research. To differentiate neology candidates from undocumented vocabulary, reliance had to be placed on intuition, knowledge of Georgian, and experience as lexicographers. The Georgian National Corpus was also consulted to check the dates of some words. Because part of the work was done manually, the methodology is considered semi-automatic and not fully automated.

Table 1: Number of lemmatised and unlemmatised words for each corpus component

	Lemmatised	**OOV lexis**
Online newspapers, magazines, online news sites, online media sites	278 137	167 070
NGOs	61 777	21 576
Governmental agencies	32 821	5 923

Undocumented words and neologisms will be added to the lexicon of the first lemmatiser so that it can recognise and lemmatise these words in the future. This tool will be used in the platform that will be developed for monitoring of neologisms in Georgian.

At the next stage of the study, to reduce the number of OOV lexis, another lemmatiser, developed by Meurer (2014) for the Georgian National Corpus (GNC, Gippert and Tandashvili 2015), was applied. The GNC contains over 200 million tokens and Meurer's lemmatiser can recognise more words than the lemmatiser applied as an exclusion source in the first phase of this research. His lemmatiser is also based on finite-state technology but contains a much larger dictionary and he developed a utility for guessing and lemmatising unknown words. In addition, lists of geographical and proper names were generated in this research, which Meurer added to his lemmatiser. After lemmatising OOV lexis with his tool, the number of unlemmatised vocabulary was reduced by 45%.

One more tool used in this research, was the Georgian word embedding platform (2022)[9] (Figure 1). The word-embedding corpus includes 1,5 billion words, making it one of the largest databases of Georgian. It allows the collection of semantically-related words. When a desired word from any language register is entered into the search bar, the word embedding presents up to thirty different words from the same register, which helps to identify more neologisms. For example, with the embedding of *ონლაინმედია / onlainmedia* ('online media'), more neologisms were extracted, such as *ინტერნეტმედია / internetmedia* ('internet media'), *ინტერნეტპორტალი / internetp'ortali* ('internet portal'), and *ინტერნეტტელევიზია / internettelevizia* ('internet television'). It should be

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

noted that the word-embedding platform is a supplementary part of the methodology, as relying entirely on it would be time-consuming and labour-intensive.

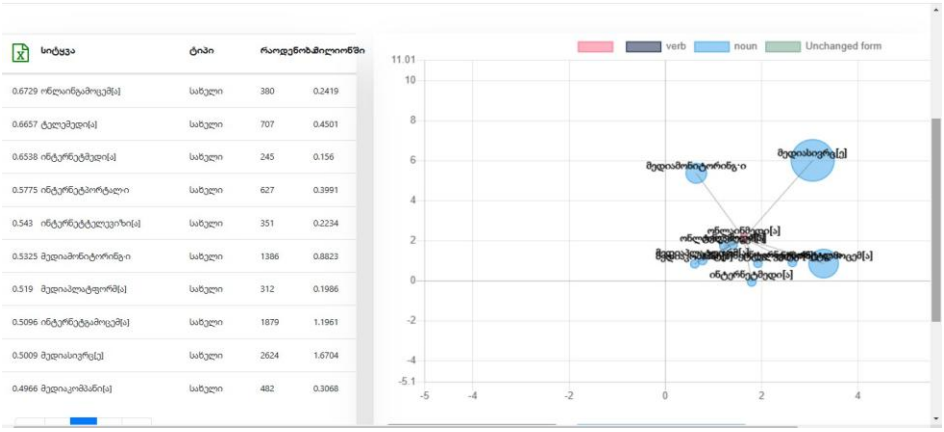


Figure 1: Georgian Word-Embedding Platform

5. Results of the study

As a result of the process described above, over 1 700 lexical neologisms in modern Georgian were identified. Most of the selected words belong to the common vocabulary of Georgian, including colloquial words and slang. Some general terms from the fields of social media, online media, and tourism were also selected. New terms from specialised fields were excluded from the study as they require detailed analysis and assistance from field professionals. The list of neologisms contains the words which entered the Georgian vocabulary during the last 20 years, and the concept of neology was expanded to words introduced in the Georgian language in the 21st century. These words cover the domains of politics, economy and finances, medicine (including esthetic medicine), tourism and hospitality, education, society and social life, agriculture, and more.

The identified neologisms will undergo thorough analysis but the preliminary study has revealed that the majority of new words in modern Georgian are anglicisms, or in other words borrowings from English. There are also hybrid words, with borrowed and Georgian roots, such as *კიბერომი / k'iberomi* ('cyberwar'), *კიბერუსაფრთხოება / k'iberusaprt'khoeba* ('cybersecurity'), *კიბერშეტევა / k'ibersheteva* ('cyber-attack'), *ონლაინგამოცემა / onlaingamotsema* ('online publication'), and *ინტერნეტგვერდი / internetgverdi* ('internet page'). The intensification of some word-forming components in Georgian were also observed, such as *ონლაინ / onlain* ('online'), *ინტერნეტ / internet* ('internet'), and *აგრო / agro* ('agro'). More examples include: *ონლაინლექსიკონი / onlainleksik'oni* ('online dictionary'),

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

ონლაინშეხვედრა / onlainshekhvedra ('online meeting'), *ონლაინსწავლება / onlainstsavleba* ('online teaching'), *ონლაინგამოცემა / onlaingamotsema* ('online publication'), *ონლაინგაზეთი / onlaingazeti* ('online newspaper'), *აგრობიზნესი / agrobiznesi* ('agro business'), *აგროტურიზმი / agroturizmi* ('agro tourism'), *აგროსაწარმო / agrosats'armo* ('agro enterprise'), and *აგროლიზინგი / agrolizingi* ('agro leasing'). These word-forming components generate many new words in modern Georgian, thus enriching the language.

Georgian is a highly synthetic language with many affixes. Some of these affixes have become very productive in modern Georgian, for example the verb-forming prefix *და- / da-* and suffix *ება / eba, და-ება / da-eba*: *დაორგანიზება / daorganizeba* ('to organise'), *დასანქცირება / dasanktsireba* ('to sanction'), *დალაიქება / dalaikeba* ('to like'), *დამესიჯება / damesijeba* ('to message'), *დაფორვარდება / daporvardeba* ('to forward'), and *დაპარკინგება / dap'ark'ingeba* ('to park').

A more recent tendency is the increased number of semantic borrowings from English, as can be observed in the terminology of social media. Alongside borrowings such as *dalaikeba* ('to like'), *gashe areba* ('to share'), *peiji* ('page'), *woli* ('wall'), *prendi* ('friend'), and *poloueri* ('follower'), there developed new polysemous meanings of Georgian words, corresponding to *like*, *share*, *page*, *wall*, *friend*, and *follower*: *მოწონება / mots'oneba* ('to like'), *გაზიარება / gaziareba* ('to share'), *გვერდი / gverdi* ('page'), and *კედელი / k'edeli* ('wall'). Native Georgian words are mostly used in written or formal speech, while borrowings remain as their colloquial synonyms. For example, the borrowing *დალაიქება / dalaikeba* ('to like') coexists with its Georgian synonym *მოწონება / mots'oneba* ('to like'), and *ფრენდი / prendi* ('friend') with its Georgian synonym *მეგობარი / megobari* ('friend'). It should be noted that while these words are not new in terms of their form, they have developed semantically and have acquired new meanings, which are used in the context of social media.

Neologisms are also found in the lower register of the language. Such are, for example, slang words *ქრაში / krashi* ('crush'), *ქრინჯი / krinji* ('to cringe'), *დასტალკვა / dastalk'va* ('to stalk'), and *სლეი / slei* ('slay'). Again, it is clear that, as with other registers, the source of modern Georgian slang is also English.

6. Conclusion

This paper presents the methodology developed at Ilia State University for the identification of neologisms in modern Georgian, which is based on the exclusion method. A study corpus was created for the project, including textual material from Georgian-language online newspapers, magazines, news websites, media sites, and websites of NGOs and governmental agencies. The corpus contains over 100 million tokens and covers the last 20 years of the development of the Georgian language. Two lemmatisation tools created for the Georgian language and, based on the finite-state technology, were applied in the study. The first lemmatiser, developed at Ilia State University and which

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

contains macrostructures of the existing academic dictionaries, was used as an exclusion list for the study. Another lemmatiser, developed for the Georgian National Corpus, can guess and lemmatise unknown words and its application enabled to reduce the number of OOV lexis by 45%.

Although the lemmatisation tools performed a significant amount of work, it is not without limitations. The tools cannot detect new meanings of existing words in the language. The process is solely based on word form and does not possess any kind of "semantic intellect", as termed by Lemnitzer (2000). As a result, new word meanings are "invisible" to the programme. For instance, the term *virus*, originally referring to a disease, has acquired a new meaning with the evolution of technology and designates computer viruses. In Georgian, it now has one more meaning and refers to popular videos that spread quickly and widely on the internet. Such changes cannot be detected by this approach.

When dealing with text processing on the internet, it is important to underline the existence of many typographical errors. The lemmatisers used in the study cannot handle these errors and treat them as separate lexical units. Additionally, many Georgian words are not documented in dictionaries, and thus they are perceived as new by the tool and appear in the list of OOV lexis. Since the focus of this study is mainly on non-lemmatised content, the manual selection process becomes more time-consuming.

At the next stage of the project, neologisms identified during the study will be analysed thoroughly and classified, an online dictionary of neologisms will be composed and published, and a website will be set up for monitoring of Georgian neologisms in the future. The online dictionary of neologisms will be an explanatory dictionary, in which every word will be defined, supplemented by example sentences from the study corpus, and will contain an etymology part, explaining its source in modern Georgian. The work on the dictionary is underway and it will be published online in 2026.

Endnotes

1. The word denotes a Russian soldier.
2. Borrowing from English Robocop, a blend of robot + cop is formed. The word originates from an American science fiction film of the same title. In Georgian it is used to designate a special forces officer.
3. The word has originated from a Ukrainian name Titushko, a former sportsman who hired people to attack pro-European demonstrators in Kiev. In Georgian, it denotes a criminal hired by the Government to attack pro-European activists.
4. This word is of German origin Sonderkommandos, referring to punitive groups created by Nazi Germany. In Georgian, it means a member of a punitive group of criminals hired by the Government.
5. https://www.oxfordlearnersdictionaries.com/definition/english/chemistry?q=chemistry
6. https://enkacorpus.iliauni.edu.ge

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

7. LancsBox (2024) is an advanced software tool designed for analysing language data and corpora, and was used for tokenising downloaded texts. It was developed by a team of researchers at Lancaster University, led by Dr. Vaclav Brezina. See: http://corpora.lancs.ac.uk/lancsbox/.
8. qartnlp.iliauni.edu.ge
9. https://wordembedding.spellchecker.ge/

References

Digital tools

Georgian Word-Embedding Platform. 2021. https://wordembedding.spellchecker.ge/ [14 November 2024]

LancsBox. 2021. http://corpora.lancs.ac.uk/lancsbox [14 November 2024]

Lemmatisation Tool for the Georgian Language. 2021. https://qartnlp.iliauni.edu.ge/upload-xls [14 November 2024]

Other literature

Barbaresi, A. 2015. *Ad Hoc and General-Purpose Corpus Construction from Web Sources*. Unpublished PhD Dissertation. Lyon: École Normale Supérieure de Lyon.

Beliashvili, T. 2015. The Function of Neologisms in French and Georgian Media Discourse. *Spekali* 9: 1-11. Tbilisi: Tbilisi State University.

Cabré Castellví, M.T. and R. Estopà Bagot. 2009. Trabajar en neología con un entorno integrado en línea: la estación de trabajo OBNEO. *Revista de Investigación Lingüística* 12: 17-38.

Cook, P. and S. Stevenson. 2010. Automatically Identifying Changes in the Semantic Orientation of Words. Calzolari, N., K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapias (Eds.). 2010. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10), Valetta, Malta, 19–21 May 2010*: 28-34. Valletta: European Language Resources Association.

Davitishvili, N. 2018. On the Integration of Anglicisms into Present-Day Georgian. *Studies in Literature and Language* 17(3): 22-27.

EDGL. 1950–1964. *Explanatory Dictionary of the Georgian Language*. Chikobava, A. (Ed.). I–VIII volumes. Tbilisi: Mecniereba.

Falk, I., D. Bernhard and C. Gérard. 2018. *The Logoscope: A Semi-Automatic Tool for Detecting and Documenting French New Words from the Linguistic Project to the Web Interface.* Research Report, University of Strasbourg.

Gippert, J. and M. Tandashvili. 2015. Structuring a Diachronic Corpus. The Georgian National Corpus Project. Gippert, J. and R. Gehrke (Eds.). 2015. *Historical Corpora. Challenges and Perspectives*: 305-322. Tübingen: Narr.

Goshkheteliani, I. and M. Kikvadze. 2017. The Influence of English Borrowings on the Georgian Language. *Journal of Teaching and Education* 7(1): 459-463.

Grochocka, M. 2011. *Lexical Creativity in English: A Corpus-Based Study*. Unpublished PhD Dissertation. Poznań: Adam Mickiewicz University.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2044 (Lexicofocus)

Halskov, J. and P. Jarvad. 2010. Automated Extraction of Neologisms for Lexicography. Granger, S. and M. Paquot (Eds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22–24 October 2009:* 405-410. Louvain-la-Neuve: Presses universitaires de Louvain.

Janicijevic, T. and D. Walker. 1997. NeoloSearch: Automatic Detection of Neologisms in French Internet Documents. Lessard, G. and M. Levison (Eds.). *The 1997 Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing, Kingston, Ontario, Canada, 3–7 June 1997*. Kingston: Queen's University.

Janssen, M. 2009. Detección de Neologismos: una perspectiva computacional. *Debate Terminológico* 5: 68-75.

Kirvalidze, N. 2017. Linguo-Cultural and Pragmatic Peculiarities of the Phenomenon of Anglicisation in Georgia. *Journal of Teaching and Education* 6(2): 287-298.

Klosa-Kückelhaus, A. and I. Kernerman. 2022. GWLN — Globalex Workshop on Lexicography and Neology: An Overview. *Lexicala* 30: 10-13.

Klosa-Kückelhaus, A. and I. Kernerman. 2023. Unregistered Words, Neologisms, and Dictionaries: An Introduction. *Lexicography* 10(2): 87-93.

Lemnitzer, L. 2000. Einleitung und Hintergrund: Die Wortwarte — auf der Suche nach den Neuwörtern von morgen.
wortwarte.de/ [19 October 2022]

Lobzhanidze, I. 2021. *Principles of Morpho-Syntactic Annotation of Georgian and Morphological Analysis of the Finite State Position*. Tbilisi: Ilia University Press.

Margalitadze, T. 2020. Language and Ecology of Culture. *Lexicographica* 36: 225-240.

Margalitadze, T. 2022. Lexicography of Georgian. Hanks, P. and G.-M. de Schryver (Eds.). 2022. *International Handbook of Modern Lexis and Lexicography*: 1-24. Berlin/Heidelberg: Springer-Verlag.

Melikishvili, D. 2014. *Systemic Morpho-Syntactic Analysis of the Georgian Verb*. Tbilisi: Tbilisi State University.

Meurer, P. 2014. Morphosyntactic Analysis of Georgian. *Georgian National Corpus.*
gnc.gov.ge.

Mtchedlishvili, M. 2019. Political Neologisms in Global Politics and the Issue of their Transposition from English into Georgian. *Spekali* 13. Tbilisi: Tbilisi State University.

Paryzek, P. 2008. Comparison of Selected Methods for the Retrieval of Neologisms. *Investigationes Linguisticae* 16: 163-181.

Rayfield, D. 2023. The Indigestible Impact of English on the Modern Georgian Lexicon. *Digital Kartvelology* 2: 14-27.

Shanidze, A. 1973. *Foundations of Georgian Grammar.* Tbilisi: Tbilisi University Press (in Georgian).

Trap-Jensen, L. 2020. Language-Internal Neologisms and Anglicisms: Dealing with New Words and Expressions in the Danish Dictionary. *Dictionaries* 41(1): 11-25.

Trask, R.L. 1996. *Historical Linguistics*. Oxford: Oxford University Press.

Yano, T. and M. Kang. 2008. Taking Advantage of Wikipedia in Natural Language Processing.
www.cs.cmu.edu/~taey/pub/wiki.pdf [23 October 2022]

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Detection and Description of Neologisms in Korean Lexicography: Methodological Issues in Corpus Balance, Word Unit Bias and LLM Assistance*

Kilim Nam, *Department of Korean Language and Literature, Yonsei University, Seoul, South Korea (nki@yonsei.ac.kr)* (*https://orcid.org/0009-0001-9358-2673*)

Soojin Lee, *International Exchange Department, Kyungpook National University, Daegu, South Korea (sjmano27@naver.com) (https://orcid.org/0009-0002-1720-1880)*
and
Hae-Yun Jung, *International Exchange Department, Kyungpook National University, Daegu, South Korea (haeyun.jung.22@gmail.com) (https://orcid.org/0009-0006-4837-2569)*

Abstract: This study explores the potential application of large language models (LLMs) in Korean neologism extraction and dictionary compilation while critically examining the limitations of existing methods, including the bias toward news-oriented data and morphological neologisms. By analysing data from news corpora alongside messenger and online post corpora, the study identifies significant limitations in current news-centred approaches, particularly in detecting the first occurrences and extracting neologisms related to everyday topics. Experimental results involving LLMs demonstrate their potential to address the limitations of news-biased neologism extraction by suggesting unregistered words from diverse web-based contexts. However, issues such as duplication and overgeneration persist. In tasks involving semantic neologism recommendation and dictionary microstructure creation, LLMs performed relatively well with high-frequency and news-biased topics when provided with additional contextual prompts, yet revealed limitations with low-frequency and non-news-biased neologisms. These findings suggest that the performance of current LLMs heavily relies on the diversity of training data and user-provided contextual information. The results of this study underscore the need for further investigation into

* A version of this paper was presented at the 6th Globalex Workshop on Lexicography and Neology (GWLN-6), held on 3 July 2024 at the University of Pretoria, Hatfield Campus, Pretoria, South Africa.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

the critical challenges in neologism research, lexicography, and corpus linguistics, as well as the role lexicography might play in enhancing the performance of LLMs.

Keywords: LEXICOGRAPHY, NEOLOGISMS, UNREGISTERED WORDS, NEWS CORPUS, SEMANTIC NEOLOGISM, REPRESENTATIVENESS, BALANCE, LEXICOGRAPHIC DATA, MACROSTRUCTURE, LARGE LANGUAGE MODELS

Opsomming: Die opsporing en beskrywing van neologismes in die Koreaanse leksikografie: Metodologiese kwessies rakende korpusbalans, vooroordeel teenoor woorddele en ondersteuning deur groottaalmodelle. In hierdie studie word die potensiële toepassing van groottaalmodelle (GTM'e) in Koreaanse neologisme-onttrekking en woordeboeksamestelling verken, terwyl die beperkings van bestaande metodes, insluitend die vooroordeel ten gunste van nuusgeoriënteerde data en morfologiese neologismes, krities ondersoek word. Deur nuuskorpusdata naas kitsboodskap- en aanlynposkorpora te ontleed, identifiseer hierdie studie beduidende beperkings in huidige nuusgesentreerde benaderings, veral in die opsporing van eerste voorkomste en die onttrekking van neologismes wat met alledaagse onderwerpe verband hou. Empiriese resultate in die gebruik van GTM'e demonstreer hul potensiaal om die beperkings van nuusbevooroordeelde neologisme-onttrekking aan te pak deur ongeregistreerde woorde uit diverse webgebaseerde kontekste voor te stel. Kwessies soos duplisering en oorgenerering bestaan egter steeds. In take wat aanbevelings rakende semantiese neologismes en woordeboekmikrostruktuurskepping behels het, het GTM'e relatief goed gevaar met hoëfrekwensie- en nuusbevooroordeelde onderwerpe wanneer hulle van addisionele kontekstuele prompe voorsien is, maar tekortkominge het ook duidelik by laefrekwensie- en nienuusbevooroordeelde neologismes geblyk. Hierdie bevindings dui daarop dat die prestasie van huidige GTM'e sterk op die diversiteit van opleidingsdata en gebruikersverskafde kontekstuele inligting staatmaak. Die resultate van hierdie studie beklemtoon die behoefte aan verdere ondersoeke na die kritieke uitdagings in neologisme-navorsing, die leksikografie en korpuslinguistiek, sowel as na die rol wat die leksikografie in die verbetering van die prestasie van GTM'e kan speel.

Sleutelwoorde: LEKSIKOGRAFIE, NEOLOGISMES, ONGEREGISTREERDE WOORDE, NUUSKORPUS, SEMANTIESE NEOLOGISME, VERTEENWOORDIGENDHEID, BALANS, LEKSIKOGRAFIESE DATA, MAKROSTRUKTUUR, GROOTTAALMODELLE

1. Introduction

Corpus research, which was pioneered in lexicography, has since developed into the methodology of corpus linguistics and has been expanded, refined, and eventually adopted by neological studies, among others. News corpora in particular have been used as authentic language data and a strong basis for the identification of institutionalised neologisms (as opposed to nonce words), the dating of their first appearances, and the investigation of their usage trends. Their status as the greatest resource for neologism collection and study has been abundantly discussed (cf. Renouf 2013; Boussidan 2013; Nam et al. 2020; Freixa and Adelstein 2013; Klosa and Lüngen 2018). However, the spread of web languages and

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

the emergence of large language models (LLMs) have a considerable impact on the creation and diffusion of neologisms, as well as on the application of language resources. In that sense, it has become crucial to re-examine not only the bias of [+formal] and [+written] news corpus, but also the neologism extraction methods centred on single-word units. This study critically reviews the methodology for Korean neologism research, which has relied on the semi-automated extraction from news corpora from 2005 to date, and explores ways to improve dictionary compilation to reflect the dynamics of language from the actual language use by different communities.

This paper is organised as follows. Section 2 reviews current methodologies for extracting neologism extraction and dictionary compilation and presents the research methodology used for this study. Three corpora have been compared and analysed, namely a news corpus of 500 million words, a 14-million-word corpus of online posts from forums and social media, and a 6-million-word instant messages (IM) corpus, which roughly span from 2020 to 2022.

Sections 3 and 4 test the potential of large language models (LLMs) for neologism detection and description as recent LLMs have begun to be discussed in language research and dictionary compilation. Their ability to identify and describe neologisms is reviewed to see if they can somewhat compensate for the limitations of existing neologism extraction methods through learning from large-scale data. Python programmes are used for corpus analysis and processing, and the detection of unregistered words draws on the largest Korean dictionary, *Urimalsaem*, as the accumulated list of neologisms to date. The study identifies three key issues related to the detection of neologisms and their lexicographic description, each of which is addressed in a section.

Firstly, Section 3.1 investigates issues related to the source genre and the news bias of neologisms. Based on domain, frequency, and distribution analyses, it is shown that neologisms mainly reflect news topics, such as politics, economy, and society, and comprise very few expressions pertaining to daily life, such as food-related terms or emotion terms. Despite their low frequencies in newspapers, everyday expressions that are frequently used by Korean speakers, especially those reflecting their attitudes or feelings on everyday topics, hold significant value as headword candidates. This section therefore argues for the need to diversify corpus resources beyond newspapers in order to collect neologisms that reflect native speakers' expressions and to build dictionaries that capture the dynamism of language.

Section 3.2 secondly examines how a comprehensive representation of the lexicon of a given era requires not only the identification of new forms, but also the analysis of frequencies, distribution, and discourse contexts. This entails the extraction and lexicographic description of semantic neologisms in addition to formal neologisms. Case studies are drawn upon to show that neologism research must be discussed in discourse context, which is closely related to phrase-level headwords, such as typical collocations and patterns.

Finally, Section 4 discusses the potential and limitations of LLMs in collecting and describing neologisms, both as an alternative to the issues raised in

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

the previous two sections, and as a new approach to lexicography. Although recent studies — including McKean and Fitzgerald (2023), Lew (2023, 2024), and De Schryver (2023) — have discussed the use of LLMs for lexicographic description, there has been little research discussing the current state of LLMs in the detection and description of neologisms or unregistered words. Two factors could explain this lack of research: the training data of LLMs is firstly not up-to-date, and secondly, neologisms only play a small part in lexicographic description. However, the detection of neologisms and unregistered words, which relies on corpora and lexicographers' intuition, still faces a number of challenges, such as the diversity in genres and topics, the exhaustiveness of detection (which tends to display a bias toward formal neologisms), and the subjectivity of neologism description. The section explores whether LLMs can contribute to the description of neologisms and unregistered words and enhance lexicographic productivity by means of concrete and extensive experiments. These experiments focus on the identification of neologisms, the recommendation of headword candidates, and the description of lexical microstructures. The experiments involve the main LLMs for Korean, namely the native Naver CLOVA[1] and the international model ChatGPT. The evaluation of their performances and the comparison between the two show how the volume of Korean data used in training, as well as the training based on existing dictionaries, can affect their ability to identify and describe neologisms.

More than twenty years ago, Sinclair (2004: 188-192) emphasised the need for the linguistics community to prepare for larger, unannotated raw corpora to contribute to the future information society. In the era of LLMs, it has become crucial to examine whether large corpora alone are sufficient for automatic identification of neologisms and unregistered words, and to balance corpora and LLMs as linguistic resources and methodologies in order to address written-language bias and complement human intuition. Thus, this paper argues that the meaningful contribution of neologism research to society in the era of LLMs entails building larger corpora, redefining corpus diversity and balance, developing new extraction methodologies, and compiling dictionaries that reflect the linguistic dynamics of native speakers.

2. Research background and methodology

2.1 Overview of the Korean Neologism Investigation Project

In terms of sources, the Korean Neologism Investigation Project (KNIP) is primarily focused on major daily newspapers, and in terms of methodology, it can be characterised as a semi-automatic approach to neologism extraction, that is, combining automatic extraction and manual collection of new form candidates. Table 1 summarises the types of sources and methodologies[2] used by the National Institute of Korean Language (NIKL) as well as by the Kyungpook National University (KNU) research team (highlighted in grey) in the neologism investigation project from its inception to date.[3]

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Table 1: Sources and methods for the investigation of neologisms by the NIKL and KNU

	Time spans	Sources	Methodology
Phase 1	1994–2001	major national daily newspapers and magazines	manual extraction
Phase 2	2002–2004	major national daily newspapers and broadcasts	not mentioned
Phase 3	2005–2010	major national daily newspapers and broadcasts	corpus, semi-automatic extraction
Phase 4	2012–2019	around 130 online news websites	corpus (web-based extractor), semi-automatic extraction
Phase 5	2020 to date	5 online news websites	corpus, semi-automatic extraction

The issues raised here are twofold: (1) a bias toward the written language of the news genre as the target of the investigation, and (2) limitations in semi-automatic analysis based on 'forms', reflecting the constraints of the underlying corpus and extraction methodology, respectively. Despite the premise that in any corpus-based research, including corpus-based lexicography, corpus representativeness and corpus balance must be ensured, the collection of neologisms has relied on mass media. The main reason for this is that mass media provide evidence of 'institutionalised' neologisms, as opposed to nonce words. Major daily newspapers in particular provide information fulfilling crucial criteria, such as the date of first occurrence, frequency and distribution. Nonetheless, collecting news-based neologisms also presents limitations, including the restricted scope of neologisms, as these mostly revolve around public affairs, and the inability to identify the first occurrence and spread of some neologisms.

(1) Examples of news-biased neologisms
   a. Domain bias: *kinkupcaynansayngkyeyciwenkum*[4] ('emergency disaster relief support'), *cwungswucheng* ('Serious Crimes Investigation Agency'), and *panmwuncemsenen* ('Panmunjeom Declaration')
   b. First appearance and spread bias: *phathpwung* ('red bean pastry') (blogs: 2012; news: 2021), *mayneekkey* ('manners shoulder') (blogs: 20 occurrences in the past year; news: 0 occurrences)

Example (1a) provides typical examples of domain-biased neologisms when these are only extracted from news sources. Such neologisms focus on topics related to public affairs, such as politics, administration, law, economy, and society. In contrast, compared to neologisms concerned with public affairs, private communication texts tend to generate and spread more neologisms related to everyday

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

topics, such as food and hobbies, or expressions of personal feelings. As a result, the frequency of expressions such as those presented in (1a) is relatively low outside of news texts. *Phathpwung* ('red bean pastry') in (1b) first appeared in blogs in 2012 but was not mentioned in newspapers until 2021; furthermore, *mayneekkey* ('manners shoulder') was mentioned more than 20 times in blogs in the past year, but not once in newspapers. These examples show the limitations of news sources in fully capturing the first appearance and spread of neologisms while these rapidly spread in online media.

As for the limitation of form-centred neologism extraction, it correlates with the main characteristics of current neologism research methods. At present, the extraction of Korean neologisms consists of semi-automatic extraction of neologisms based on the comparative analysis of neologism candidates from a large-scale news corpus against existing lexical resources. The identification of neologisms based on such a form-centred comparative analysis usually means dismissing semantic neologisms or grammatical neologisms, wherein new meanings are created even though they do not involve new forms (Nam 2020; Nam 2021; Nam and An 2023). Therefore, phenomena such as semantic extension or acquisition of metaphorical meanings through new collocational combinations, as illustrated in Example (2), are more often than not overlooked in the extraction and lexicographic description of neologisms.

(2) Examples of semantic neologisms:
   a. *phokphwung* ('storm'): (1) storm warning, tropical storm; (2) storm (as in 'extreme') growth/tears/eating
   b. *konghwakwuk* ('republic'): (1) democratic republic, Hellenic Republic; (2) apartment/Seoul/real estate republic

The creativity of native speakers includes not only grammatical creativity as discussed since Chomsky (1957), but also morphological creativity and semantic creativity. Specifically, these refer the creation of new words or sentence structures, and the creation of new meanings through novel collocations or metaphors, respectively. In this context, fundamental methodological changes seem crucial to ensure that these creative meanings are no longer overlooked in neologism extraction and dictionary description. These two issues, namely the bias towards news language and the bias towards formal neologisms, are further discussed in Sections 3.1 and 3.2, respectively.

2.2 Methodology

This study focuses on large-scale neologism collection efforts combined with corpus use from 2005 to the present and surveys a total of 11 neologism datasets (2005, 2008–2010, and 2012–2022). In addition, the study examines two other corpora, which consist of distinct registers from the news and were predominantly used in previous neologism studies, to assess the adequacy of the linguistic

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

resources and methodology used for neologism collection. These corpora are the *Instant Messaging Corpus* (IM) and *Online Post Corpus* (OP) compiled by NIKL as part of the *Modu Corpus* (lit. 'All Corpora'). Additionally, the *Newspapers Corpus*, also from the *Modu Corpus*, was used for comparison. Table 2 shows the size of each corpus in *ecel* unit, which is the Korean word unit based on word spacing.[5]

Table 2: Size of the corpora under study in *ecel* (space unit word)

Source	Corpus	**Time Segment 1 (01-07-2019–30-06-2020)**	**Time Segment 2 (01-07-2020–30-06-2021)**	**Time Segment 3 (01-07-2021–30-06-2022)**	**Total**
NIKL	News	167 286 198	225 402 761	100 892 770	493 581 729
	IM	2 826 229	2 834 967	788 620	6 449 816
	OP	3 206 220	4 969 648	6 371 497	14 547 365

These corpora were analysed to identify issues related to the biases of linguistic resources predominantly used in previous neologism investigations. In other words, this study examines whether neologisms were omitted and whether certain neologisms first appeared in media or genres other than news outlets. The datasets shown in Table 2 actually have substantial differences in terms of time spans; as a result, the study focuses on a relatively short time frame of three years, during which all three datasets overlap. The selection of this time frame aligns with the characteristics of the neologisms focused on in this study. Neologisms constitute a highly dynamic lexical category, exhibiting emergence and disappearance within very short periods of time. As such, the analysis of the subtle dynamics of neologisms calls for a short-term diachronic approach,[6] as opposed to traditional diachronic studies looking into changes over centuries.

3. Korean neologism research: Current status and corpus linguistic issues

3.1 The bias of news language and challenges in neologism extraction

The advantage of using newspapers and news corpora for collecting neologisms lies not only in their widespread dissemination across the Korean-speaking community, but also in their capacity to reflect societal changes quickly. Nonetheless, it is evident that news data cannot fully represent the linguistic reality of Korean native speakers. Therefore, biases related to news data can also be found in the results of the neologism investigations that are based on such data. For instance, Lee (2022) analysed the semantic categories of neologisms from 2015 to 2021 and found a high proportion of neologisms in the categories of

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Politics, Economy, and Society, correlating such results to the limitation of news text genre as the primary source for neologisms.

In the present study, the IM and OP corpora have been analysed for comparison with news sources, focusing on the date of first occurrence and topic diversity in order to identify biases related to written language. If a given neologism has been actively used in other domains long before it appeared in the news texts, it becomes evident that the timing of first occurrences needs to be revisited. In the same vein, if neologisms from news sources are biased toward topics pertaining to public affairs due to the [+public] nature of the news genre, the question of collecting neologisms that emerge and are used in [+private] domains also needs to be addressed.

As seen in Table 2, the *Newspaper Corpus*, *IM Corpus*, and *OP Corpus* were divided into three time segments of one year each. The second time segment (July 2020–June 2021) coincides with the time frame for the 2021 neologisms, and the third time segment (July 2021–June 2022) partially coincides with the time frame for the 2022 neologisms.[7] When analysing non-news genre data (i.e. instant messaging texts and online posts) separately, it appeared that a few neologisms collected in 2021 and 2022 occurred in the *IM Corpus* and the *OP Corpus* at earlier date than the extraction time frame, as shown in Table 3.[8]

Table 3: Neologisms of 2021 and 2022 that appear in IM and OP before News

	Neologism	First occurrence within Time Segment 1		First occurrence within Time Segment 2	
		IM	Online posts	IM	Online posts
2021 neologisms	*ttasangsang* ('double top stock')	✓			
	ssalmek ('items-for-rice')	✓			
	ccomccomttali ('teeny-weeny bits')	✓			✓
2022 neologisms	*phathpwung* ('red bean pastry')	✓			✓

Based on the News Corpus, the neologism *phathpwung* ('red bean pastry') was extracted within the time frame for the 2022 neologism collection. However, Table 3 shows it first appeared within the Time Segment 1, rather than Time Segment 2. Of these four neologisms, all but *ccomccomttali* ('teeny-weeny bits') share the common feature of being abbreviations, and only *ttasangsang* ('double top stock'), which falls under the domain of Economy, belongs to the semantic domains of Politics, Economy, and Society.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Table 4: First occurrence dates of *ttasangsang* ('double top stock'), *ssalmek* ('items-for-rice'), *ccomccomttali* ('teeny-weeny bits), *phathpwung* ('red bean pastry') in News and Blogs

	Neologism	Naver News in-links	Naver News out-links	Blogs
2021 neologisms	*ttasangsang* ('double top stock')	03-07-2020	06-07-2020	17-11-2019
	ssalmek ('items-for-rice')	10-07-2020	14-10-2021	07-08-2017
	ccomccomttali ('teeny-weeny bits')	31-01-2021	09-12-2020	27-02-2017
2022 neologisms	*phathpwung* ('red bean pastry')	09-12-2021	27-12-2020	13-02-2012

Table 4 shows the first occurrence dates of the four neologisms in Naver News in-link data as the timing criterion for neologism extraction and compares with their appearances in Naver News out-link data and blogs.[9] The 2022 neologism *phathpwung* ('red bean pastry') should be included within the time frame of the 2021 neologisms when expanding the scope to Naver News out-link data, even though the sources are still limited to the news genre. If the scope is further extended to blog data, the neologism should have been collected as a 2012 neologism as it appeared in blogs as early as February 2012. As the other three neologisms show the same tendencies, it could be assumed that the first occurrence dates of many other neologisms would be significantly earlier if the timing criterion was expanded to sources other than news.[10] Given the popularity and diversity of speakers in IM platforms (Huang and Nam 2023) and considering the multiplicity and receptivity of blogs as both interpersonal and mass communication (Smyk-Bhattacharjee 2009), web out-links need to be taken into consideration as a means to reflect the linguistic realities of native speakers that could potentially offer a more effective alternative to news media.

To explore the issue of written register bias further, derivatives based on the same suffix *-sulep-* are examined and compared in terms of occurrences and distribution across news, IM, and OP corpora. The suffix *-sulep-* is used to create adjectives meaning 'being like/having the quality of' the noun it is affixed to, and was chosen for its established high productivity, thereby providing a particularly suitable case for analysis. The suffix *-sulep-* ('-like') can be easily combined within rules and without significant restrictions with most types of nouns, including common nouns, proper nouns, native Korean words, and loanwords. As a result, many derivatives are nonce-words or short-lived expressions that do not undergo the process of institutionalisation. However, the regularity of derivation the suffix presents can differ depending on the genre or use domain.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Table 5 presents the number of occurrence types and the size of the corpus of unregistered *-sulep-* derivatives across news, IM, and OP corpora.

Table 5: Occurrence frequency for unregistered *-sulep-* derivatives by text genre

	News	IM	OP
Number of types	113	142	245
Corpus scale (in *ecel*)	220 million	10 million	10 million

The *News Corpus* is by far the largest corpus, followed by the *OP Corpus* and the *IM Corpus* in that order. However, the number of unregistered *-sulep-* derivatives is the lowest in the *News Corpus*, the largest number being found in the *OP Corpus*. In other words, many more *-sulep-* derivatives are created in a corpus that is much smaller than the *News Corpus*. This underlines the close correlation between the characteristics of the text genre and the creation of new words. Unlike newspapers, which tend to use a more formal and prescriptive language, online posts and instant messages allow users to deviate from norms freely, to play with language, and to create new expressions. Therefore, to capture a wide range of new words, it is necessary to look into diverse linguistic resources that take into account the characteristics of text genres and linguistic creativity, rather than solely relying on the size of the corpus.

(3) Examples of occurrences that are common to news and IM/OP genres (20 in total)
   a. *pwukhansulepta* ('North Korean-like') and *hankwuksulepta* ('South Korean-like')
   b. *pokswungasulepta* ('peach-like') and *ppangcipsulepta* 'bakery-like')
   c. *opesulepta* ('over-ish') and *cwacelsulepta* ('frustration-inducing')
   d. *salamsulepta* ('people-like') and *sencinkwuksulepta* ('developed country-like')

(4) Examples of occurrences exclusive to IM/OP genres (307 in total)[11]
   a. *tiolsulepta* ('Dior-like'), *meylukheylsulepta* ('Merkel-like'), and *sewulsulepta* ('Seoul-like')
   b. *chokhosulepta* ('chocolaty') and *haympekesulepta* ('hamburger-like')
   c. *kkamccaksulepta* ('surprise-inducing') and *ttiyongsulepta* ('boing-like')
   d. *kwiyepppoccaksulepta* ('cutie pie-like') and *kokwumasulepta* ('sweet potato-like')
   e. *cwummasulepta* ('auntie-like') and *cwung2sulepta* ('teenage-like'); *kemchalsulepta* ('prosecution-like') and *kwuninsulepta* ('military-like')
   f. *choposulepta* ('beginner-like') and *weneminsulepta* ('native-like')

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

g. *keyimsulepta* ('gaming-like'), *okhelthusulepta* ('occult-ish'), and *weypthwunsulepta* ('webtoon-like')
h. *kolmoksulepta* ('backstreet-like'), *khapheysulepta* ('café-like'), and *hyuyangcisulepta* ('vacation spot-like')

(5) Examples of exclusive 'Newspaper' types (74 in total)
a. *khakhaosulepta* ('Kakao-like') and *yunsekyelsulepta* ('Yoon Suk-Yeol-like')
b. *panillasulepta* ('vanilla-like') and *sukhonsulepta* ('scone-like')
c. *kammyengsulepta* ('impression-making') and *ongcolsulep* ('narrow-minded-ish')
d. *wisensulepta* ('hypocrisy-like') and *huymangsulepta* ('hope-like')
e. *haksayngsulepta* ('student-like') and *kongmwuwensulepta* ('civil servant-like')
f. *loksulepta* ('rock-like') and *khenchyulisulepta* ('country-like')

The *-sulep-* derivatives presented in Examples (3) to (5) have been grouped according to the base type *-sulep-* is suffixed to: (a) corresponds to proper nouns, (b) food-related terms, (c) emotion-related terms, and (d) object attributes or characteristics, all four categories being common to news, IM and OP genres. On the other hand, (4.e) refers to terms denoting [person] terms based on age or occupation, while (4.f) corresponds to [person] terms describing their personality or characteristics. Example (4.g) refers to contents consumed for hobbies or entertainment, and (4.h) groups terms whereby the suffix is combined with a place-related base. Just as (4.e) and (4.f), the bases in (5.e) represent [person] terms pertaining to profession, personality, or characteristics; the main difference is that (5.e) has a limited range of examples compared to the much more varied types observed in (4.e) and (4.f). Example (5.f) is somewhat akin to (4.g) in that the bases denote it subgenres of [music] as cultural contents to be enjoyed or consumed.

Among the types of *-sulep-* derivatives that appeared exclusively in the news genre, proper noun bases mainly pertain to politicians, celebrities, or companies' names, and emotion-related bases are terms frequently found in formal written language, such as *kammyeng* ('impression') and *ongcol* ('narrow-mindedness'). In contrast, the derivatives appearing only in non-news genres, often include combinations with food-related nouns — e.g., *chokho* ('choco') and *haympeke* ('hamburger') — and with terms related to leisure time — e.g. *keyiming* ('gaming'), *okhelthu* ('occult'), and *weypthwun* ('webtoon'). In addition, bases denoting emotions, such as the root *kkamccak* ('surprise') or the onomatopoeia *ttiyong* ('boing'), further highlight the difference from derivatives exclusive to the news genre.

These findings suggest that the thematic characteristics of the text genre can have a significant impact on the type of vocabulary collected. For instance, the news genre predominantly reflects topics such as politics, economics, and society, whereas IM and OP genres focus on everyday topics, such as food or hobbies. This confirms the need to diversify domains and genres beyond news sources to more fully represent the linguistic creativity of native speakers.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

3.2 The bias of formal neologisms

From the start, KNIP has sought to detect semantic neologisms, defined as existing word forms used with a different meaning, along with formal neologisms. Accordingly, the early neologism reports marked headwords with the symbol ♣ or added phrases such as '♣ different meaning' or '※ different meaning' at the end of the entry to indicate that the headword was a semantic neologism, as shown in Example (6).

(6) *kancephwapep* 'indirect statement' (*1994 Neologism Report*, p. 3). 'In the early 1980s, Castro announced, by means of indirect statements made in state-run newspapers, the government's position that Cuban residents could emigrate overseas without interference from authorities.' (*Joseon Ilbo*, 94.08.21, p. 2)
♣ different meaning

In total, there are around 260 examples of semantic neologisms across the Neologism Reports, which represent a small number compared to formal neologisms. The yearly trends in the collection of semantic neologisms are presented in Table 6, with representative cases highlighted in grey.

Table 6: Number of semantic neologisms in the neologism reports

Year	1994	1995	2000	2001	2002	2003	2004	2005	2008	2009	2010	2012	total
Semantic neologisms	45	26	53	25	36	24	41	3	3	3	3	1	263

What can be immediately observed from Table 6, is the significant decrease in the number of semantic neologisms collected from Phase 3 onwards, that is, following the introduction of semi-automatic extraction in 2005, compared to the first and second phases, which heavily relied on manual collection. Furthermore, no more semantic neologisms have been collected from Phase 4 (2012–2019), when web-based automatic extraction was implemented, through to Phase 5 (2020–). While the development of Korean neologisms research can thus be divided into five phases, progressing from manual extraction of neologisms to extraction through static corpora and the use of dynamic time-series corpora, the interest in semantic neologisms has either stagnated or become even more passive.

In fact, the collection of semantic neologisms even seemed to retrogress, with more semantic neologisms included in the datasets during the manual collection phase than during the semi-automatic extraction phase. Moreover, the consideration of semantic neologisms has further weakened with the use of large-scale corpora, since semantic neologisms, having the same form as existing words, are ipso facto excluded from the list of neologism candidates, which is from the very first stage of the neologism investigation. Recent studies have nonetheless

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

looked into the automatic identification and detection of semantic changes in large corpora (cf. Boussidan 2013; Renouf 2013; Nam et al. 2018; Nam et al. 2019). While automatic and semi-automatic identification and detection of meanings have been proven feasible, it is still difficult to propose methods that can be directly applied to actual collection tasks. Nevertheless, the need remains to establish criteria for identifying the emergence and first occurrence date of new meanings, just as for formal neologisms. Continuous research and discussion are thus needed for determining what constitutes semantic novelty and for identifying the first occurrence of semantic neologisms.

4. Review of Korean LLMs for the lexicographic compilation of neologisms

4.1 Korean neologism research and experiments using LLMs

A number of experimental studies have recently been conducted to examine how LLMs can contribute to the automation of dictionary compilation and increase lexicographic productivity in line with lexicographic advancements, such as the use of corpora and computers, the development of editing tools, and the participation of dictionary users. McKean and Fitzgerald (2023) explore the potential and limitations of LLMs in lexicography by testing the performance of ChatGPT in headword recommendation and entry description. Other studies, including De Schryver (2023) and Lew (2023, 2024), discuss the impact and potential applications of LLMs in dictionary compilation extensively. A common conclusion is that although LLMs have shown considerable promise for automating lexicography, significant issues remain regarding accuracy, such as incorrect examples and hallucinations. However, most of these studies have not included discussions on the detection of neologisms and unregistered words, which play an important role in macrostructure expansion, nor on the description of neological headwords.

In light of this underexplored area, this section discusses the potential applicability of LLM-based dictionary compilation to the detection and description of Korean neologisms and unregistered words. As seen in Section 3, there are clear limitations in the collection and description of neologisms, having relied heavily on corpora and human intuition. Two scenarios could thus unfold from experimenting with LLMs in regard to the description of neologisms and unregistered words. Firstly, LLMs may be able to overcome or complement the limitations of existing neologism detection and description methodologies. In this case, LLMs could serve as a new methodology to address the shortcomings of traditional methodologies, namely the reliance on human intuition and corpora, thereby significantly improving lexicographic productivity. The second scenario is the opposite outcome: If LLMs are found unreliable in recommending and describing new and unregistered words, the role of lexicographers with 'intuition' and the construction of lexicographic databases will become even more critical. Moreover, the importance of lexicographic databases and descriptions

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

that can compensate for the weaknesses and limitations of LLMs will need to be discussed in greater depth.

For the experiments of detection and description of Korean neologisms or unregistered words,[12] this study uses OpenAI's ChatGPT (GPT-4 version), the most widely known generative AI and international model, and Naver CLOVA X (HyperCLOVA X version), a domestic model.[13] For both AI models, the experiments consisted of entering prompts on the user platforms.

In Section 4.2, LLMs are tested on the recommendation of neologism candidates spanning a broad range of genres to address the news bias discussed in Section 3.1. In Section 4.3, LLMs are prompted on the definition as well as the description of the semantic neologisms.

4.2 Experiments on news bias: Detection of neologisms and unregistered words

Section 3 examined the issue of genre bias in the semi-automatic extraction methods based on news-centred corpora traditionally used in neologism research, as demonstrated by the cases of unregistered *-sulep-* ('-like') derivatives. Here the question is whether LLMs trained on large-scale data encompassing various sources, such as encyclopaedias like Wikipedia, books, academic articles, news, blogs, forums, and other websites, can serve as an alternative.

To answer this, the two LLMs that are most utilised by Korean users (ChatGPT and CLOVA X) have first been tested to assess whether they can effectively perform the task of recommending neologisms and unregistered words. The query was for them to recommend a list of *-sulep-* derivatives not found in existing dictionaries, as shown in Example 7. In the case of ChatGPT, the LLM was trained on Korean data from the state-led *Standard Korean Language Dictionary* and the headword lists of the Korean dictionaries provided by Naver[14] and Daum[15] portals. Similarly, CLOVA X learned from the headword lists of the *Standard Korean Language Dictionary*, *Urimalsaem*, and the *Korea University Korean Dictionary*, all provided by Naver Dictionary Service.[16]

(7) Prompt:[17]
Make a list of 35 derivatives with the suffix *-sulep-* '-like'. Only include words that meet the following conditions:
- Not already included in an existing dictionary.
- Not AI-invented: they must be words that people have actually used.

(8) Generative AI answer: ChatGPT
a. {*chinkwu* 'friend', *kamseng* 'feeling', *hyensil* 'reality', *inkan* 'human', *kacok* 'family', *chengchwun* 'youth', *phulo* 'pro', *pinthici* 'vintage', *koyangi* 'cat', *hipsuthe* 'hipster', *wucwu* 'space', *semin* 'commoner', *tokthuk* 'unique', *ipeynthu* 'event', *centhong* 'tradition', *hyuka* 'vacation', *tosi* 'city', *yelceng* 'passion', *mohem* 'adventure', *chwimi* 'hobby', *yelceng* 'passion', *yehayng*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

'travel', *milay* 'future', *cengcik* 'honesty', *yuhayng* 'fashion', *chopo* 'beginner', *tokse* 'reading', *cayen* 'nature'}*sulepta* '-like'

b. {*cayu* 'freedom', *sonye* 'girl', *phyenghwa* 'peace', *kamtong* 'emotion', *chincel* 'kindness', *sikol* 'countryside', *cayen* 'nature', *pimil* 'secret'}*sulepta* '-like'

Of the 35 derivatives suggested by ChatGPT, *yelceng* ('passion') is suggested twice, and 8 suggestions are already included in the Korean language dictionary *Urimalsaem* (8.b), reducing the list of suggestions to 26 unregistered derivatives. Additionally, 12 of them can be found in the *News*, *IM*, and *OP* corpora under study (Example 9).

(9) ChatGPT suggestions found as unregistered words in the corpus under study

a. OP: {*kamseng* 'feeling', *koyangi* 'cat', *milay* 'future', *pinthici* 'vintage', *wucwu* 'space', *centhong* 'tradition', *chengchwun* 'youth', *chopo* 'beginner', *phulo* 'pro'}*sulepta* '-like'

n. IM: *seminsulepta* 'commoner-like'

d. News: *milay* 'future', *yuhayng* 'fashion', *centhong* 'tradition', *chengchwun* 'youth', *hipsuthe* 'hipster'}*sulepta* '-like'

Out of these, {*milay* 'future', *centhong* 'tradition', *chengchwun* 'youth'}*sulepta* '-like' are common to both the *OP Corpus* and the *News Corpus*.

(10) ChatGPT suggested unregistered words with actual web usage[18]
{*chinkwu* 'friend', *hyensil* 'reality', *inkan* 'human', *kacok* 'family', *tokthuk* 'unique', *ipeynthu* 'event', *hyuka* 'vacation', *yelceng* 'passion', *mohem* 'adventure', *chwimi* 'hobby', *yehayng* 'travel', *cengcik* 'honesty', *tokse* 'reading'}*sulepta* '-like'

While {*tokthuk* 'unique', *ipeynthu* 'event', *cengcik* 'honest', *tokse* 'reading'}*sulepta* '-like' are likely nonce-words given their fairly low Google search frequencies, it is nonetheless worth noting that none of the 35 ChatGPT suggestions were "invented" by ChatGPT as actual forms used by Korean speakers. In other words, both Examples (9) and (10) can be considered as headword candidates, although further research on their actual frequencies and usage patterns may be needed.

(11) Generative AI Answers: CLOVA X

a. *kwiyemsulepta* 'lovable', *angcungsulepta* 'cute', *messulepta* 'stylish', *nekulewumsulepta* 'magnanimous', *alumtawumsulepta* 'beauty-like', *calangsulepta* 'proud', *mancoksulepta* 'satisfying', *tahayngsulepta* 'lucky', *kekcengsulepta* 'worrisome',

b. *cengtapta* 'affectionate'/*cengtapta* 'affectionate', *alumtawum* 'beauty'/*alumtawum* 'beauty',

c. *sulkilopta* 'wise', *sulkilowum* 'wisdom'/*sulkilowum* 'wisdom',

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

d. *hayngpoksulewum* 'happiness'/*hayngpoksulewum* 'happiness'/*hayngpoksulewum* 'happiness', *salangsulewum* 'loveliness'/*salangsulewum* 'loveliness', *cayensulewum* 'naturalness'/*cayensulewum* 'naturalness', *poksulewum* 'fortune'/*poksulewum* 'fortune', *kwiyemsulewum* 'lovability'/*kwiyemsulewum* 'lovability', *angcungsulewum* 'cuteness'/*angcungsulewum* 'cuteness', *messulewum* 'stylishness'/*messulewum* 'stylishness'
e. *calangsulewum* 'proudness', *mancoksulewum* 'satisfaction', *tahayngsulewum* 'luckiness', *kekcengsulewum* 'worry'

The first issue with CLOVA X answer is that it included not only *-sulep-* derivatives but also *-tap-* (11.b) and *-lop-* (11.c) derivatives. While the latter two can be considered synonymous to *-sulep-* ('-like'), it was not the task requested, suggesting that CLOVA X is likely to include unwanted noise in the results when asked to recommend unregistered words or neologisms as headword candidates. The second issue is the high number of duplicates.[19] Although CLOVA X suggested a list of 35 items as requested by the prompt, 10 suggestions are actually repeated two or three times as can be seen in (11.b to 11.d). Another problematic aspect of CLOVA X results is that 6 items are *-sulep-* derivatives in the adjectival form *-sulepta* (11.a). Most of the other derivatives returned by CLOVA X are the nominalised form *-sulewum* (11.d and 11.e). This again leads to the issue of duplicates as shown in Example (12), where pairs such as *kekcengsulepta* 'worrisome'/*kekcengsulewum* 'worry' are simply two word forms of the same lexeme.

(12) Adjective/noun duplicates suggested by CLOVA X
kekcengsulepta 'worrisome'/*kekcengsulewum* 'worry', *kwiyemsulepta* 'lovable'/*kwiyemsulewum* 'lovability', *tahayngsulepta* 'lucky'/*tahayngsulewum* 'luckiness', *mancoksulepta* 'satisfying'/*mancoksulewum* 'satisfaction', *messulepta* 'stylish'/*messulewum* 'stylishness', *angcungsulepta* 'cute'/*angcungsulewum* 'cuteness', *calangsulepta* 'proud'/*calangsulewum* 'proudness'

Ultimately, after removing all the duplicates and noise, the *-sulep-* derivatives suggested by CLOVA X only amount to 13 items, listed in Example (13).

(13) CLOVA X suggestions of *-sulep-* derivatives
a. {*kekceng* 'worry', *kwiyem* 'love', *tahayng* 'luck', *mancok* 'satisfaction', *mes* 'style', *angcung* 'cuteness', *calang* 'pride'}*sulepta* '-like'
b. {*nekulewum* 'magnanimity', *alumtawum* 'beauty'}*sulepta* '-like'
c. {*pok* 'fortune', *salang* 'love', *cayen* 'nature', *hayngpok* 'happiness'}*sulewum* 'likeness'

Of the 13 suggestions in Example (13), only *kwiyemsulepta* ('lovable'), *nekulewumsulepta* ('magnanimous'), and *alumtawumsulepta* ('beauty-like') are unregistered words, the other 10 items being included in *Urimalsaem*. However, only *kwiyemsulepta*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

('lovable') can be confirmed as being used by Korean speakers on the web. The other two derivatives are nominalised forms of adjectives and the combination of such forms with the suffix *-sulep-* is rather awkward in Korean. In fact, the search query of *nekulewumsulepta* ('magnanimous') in browsers returned no results and the few web occurrences of *alumtawumsulepta* ('beauty-like') seem to be artificially generated text rather than examples of natural Korean usage.

This experiment of querying recommendations of unregistered *-sulep-* derivatives showed that ChatGPT performed better than CLOVA X. Further investigation is yet needed to determine whether the results would differ with prompts that include additional information regarding the combinatory constraints of the suffix *-sulep-*.

4.3 Identification and description of semantic neologisms

To overcome the bias of formal neologisms, the adequacy of ChatGPT and CLOVA X was tested by prompting the tools to recommend 10 headword candidates that qualified as semantic neologisms (i.e., words that are already in the dictionary but are actually used with different meanings). Neither LLM could provide satisfactory suggestions, so they were tested again to examine their ability to identify different meanings of a word based on the given contextual information. To this end, the LLMs were provided with examples of different usages of the well-known semantic neologism *kokwuma* ('sweet potato'), collected from newspapers and blogs, as shown in Example (14).

(14) Prompt:
These days, the word *kokwuma* 'sweet potato' is used with two different meanings. Based on the following examples where *kokwuma* 'sweet potato' is used, write the definitions for the two different senses.
Examples:
- In the past, after harvesting barley, we planted *kokwuma*.
- Boil the *kokwuma* and peel them.
- It was a well-made and interesting drama, but it had a curious balance of '*kokwuma*' and 'cider'.
(…)

(15) Generated AI answers
a. ChatGPT: Meaning of plant and food/figurative meaning (expressing frustration)
b. CLOVA X: Sweet potato as a crop produce/Sweet potato as a metaphor for a frustrating situation

As seen in Example (15), both models successfully distinguished the two meanings of *kokwuma* ('sweet potato'). However, as they also classified the examples according to the meaning of *kokwuma*, ChatGPT correctly classified the sen-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

tences under each meaning, whereas CLOVA X misclassified two sentences under the metaphorical sense instead of the literal sense and omitted one sentence. These results are based on the prompt provided in Example (16), which was given to both LLMs to generate definitions. Examples (17) to (19) compare the definitions of *kokwuma* drafted by the two LLMs and the dictionary entry written by a lexicographer in *Urimalsaem.*

(16) Prompt:
Write a dictionary entry for *kokwuma* 'sweet potato' as a headword for a Korean dictionary; make sure to include a definition and examples.

(17) Generated AI answer: ChatGPT
noun
1. [Plant] A perennial vine belonging to the morning glory family (…)
2. [Metaphorical expression] A term used metaphorically to describe a situation or person's behavior that feels frustrating and distressing. It is commonly used when a story develops slowly or a situation is not progressing.
   Examples:
   - [Plant] In the past, after harvesting barley, we planted *kokwuma*. | Boil the *kokwuma* and peel them.
     (…)
   - [Metaphorical expression] It was a well-made and interesting drama, but it had a curious balance of '*kokwuma*' and 'cider'.
     (…)

(18) Generated AI answer: CLOVA X
noun
1. [Plant] A perennial plant in the Convolvulaceae family (…).
   - In the past, after harvesting barley, we planted *kokwuma*.
   - This year, we planted three types of *kokwuma*.
2. [Crop produce] A root produce in the Convolvulaceae family (…). | Rich in carbohydrates and dietary fiber (…). | It is mainly eaten steamed or roasted (…)
   - Boil the *kokwuma* and peel them.
   - We still have leftover *kokwuma* from last year. I don't know why my mom gave us so many *kokwuma* ... I made *kokwuma* pizza to use them up.

(19) *Urimalsaem entry*
[001] [Plant] A perennial plant in the Convolvulaceae family (…)
   - The dewdrops on the *kokwuma* vines shone like beads. Yongnyeo pulled out a vine. (Han Seung-won, *Tidal Wave*)
     (…)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

[002] The tuberous root of *kokwuma*. It is commonly elongated and oval-shaped, rich in starch, and used for food and industrial purposes.
- Boiling *kokwuma*.

(…)

[003] A term used to describe a person who is inflexible and frustrating or a situation that is frustrating and not going as planned, likened to eating sweet potatoes and feeling choked.
- Recently, netizens have been using *kokwuma* to describe frustrating situations that feel stifling and *cider* to describe situations that provide a refreshing relief. (Sports Seoul, April 2016)

(…)

As can be seen in Example (17), ChatGPT described separately the literal meaning and the metaphorical meaning of *kokwuma* ('sweet potato') in its dictionary entry proposal. Moreover, it reused the examples provided in the previous prompt. CLOVA X, on the other hand, omitted the dictionary description for the metaphorical sense of *kokwuma* and only wrote the definitions of *kokwuma* as a plant and crop produce, which correspond to the entries [001] and [002] in *Urimalsaem*. Just as ChatGPT, CLOVA X reused the examples provided in the previous prompt. Both models were successful in identifying the literal and metaphorical meanings of the well-established semantic neologism *kokwuma*. However, in terms of dictionary entry compilation, ChatGPT demonstrated better performance by accurately completing the task, although its sole reliance on the input examples could be seen as both a strength and a limitation. Conversely, CLOVA X failed to describe the metaphorical meaning of *kokwuma* in its dictionary entry proposal, and was thus unsuccessful in completing the task of describing the semantic neologism as prompted.

Finally, an additional experiment was conducted where no specific information was given to the LLMs besides prompting them to describe the lexicographic information for a neologism X and then reviewing the outputs. This was unlike the previous experiments where detailed constraints and context were provided in the prompts. The additional neologisms tested included the high-frequency semantic neologism *phokphwung* ('storm') and the low-frequency semantic neologism *yengcep* ('welcome'), the news-biased neologism *caychohwan* ('Reconstruction Excess Profit Refund System'), and the non-news-biased neologism *ccwulthayng* ('tense character', used in the gaming domain). As a result of these additional experiments, CLOVA X could identify and describe the meaning of the news-biased neologism *caychohwan*, but returned inadequate results for the other neologisms. ChatGPT generated relatively appropriate definitions for the high-frequency semantic neologism and the news-biased neologism, but yielded limited results, either lacking in information or containing errors, for the low-frequency semantic neologism and the non-news-biased neologism. In particular, the models failed to describe the figurative meaning of the low-frequency semantic neologism *yengcep*, and in the case of *ccwulthayng*, the initial

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

results did not align with the intent of the query. Only after providing additional contextual information did the models present appropriate results. This highlights that LLMs are highly context-dependent and heavily rely on the diversity and accuracy of their training data. Therefore, to increase the reliability of LLMs in detecting and describing neologisms, training must incorporate up-to-date data and reflect genre diversity. Additionally, improvements are needed to enable the models to infer meanings even in situations where contextual information is not provided.

Thus far, the potential for utilising LLMs in extracting and describing Korean neologisms have been examined. The experimental results provided by ChatGPT and CLOVA X indicate that LLMs show potential as useful tools for detecting unregistered words and neologisms that are often omitted from traditional news-centred corpora. However, issues such as noise in the list of recommended neologisms, limitations in describing semantic neologisms, and duplicate outputs remain challenges to be addressed. The performance of LLMs heavily depends on the recency and linguistic diversity of their training data, necessitating ongoing efforts to train on balanced datasets and improve their algorithms.

5. Conclusion

This study analysed the current state and issues in neologism research, which represents the lexical creativity of native speakers, and examined the potential introduction of LLMs as an alternative. Two main issues in current neological research and lexicographic description are discussed: firstly, the analysis centred on online news data, which is widely used in Korea and other linguistic regions, highlighting the issue of written language bias; and secondly, the challenge of fully automatic extraction and description of semantic neologisms, which remains impossible at present. This study discussed the limitations of traditional neologism analysis, by analysing a corpus composed of sub-corpora from news data, IM data, and OP. Additionally, the study aimed to assess the potential of two major LLMs commercially available in Korea as alternatives. The findings can be summarised as follows:

Firstly, the analysis of news, IM, and OP revealed that the current methodology for neologism extraction and lexicographic description, centred on online news sources, presents limitations in identifying the first occurrence date and detecting neologisms related to everyday discourses. In particular, an analysis of the patterns of *-sulep-* ('-like') derivatives demonstrated that the same suffix led to the formation of different derivative neologisms across the three corpora. This finding highlights the need to expand neologism extraction to a wider range of genres and use domains.

Secondly, it was confirmed that the process of neologism extraction based on large corpora did not contribute to the extraction and lexicographic description of semantic neologisms. When analysing the proportion of semantic neologisms

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

in Korean neologism extraction since 1994 and examining current dictionaries, it appeared that semantic neologisms remain a blind spot of neologism research.

Thirdly, this study evaluated the ability of LLMs to detect and describe neologisms from both macrostructural and microstructural perspectives. In terms of the macrostructure, ChatGPT demonstrated the ability to complement news bias in the headword recommendation process by providing various usage examples from spoken language, messengers, blogs, and so forth. However, both models exhibited duplication and overgeneration errors to varying degrees (approximately 26% for ChatGPT, and 63% for CLOVA X), confirming that the involvement of human lexicographers is still essential. Regarding the microstructure, the ability to identify new meanings and describe them was evaluated, though it should be noted that this evaluation was conducted on a very limited sample. For the task of recommending semantic neologisms, neither of the models produced satisfactory results for open-ended queries. In situations where example sentences were provided, ChatGPT performed relatively accurately, succeeding in identifying and describing the meanings of high-frequency neologisms and news-biased neologisms. However, the performance was much weaker for low-frequency neologisms and non-news-biased neologisms.

This study differs from other studies in its discussion of the lexicographic potential of LLMs by focusing on the limitations of current neologism research and targeting a broader scope of neologisms. Overall, LLMs have shown somewhat satisfactory results in extracting and describing high-frequency neologisms and semantic neologisms, which will contribute to improving the productivity of dictionary description of neologisms in the future. The handling of neologisms that appear in genres and usage domains beyond the scope of LLM training, as well as low-frequency neologisms, remains quite insufficient though, and the lexicographic description varies significantly depending on the model. This is related to the inherent issues in the learning mechanisms of LLMs. Additionally, this study aimed to examine the correlation between the volume of Korean data used for training and the performance of neologism description through a comparison of domestic and foreign models, with results showing that there was no significant difference between them. The results of this study suggest that future research in neology studies, as well as lexicography and corpus linguistics, is needed. Discussions on how lexicography can contribute to the improvement of LLM performance are also necessary.

Endnotes

1. Naver CLOVA is reported to have trained on 6 500 times more Korean data compared to ChatGPT (https://www.donga.com/news/article/all/20230227/118100451/1).
2. Regarding the Corpus Query System, the research team in Phase 3 used the 'Neologism Investigation Program'. In Phase 4, the 'Neologism Extractor', a tool of Urimalsaem available only to expert account holders, was used. In Phase 5, the research team has been using a Python programme to extract neologisms.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

3. This table is a revised and supplemented version of Table 1 in Nam et al. (2022: 84), which adds the methodology of the neologism investigation conducted by the Language Information Research Center at KNU. The methodology follows a similar approach to the NIKL neologism research project after its termination.
4. Transcriptions of Korean follow the Yale Romanization System.
5. For the *IM Corpus*, only past conversations provided by the participants themselves were included in the corpus, and data collected by 'collection bots' or 'chatbots' were excluded to observe natural language use better. For the *Newspaper Corpus*, only a portion of the materials were included for comparison with other sources. However, if there is a need to examine the usage trends of a word over a longer time span, the entire corpus may be utilised.
6. The short-term diachronic research methodology has been described as a 'paradigm shift' in neologism research (Boussidan 2013) because previous studies on language change mainly focused on long-term trends over centuries. Although studies of language change have been conducted in traditional diachronic research, the development of corpora and the greater availability of language resources have brought more attention to short-term diachronic research. This method enables the study of the dynamics of a word or words, from the time the word or words first emerge, to the period of rapid frequency increase, the process of becoming systematised in the language, and the time when usage declines as the term fades from people's awareness. For a detailed discussion, see Boussidan (2013), Nam et al. (2019), and Lee (2024).
7. The time frame for the 2021 neologisms is 1 July 1 2020 to 30 June 2021, and for the 2022 neologisms 1 July 2021 to 30 June 2022. While Time Segment 1 partially overlaps with the investigation time frame for the 2020 neologisms, the corpus used in this study did not reveal any instances of 2020 neologisms appearing earlier than the investigation period. Therefore, only the 2021 and 2022 neologisms are considered here.
8. Words that also appeared in the *Newspaper Corpus* during the same time period are excluded.
9. Out-links redirect users to the homepages of news outlets and have different metadata structures depending on the content producer. Conversely, in-links connect to pages within Naver New, so their format, converted to Naver News' metadata structure, offers the advantage of being much simpler and more stable for data processing.
10. According to Nam et al. (2024), who examined the appearance trends of the 2023 neologisms collected from news sources in news and blogs, more than three quarters of the neologisms in the study appeared first in blogs and then in news sources.
11. These are the types that did not appear in newspapers but appeared in IM or OP. A total of 21 types that did not appear in newspapers were found in both IM and OP.
12. All experiments in this study were conducted in June 2024 and also involved Google's Gemini, in parallel to ChatGPT and CLOVA X. However, despite using the same prompts, Gemini frequently failed to properly address the requests or suggested unnatural vocabulary items that are unlikely to be created by native Korean speakers. Given the limitations of space and the need for concise discussion, the results from Gemini experiments are omitted from this paper.
13. Naver CLOVA X is a large-scale language model (LLM) AI released by Naver on 24 August 2023, which claims to have trained on more than 6 500 times more Korean data than ChatGPT at the time of development (February 2023). The comparison seems to have been made with GPT-3.5, the version preceding GPT-4, which is used in this study. Specific data on how much Korean data GPT-4 was trained on, compared to GPT-3.5, has not been disclosed.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

14. https://www.naver.com/
15. https://www.daum.net/
16. https://ko.dict.naver.com/#/main
17. As a reminder, the answers generated by AI are variable depending on the prompt. The prompts used in this study were revised and refined several times to obtain answers suitable for the research objectives.
18. Only the forms *-sulepta* (basic form) and *-sulewun* (conjugated form) were searched.
19. Duplicates were also found in ChatGPT results, although less frequently; this suggests that duplicates are a common problem in AI-generated results and not specific to CLOVA X. However, in the experiments for this study, CLOVA X produced significantly more duplicate words than ChatGPT.

References

Dictionaries

Korea University Korean Dictionary. https://ko.dict.naver.com/#/main
Standard Korean Language Dictionary. https://stdict.korean.go.kr/main/main.do
Urimalsaem. https://opendict.korean.go.kr/main

Neologism reports

Kim, H. et al. 2005. *Neologisms of 2005*. Seoul: National Institute of Korean Language.
Lee, S. et al. 2010. *Neologisms of 2010.* Seoul: National Institute of Korean Language.
Nam, K. et al. 2012. *Neologisms of 2012*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2013. *Neologisms of 2013*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2014. *Neologisms of 2014*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2015. *Neologisms of 2015*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2016. *Neologisms of 2016*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2017. *Neologisms of 2017*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2018. *Neologisms of 2018*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2019. *Neologisms of 2019*. Seoul: National Institute of Korean Language.
Nam, K. et al. 2021. *Neologisms of 2020. A New Language of the COVID-19 Pandemic.* Seoul: Hankwukmwunhwasa.
Nam, K. et al. 2022. *Neologisms of 2021. COVID-19: Are We Going Back to Normal?* Seoul: Hankwukmwunhwasa.
Nam, K. et al. 2023. *Neologisms of 2022. What New Words Have Koreans Created and Used in this Era of Generative AI?* Seoul: Hankwukmwunhwasa.
National Institute of Korean Language. 1994. *A Study of Neologisms*. Seoul: National Institute of Korean Language.
National Institute of Korean Language. 1995. *A Study of Neologisms*. Seoul: National Institute of Korean Language.
National Institute of Korean Language. 1996. *A Study of Neologisms*. Seoul: National Institute of Korean Language.
Park, Y. et al. 2000. *Neologisms of 2000*. Seoul: National Institute of Korean Language.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Park, Y. et al. 2001. *Neologisms of 2001*. Seoul: National Institute of Korean Language.
Park, Y. et al. 2002. *Neologisms of 2002*. Seoul: National Institute of Korean Language.
Park, Y. et al. 2003. *Neologisms of 2003*. Seoul: National Institute of Korean Language.
Park, Y. et al. 2004. *Neologisms of 2004*. Seoul: National Institute of Korean Language.
Soh, K. et al. 2009. *Neologisms of 2008*. Seoul: National Institute of Korean Language.
Soh, K. et al. 2010. *Neologisms of 2009*. Seoul: National Institute of Korean Language.

Other literature

Boussidan, A. 2013. *Dynamics of Semantic Change: Detecting, Analyzing and Modeling Semantic Change in a Corpus of Short Diachrony.* Unpublished PhD Dissertation. Lyon: Université de Lyon.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

De Schryver, G.-M. 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.

Freixa, J. and A. Adelstein. 2013. Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Martínez, M.J., J.A. Hinojosa and F.J. Ruiz. (Eds.). 2013. *Proceedings of the Congreso Internacional El Diccionario: Neología, lenguaje de especialidad, computación, Mexico City, Mexico, 28–30 October 2013*: 1-12. Mexico: El Colegio de México.

Huang, Y. and K. Nam. 2023. Unregistered Words in Everyday Language and a Proposal for Their Optimal Lexicographic Microstructure. *Lexicography* 10(2): 94-116.

Klosa, A. and H. Lüngen. 2018. New German Words: Detection and Description. Čibej, J., V. Gorjanc, I. Kosem and S. Krek (Eds.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, EURALEX 2018, Ljubljana, Slovenia, 17–21 July 2018*: 559-569. Ljubljana: Ljubljana University Press.

Lee, S. 2022. A Study of 2015–2020 Neologisms in Korean Language — Focusing on Neologism of 'Society', 'Economy', and 'Politics/Policies'. *Eo Mun Lon chong (Korean Language and Literature)* 94: 99-131.

Lee, S. 2024. *A Corpus Linguistic Study of Korean Neologisms from a Short-term Diachronic Perspective.* Unpublished PhD Dissertation. Daegu: Kyungpook National University.

Lew, R. 2023. ChatGPT as a COBUILD Lexicographer. *Humanities and Social Sciences Communications* 10(1): 1-10.

Lew, R. 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11(1): 1-8.

McKean, E. and W. Fitzgerald. 2023. The ROI of AI in Lexicography. *Proceedings of the 16th International Conference of the Asian Association for Lexicography: Lexicography, Artificial Intelligence, and Dictionary Users, ASIALEX 2023, Seoul, South Korea, 22–24 June 2023:* 18-27. Seoul: Yonsei University.

Nam, K. 2020. Issues of Frequency in Korean Neologisms. *Korean Semantics* 68: 213-239.

Nam, K. 2021. Overcoming Issues in Frequency-based Extraction and Lexicographic Inclusion of Korean Neologisms: A Triangulation Approach. *Lexicography* 8(1): 3-31.

Nam, K. and J. An. 2023. A Lexicographical Study on Phrase-unit Serialised Neologisms [N + Seday 'Generation'] — From 'Computer Generation' to 'Corona Generation' to 'MZ Generation'. *Journal of Korealex* 42: 33-66.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2045 (Lexicofocus)

Nam, K., S. Lee and J. Choi. 2018. Research Trends and Issues on Semantic Neology Using Web Corpus. *Journal of Korealex* 31: 55-84.

Nam, K., S. Lee and H.Y. Jung. 2020. The Korean Neologism Investigation Project: Current Status and Key Issues. *Dictionaries. Journal of the Dictionary Society of North America* 41(1): 105-129.

Nam, K., S. Lee and B. Kang. 2019. The Linguistics of 'Frame' in Korean. *Korean Semantics* 65: 135-163.

Nam, K., H. Song, J. An and B. Kang. 2024. Where Do Neologisms Originate and Spread? — A Critical Review of Neologism Extraction Focused on News Media. *Language Facts and Perspectives* 63: 5-35.

Nam, K., H. Song, J. Choi and S. Lee. 2022. *A Study on Modern Neologism*. Seoul: Hankukmunhwasa.

Renouf, A. 2013. A Finer Definition of Neology in English: The Life-cycle of a Word. Hasselgård, H., J. Ebeling and S. Oksefjell Ebeling (Eds.). 2013. *Corpus Perspectives on Patterns of Lexis*: 177-208. Amsterdam: John Benjamins.

Sinclair, J.M. 2004. *Trust the Text: Language, Corpus and Discourse.* London: Routledge.

Smyk-Bhattacharjee, D. 2009. *Lexical Innovation on the Internet — Neologisms in Blogs*. Unpublished PhD Dissertation. Zurich: University of Zurich.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Terminologiebestuur binne die Suid-Afrikaanse konteks

Michelle Goosen, *Departement Afrikatale, Universiteit van Pretoria, Pretoria, Suid-Afrika (michelle.goosen@up.ac.za) (https://orcid.org/0000-0001-7465-9113)*
en
Elsabé Taljard, *Departement Afrikatale, Universiteit van Pretoria, Pretoria, Suid-Afrika (elsabe.taljard@up.ac.za) (https://orcid.org/0000-0002-4507-1633)*

Opsomming: Hierdie artikel ondersoek die sistematiese bestuur van terminologiese inligting in Suid-Afrika deur middel van terminologiebestuursisteme. Daar word argumenteer vir 'n gestruktureerde benadering tot terminologiebestuur binne verskillende kontekste, naamlik op makro-, meso- en mikrovlak. Daar word spesifiek gefokus op terminologiebestuur binne die kommersiële konteks wat binne 'n organisasie produktiwiteit kan verhoog, doeltreffende kommunikasie kan verseker en opbrengs op belegging kan verbeter. Die belang van gebruikersgeoriënteerde terminologieprodukte word beklemtoon waar daar spesifiek aandag aan die profiel van die teikengebruiker gegee word. Verskeie tipes sagteware vir terminologiebestuur, d.i. losstaande, internetgebaseerde en mobiele toepassings, word bespreek. Hierdie bespreking sluit ook die kennisgebaseerde benadering tot terminologie in. Die artikel stel 'n praktiese terminologiewerkvloeidiagram voor wat geskik is vir die veeltalige Suid-Afrikaanse konteks. Dit wys op die noodsaaklikheid van verdere navorsing, institusionele ondersteuning en gebruikersbetrokkenheid om 'n volhoubare terminologie-infrastruktuur in Suid-Afrika te vestig.

Sleutelwoorde: TERMINOLOGIEBESTUUR, TERMINOLOGIEBESTUURSISTEME, TERMINOLOGIEPRODUKTE, KOMMERSIËLE KONTEKS, OPBRENGS OP BELEGGING, MAKRO-, MESO EN MIKROVLAK, TEIKENGEBRUIKER, SAGTEWARE, KENNISGEBASEERDE BENADERING TOT TERMINOLOGIE, TERMINOLOGIEWERKVLOEIDIAGRAM

Abstract: Terminology Management in the South African Context. This article examines the systematic management of terminological information in South Africa through the use of terminology management systems. It argues for a structured approach to terminology management across different contexts, namely at macro, meso, and micro level. The focus is specifically on terminology management within the commercial context, where it can enhance productivity within an organisation, ensure effective communication, and improve return on investment. The importance of user-oriented terminological products is emphasised, with specific attention given to the profile of the target user. Various types of software for terminology management, namely standalone, internet-based, and mobile applications, are discussed. This discussion also includes the knowledge-based approach to terminology. The article proposes a practical terminology workflow diagram suited to the multilingual South African context. It highlights the necessity of further research, institutional support, and user involvement to establish a sustainable terminology infra-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

structure in South Africa.

Keywords: TERMINOLOGY MANAGEMENT, TERMINOLOGY MANAGEMENT SYSTEMS, TERMINOLOGY PRODUCTS, COMMERCIAL CONTEXT, RETURN ON INVESTMENT, MACRO, MESO AND MICRO LEVEL, TARGET USER, SOFTWARE, KNOWLEDGE-BASED APPROACH TO TERMINOLOGY, TERMINOLOGY WORKFLOW DIAGRAM

1. Inleiding

Terminologiebestuur kan gedefinieer word as die sistematiese proses waartydens terminologiese inligting versamel, geprosesseer, beskryf en aangebied word. Hierdie aktiwiteite word hoofsaaklik aan die hand van 'n terminologiebestuursisteem (Eng: *Terminology Management System, TMS*) uitgevoer (Steurs et al. 2015: 224, Warburton 2015). Na gelang van die terminologiebestuursprogram word die terminologiese inligting in die vorm van byvoorbeeld, glossariums, termlyste, terminologiese databasisse of terminologiese kennisbasisse aangebied. Die terminologiese inligting word uiteindelik deur die teikengebruiker van 'n terminologieproduk d.m.v. 'n standaard webblaaier en/of 'n toepassing (Eng: *application*) geraadpleeg. Hoewel terminologiebestuur binne 'n eentalige konteks kan geskied, is dit algemene praktyk om terminologiese inligting — veral binne 'n veeltalige bestel — vanuit die brontaal na een of meer teikentale te vertaal. Terminologiewerk en vertaling word derhalwe as 'n geïntegreerde praktyk beskryf. Vanweë die noue verband tussen terminologie en vertaling word aspekte wat met terminologiebestuur verband hou, dikwels oor die hoof gesien. Bowker (2020: 262) meld:

> [Terminology management] is often hidden or blended with other tasks ... it might fall under another heading, it might be carried out by someone with a different job title, and it might not be billed separately but instead be included in the overall price of translation or revision.

Dit is daarom dikwels moeilik om vas te stel hoeveel tyd spesifiek aan aspekte wat met terminologiebestuur verband hou, bestee word, of hoeveel opbrengs op belegging (Eng: *return on investment*) deur terminologiebestuur gegenereer word. Werkvloei binne terminologiebestuur kan onder andere bepaal deur wié terminologiese take uitgevoer word en wát hierdie take behels. Sodoende kan daar ook vasgestel word watter persentasie van inkomste deur terminologiebestuur opgelewer word. Binne die literatuur vir terminologie is daar egter weinig inligting oor terminologiewerkvloei en word dit dus nie noodwendig in terminologiebestuur toegepas nie (Warburton 2015: 386, Bowker 2020: 274).

Net soos leksikografie is terminologie 'n gebruikersgeoriënteerde dissipline en as sulks moet terminologieprodukte volgens die profiel van die teikengebruiker saamgestel word en uiteindelik aan die behoeftes van dié gebruiker voldoen. Dit blyk egter dat die profiel van die teikengebruiker van 'n terminologieproduk selde binne terminologiebestuur in ag geneem word en terminolo-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

giese hulpbronne word derhalwe dikwels as onvoldoende beskou (Durán-Muñoz 2012: 78-79, Bowker 2020: 274).

Die doel van hierdie artikel is om in die eerste plek onderskeid te tref tussen die verskillende kontekste waarbinne terminologiebestuur uitgevoer kan word. In hierdie bespreking word daar gefokus op terminologiebestuur binne die kommersiële konteks. Om voorsiening te maak vir terminologiebestuur wat binne verskillende kontekste vir verskillende doeleindes uitgevoer word, word daar in die tweede plek voorgestel dat terminologiebestuur op makro-, meso- of mikrovlak gedoen kan word. In die derde plek word belangrike vrae gestel oor die profiel van die potensiële teikengebruiker van 'n terminologieproduk. Vanweë die gebrek aan terminologiese databasisse in Suid-Afrika, bestaan die moontlikheid dat (bestaande) Suid-Afrikaanse terminologiebestuursprogramme nie noodwendig terminologiebestuursisteme vir terminologiebestuur gebruik nie. Daar word dus vierdens onderskeid getref tussen die verskillende tipes beskikbare sagteware vir terminologiebestuur. Hierdie oorsig sluit ook die kennisgebaseerde benadering tot terminologie in — 'n benadering wat nog nie binne die breër konteks van terminologie in Suid-Afrika geïmplementeer is nie. Om die verskillende aktiwiteite binne terminologiebestuur te optimaliseer, word daar laastens 'n terminologiewerkvloeidiagram vir die Suid-Afrikaanse konteks voorgestel.

2. Terminologiebestuur binne verskillende kontekste

Terminologiebestuur kan binne verskillende kontekste vir verskillende doeleindes uitgevoer word. Terminologiebestuur as deel van taalbeplanning word byvoorbeeld gedoen om aan die taalbehoeftes van taalgemeenskappe te voldoen wat op sosiale, kulturele, ekonomiese, opvoedkundige én politiese vlak voorkom. Terminologiebestuur vir akademiese doeleindes bevorder op sy beurt die teikengebruiker van 'n terminologieproduk se geletterdheidsvlakke en maak voorsiening vir gespesialiseerde kommunikasie binne 'n bepaalde vakgebied. Daar is egter 'n groeiende behoefte onder veeltalige taalgemeenskappe aan terminologiebestuur buite hierdie twee kontekste. Kundiges verwys na terminologiebestuur buite taalbeplanning en die akademiese omgewing as terminologiebestuur binne kommersiële omgewings (Warburton 2015, Warburton 2024, Bowker 2020). Binne die literatuur vir terminologie in Suid-Afrika is daar tot dusver weinig navorsing oor terminologiebestuur in kommersiële omgewings beskikbaar. Daar word derhalwe in die hieropvolgende bespreking op terminologiebestuur binne die kommersiële konteks gefokus.

Terminologiebestuur binne die kommersiële konteks

Terminologiebestuur binne 'n kommersiële konteks word eerstens gedoen om aan die praktiese, gespesialiseerde kommunikatiewe behoeftes van 'n organisasie op sowel interne as eksterne vlak te voldoen. Hierdie terminologiese inligting fokus op 'n unieke handelsmerk, diens of produk wat in byvoorbeeld bemar-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

kingsveldtogte, kliëntediens en produkinligting, byvoorbeeld 'n gebruikershandleiding, gebruik word. Binne hierdie konteks verwys kommunikatiewe behoeftes op interne vlak na gespesialiseerde kommunikasie wat binne 'n gegewe organisasie plaasvind. Aan die ander kant verwys kommunikatiewe behoeftes op eksterne vlak na gespesialiseerde kommunikasie tussen 'n organisasie en die kliënt van 'n gegewe handelsmerk, diens of produk. In praktiese terme sou dit beteken dat 'n kommersiële organisasie soos byvoorbeeld 'n bank, 'n mediesefondsskema of 'n selfoonmaatskappy terminologiese inligting op sistematiese wyse versamel, prosesseer en dissemineer.

'n Volgende belangrike rede waarom terminologiebestuur binne 'n kommersiële konteks uitgevoer word, is opbrengs op belegging. Vanuit 'n besigheidsperspektief bespaar terminologiebestuur kostes en tyd en bevorder produktiwiteit. Dit kan, aldus Allard (2012: 38-39), soos volg opgesom word:

— Dit bevorder die korrekte gebruik van terme.
— Dit bevorder die kwaliteit van die finale weergawes van tekste soos byvoorbeeld installeringshandleidings, gebruikershandleidings en lyste wat algemene vrae en hulle antwoorde bevat (Eng: *Frequently Asked Questions*, afgekort as FAQ) wat binne 'n organisasie geskep word.
— Dit verminder tyd en moeite wat bestee moet word aan verbeteringe wat aangebring moet word aan byvoorbeeld tekste wat geskep is.
— Dit versterk 'n organisasie se handelsmerk en geloofwaardigheid wat verseker dat konsekwente handelsmerkboodskappe aan kliënte oorgedra word.
— Dit vergemaklik die proses van gehaltebeheer wat op byvoorbeeld produkte gedoen moet word.
— Dit bevorder doeltreffende kommunikasie op interne en eksterne vlak tussen spesialiste, semi-spesialiste en leke.
— Binne kliëntediens verminder terminologiebestuur die aantal oproepe wat deur kliënte aan 'n organisasie gemaak word. Deur die korrekte terminologie te gebruik word enige dubbelsinnigheid uit die weg geruim en verstaan kliënte hoe 'n bepaalde produk of diens werk, sonder enige addisionele ondersteuning van 'n organisasie se kliëntediens. Indien 'n gebruikershandleiding na byvoorbeeld die *aan/af-knoppie* én die *aan/af-skakelaar* vir dieselfde konsep verwys, kan dit verwarring veroorsaak en is die kliënt genoodsaak om die kliëntediens te skakel.
— Dit verminder die risiko van produkmislukking — en potensiële aanspreeklikhede wat met produkmislukking verband kan hou — wat deur verkeerde, dubbelsinnige of inkonsekwente terminologie veroorsaak kon word.

Die opbrengs op belegging word afsonderlik binne 'n gegewe kommersiële konteks gemeet, aangesien die winste deur 'n spesifieke produksieomgewing bepaal word (Warburton 2014: 649). Daar is dus nie 'n spesifieke formule om die opbrengs op belegging te bepaal wat binne enige kommersiële konteks van terminologiebestuur toegepas kan word nie. Daar is wel faktore wat dit beïnvloed. Hieronder word drie faktore uitgelig, soos deur Warburton (2024: 9) geïdentifiseer.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Globale verspreidingskanale en verskillende vorme van media, byvoorbeeld webblaaie, mobiele toepassings en sosiale media, wat gebruik word vir die beskikbaarstelling van terminologie verseker dat kliënte van verskillende agtergronde 'n sekere produk of diens verstaan en doeltreffend kan gebruik. 'n Volgende belangrike aspek is sowel die hoeveelheid inhoud wat geproduseer en vertaal word as die getal teikentale waarin die inhoud vertaal word. Korrekte (en geverifieerde) terme kan weer gebruik word wat duplisering verminder en verseker dat die korrekte terme in die teikentale vertaal word. Laastens het die gebruik van rekenaarprogrammatuur soos terminologiebestuur- en vertaalsisteme ook 'n uitwerking op die opbrengs op belegging binne die kommersiële konteks. 'n Terminologiebestuursisteem stel die terminoloog in staat om 'n groot hoeveelheid terminologiese inligting in verskillende tale tegelykertyd te bestuur. Sekere vertaalsisteme voer die vertaalproses semi-outomaties uit. Dit beteken dat die vertaalproses aansienlik vinniger uitgevoer kan word en 'n produk of diens kan binne die gegewe tyd beskikbaar gestel word. Taalbehoeftes van taalgemeenskappe wat op sosiale vlak voorkom kan deur terminologiebestuur binne die kommersiële konteks vervul word. Met betrekking tot die Suid-Afrikaanse bestel is daar egter 'n groot leemte in terminologiebestuur binne die kommersiële konteks. Afgesien van die ooglopende redes waarom terminologiebestuur binne dié konteks nie noodwendig uitgevoer word nie, soos byvoorbeeld onkundigheid oor wat terminologiebestuur is, die onderskatting van die waarde van terminologiebestuur en onwilligheid of huiwering om terminologiebestuur te implementeer (Sauberer 2018: 10), is daar 'n gebrek in Suid-Afrika aan die nodige gespesialiseerde kundigheid om terminologiebestuur binne die kommersiële konteks te bedryf. Die Suid-Afrikaanse banksektor is 'n goeie voorbeeld in dié verband. Hierdie sektor bedien 'n aansienlike deel van die (veeltalige) Suid-Afrikaanse gemeenskap — volgens *The Banking Association South Africa* se 2021-verslag het 85% van volwassenes in daardie jaar toegang tot 'n bankrekening gehad. Gegewe die sentrale rol wat toegang tot hoë kwaliteit finansiële inligting vir die deursnee Suid-Afrikaanse burger het, sou mens 'n besondere poging van die kant van finansiële instellings verwag om hierdie inligting in kliënte se sterkste taal aan hulle beskikbaar te stel. Dit sluit in toegang tot korrekte en konsekwente gebruik van relevante terminologie. Daar is egter geen blyke van sodanige pogings nie. Vergelyk in hierdie verband die top vyf finansiële instellings in Suid-Afrika (Evans 2024). *Standard Bank Group*, *Absa*, *FirstRand*, *Nedbank* en *Investec* se terminologiese inligting word slegs in Engels gelys en is beperk tot 'n spesifieke aspek vir 'n spesifieke teikengebruiker binne 'n gegewe finansiële instelling[1]. *Standard Bank Group* se glossarium hou byvoorbeeld verband met buitelandse valuta terwyl *Absa* se glossarium op beleggingsbestuur fokus.

Terminologiebestuur binne verskillende raamwerke

Terminologie word tans deur sowel staatsinstellings as privaatinstellings op 'n sistematiese of ad hoc-basis binne formele én informele raamwerke uitgevoer. Sommige terminologieontwikkelingsinisiatiewe is losstaande projekte, terwyl

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

ander projekte deel uitmaak van geïntegreerde terminologieontwikkelingsinisiatiewe. Vanweë die gefragmenteerde aard van dié terminologiebestuursprogramme, is dit moeilik om 'n akkurate oorsig te gee van die werk wat alreeds gedoen is (Antia 2015: 470). Binne die literatuur vir terminologiebestuur word daar tans nie onderskeid tussen terminologiebestuur binne verskillende kontekste vir verskillende doeleindes getref nie. Vir die doel van hierdie artikel word daar wel onderskei tussen terminologiebestuur op makro-, meso- en mikrovlak. Hierdie onderskeid word vervolgens kortliks bespreek en aan die hand van enkele voorbeelde binne die Suid-Afrikaanse konteks geïllustreer.

Makrovlakterminologiebestuur

Makrovlakterminologiebestuur word op nasionale vlak uitgevoer. Terminologiebestuur op dié vlak word binne die konteks van taalbeplanning gedoen en word nie noodwendig vir die doel van opbrengs op belegging uitgevoer nie. Binne terminologiebestuur op makrovlak word die taalbehoeftes van die teikengebruikers van terminologieprodukte hoofsaaklik deur 'n regering geïdentifiseer en vervul. In Suid-Afrika is die Departement van Sport, Kuns en Kultuur (https://www.dsac.gov.za/) verantwoordelik vir terminologiebestuur op makrovlak. Daar word ten doel gestel om terminologiese inligting soos terme, verklarende inligting, gebruiksvoorbeelde en bronpublikasies in verskeie tale in die vorm van termlyste aan die teikengebruikers van terminologieprodukte beskikbaar te stel[2].

Mesovlakterminologiebestuur

Mesovlakterminologiebestuur word deur staatsondernemings onderneem en kan in samewerking met ander gesaghebbende liggame, soos byvoorbeeld universiteite, uitgevoer word. Terminologiebestuur deur staatsondernemings (Eng: *state-owned enterprises*) en ander openbare entiteite en terminologiebestuur vir akademiese doeleindes word ook op mesovlak uitgevoer. Hoewel daar aanvaar word dat 'n regering en sy departemente terminologiese databasisse of terminologiese kennisbasisse aan die publiek beskikbaar moet stel (Warburton 2024), kan staatsondernemings en publieke entiteite terminologiebestuur op dié vlak doen om terminologiese inligting intern te gebruik en/of om hierdie inligting aan die publiek beskikbaar te stel om in spesifieke taalbehoeftes van taalgemeenskappe te voorsien. Nie alle Suid-Afrikaanse staatsondernemings en publieke entiteite wat terminologiebestuur toepas se terminologiese inligting is geredelik beskikbaar nie. Die Suid-Afrikaanse Inkomstediens (SAID) (Eng: *South African Revenue Service* (SARS)) is een van die enkele staatsondernemings wie se terminologiese inligting deur die publiek geraadpleeg kan word. Die SAID stel etlike terminologieprodukte aan die teikengebruikers, d.i. belastingbetalers, beskikbaar. Die teikengebruikers het toegang tot 'n Engelse glossarium wat sekere terme bevat wat op die SAID se webtuiste gebruik word. Die teikengebruikers kan ook 'n veeltalige termlys raadpleeg wat 450 finansiële terme in die 11 amptelike tale bevat wat 'n wye spektrum van finansiële terminologie dek. Die veeltalige

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

termlys is saamgestel met die oog daarop om die tale wat hulpbronarm is, te bevorder, en aan die teikengebruikers van hierdie terminologieproduk se spesifieke taalbehoeftes wat op sosiale vlak voorkom, te voldoen.

Mikrovlakterminologiebestuur

Terminologiebestuur op mikrovlak word hoofsaaklik binne 'n kommersiële konteks uitgevoer waar die terminologiese inligting op 'n unieke handelsmerk, diens of produk fokus. Terminologiebestuur op dié vlak word gedoen om gespesialiseerde kommunikasie binne 'n maatskappy of organisasie en tussen 'n gegewe maatskappy of organisasie en sy kliënte te bevorder. Mediese skemas, byvoorbeeld, stel op verskillende maniere terminologiese inligting aan die teikengebruikers beskikbaar. Vergelyk in hierdie verband *Discovery Medical Scheme* en *KeyHealth Medical Scheme*. *Discovery Medical Scheme* se hoofbrosjure lys sekere sleutelterme en verklarende inligting wat in dié brosjure gebruik word. Dit stel ook'n lys met terminologiese inligting op hulle webtuiste aan teikengebruikers beskikbaar. *Keyhealth Medical Scheme* se brosjures bevat elk 'n glossarium wat aangevul word deur 'n addisionele glossarium wat op dié skema se webtuiste beskikbaar is. Die terminologieprodukte wat deur hierdie twee mediese skemas beskikbaar gestel is, is slegs in Engels beskikbaar. Dit val dus te betwyfel of daar aan die taalbehoeftes van die teikengebruikers voldoen word wie se sterkste taal nie Engels is nie.

3. Profiel van die teikengebruiker van 'n terminologieproduk

Binne die leksikografie is omvattende navorsing oor die profiel van 'n potensiële teikengebruiker van sowel gedrukte as aanlyn leksikografiese produkte reeds gedoen, sien byvoorbeeld Bergenholtz en Tarp 1995, Tarp 2004, Gouws en Prinsloo 2005, Bergenholtz et al. 2009, De Schryver 2009, Bothma et al. 2016. Binne die funksionele benadering tot leksikografie word leksikografiese produkte volgens die profiel van 'n spesifieke teikengebruiker saamgestel. Aspekte soos die taalvaardighede, taalbehoeftes, ensiklopediese kennis, naslaanvaardigheid én rekenaargeletterdheidsvlakke van die potensiële teikengebruiker het 'n definitiewe en 'n uiteenlopende invloed op die leksikografiese proses van enige gegewe leksikografiese produk (Bergenholtz en Tarp 1995: 20-21, Gouws en Prinsloo 2005: 25).

'n Terminologieproduk moet ook volgens die profiel van die potensiële teikengebruiker saamgestel word. Daar word egter weinig navorsing oor die profiel van die potensiële teikengebruiker van terminologieprodukte gedoen. Navorsing wat wel uitgevoer is, het bevind dat bestaande terminologieprodukte nie noodwendig vir 'n spesifieke teikengebruiker ontwikkel is nie en dus nie aan die behoeftes van die teikengebruiker van sodanige produkte voldoen nie: "... user studies in terminology are relatively rare and those that exist have uncovered various types of dissatisfaction with existing terminology resources" (Bowker 2020: 274). Voorts toon dié studies dat teikengebruikers 'n behoefte aan veeltalige

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

terminologiese hulpbronne het, dat huidige terminologieprodukte nie oor gevorderde soekfunksies beskik om die teikengebruiker te help om byvoorbeeld gevorderde soektogte binne 'n terminologiese databasis of terminologiese kennisbasis uit te voer nie en dat daar 'n gebrek is aan 'n funksie waar die teikengebruiker terugvoering oor die terminologiese inligting kan gee (Durán-Muñoz 2010, Bowker 2020: 274). Dieselfde aspekte wat van belang is vir die konseptualisering van 'n algemene leksikografiese hulpbron of produk moet ook in ag geneem word in die konseptualisering van 'n terminologieproduk.

Belangrike vrae wat beantwoord moet word, sluit in: Wat is die teikengebruiker se sterkste taal? Watter (vak)gebiede se terminologie moet in die teikengebruiker se sterkste taal ontwikkel word? Moet die terminologie vir die doeleindes van dekodering of enkodering ontwikkel word? Watter terminologiese inligting moet ingesluit word? Is die teikengebruiker 'n vakspesialis, 'n semispesialis of 'n leek? Sal die teikengebruiker die gebruikerskoppelvlak kan gebruik? Indien nie, hoe kan hierdie behoefte aangespreek word? Hierdie vrae kan uitgebrei en aangepas word na gelang van die ontwikkeling van 'n terminologieproduk vir 'n spesifieke teikengebruiker.

4. Terminologiebestuursisteme

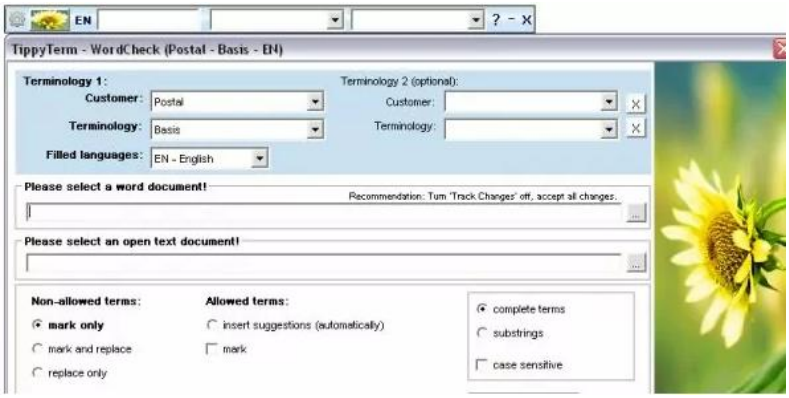
Die vroegste dokumentering van terminologie in Suid-Afrika is so vroeg as die 1950's op indekskaarte gedoen (Alberts 2010: 609). Hoewel dit in daardie tydstip die mees praktiese werksmetode was, het indekskaarte weggeraak en mettertyd onduidelik/onleesbaar geraak. Meer as 'n halfeeu later word terminologiebestuursisteme vir terminologiebestuur gebruik.

Terminologiebestuursisteme kan opgedeel word in die sisteemagterkant (Eng: *back-end*) en die sisteemvoorkant (Eng: *front-end*). Die versameling, prosessering, integrering en deurlopende bywerking van terminologiese inligting maak deel uit van die sisteemagterkant. Aan die voorkant word die terminologiese inligting deur die teikengebruiker d.m.v. 'n standaard webblaaier of 'n toepassing (Eng: *application*) geraadpleeg.

Binne terminologiebestuursisteme word daar onderskeid getref tussen losstaande en internetgebaseerde sagteware, en mobiele terminologiese toepassings. Laasgenoemde staan ook bekend as *m-terminologie* of *TMS applications*. Hierdie drie tipes sagteware word vervolgens bespreek.

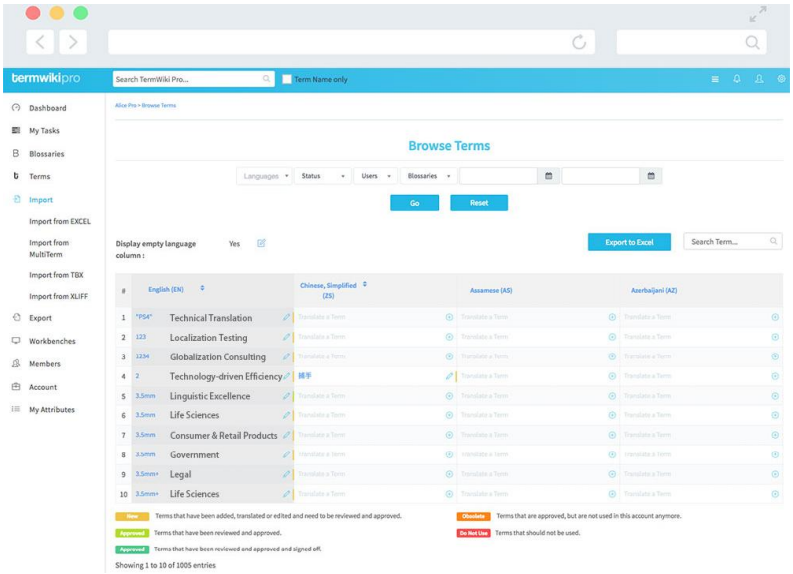
Losstaande sagteware, soos byvoorbeeld *AnyLexic*, *Lingo 4.0* en *TippyTerm* word op 'n rekenaar of 'n skootrekenaar geïnstalleer en werk onafhanklik van enige ander sisteem, d.i. 'n ander elektroniese toestel of 'n internetverbinding. Hierdie tipe sagteware bied slegs toegang aan een gebruiker op 'n keer tot 'n gegewe terminologiese databasis. Gevolglik kan slegs een gebruiker daarop werk. Hoewel die gebruiker altyd toegang tot sy/haar terminologiese databasis sal hê, is hy/sy self verantwoordelik vir rugsteunkopieë van die terminologiese inligting wat in sodanige databasis opgeneem word. Losstaande sagteware word hoofsaaklik deur individue soos tolke en vertalers asook binne kleinskaalse terminologieprojekte gebruik.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)



Figuur 1: 'n Gedeelte van die sisteemagterkant van *TippyTerm*

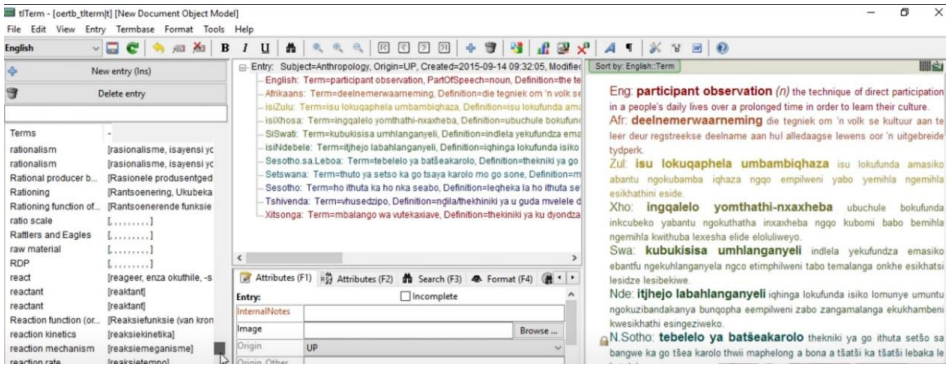
Die eerste tipe internetgebaseerde sagteware, soos byvoorbeeld *Evoterm*, *TermWeb*, *TermWiki Pro*, *QTerm* en *WebTerm* bied die gebruiker toegang tot die sagteware d.m.v. 'n internetverbinding en 'n webblaaier. Die gebruiker kan enige tyd, op enige plek toegang tot die sagteware verkry, mits hulle 'n betroubare internet-verbinding gebruik. Aangesien die data op 'n sentrale bediener (Eng: *server*) gestoor word, is die gebruiker nie verantwoordelik vir enige rugsteunkopieë nie. *Autshumato Terminology Management System*, 'n terminologiebestuursisteem wat plaaslik ontwikkel is, is ook 'n voorbeeld van internetgebaseerde sagteware wat platformonafhanklik is.



Figuur 2: 'n Gedeelte van die sisteemagterkant van *TermWikiPro*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

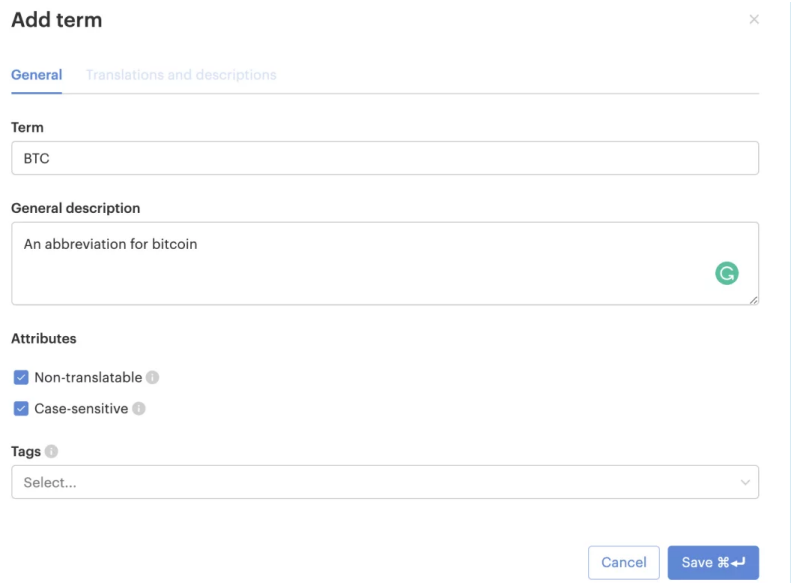
Die tweede tipe internetgebaseerde sagteware, soos byvoorbeeld *Acrolinx*, *tlTerm*[3] en *XBENCH.NET* vereis dat die sagteware-pakket op 'n rekenaar of 'n skootrekenaar geïnstalleer word en d.m.v. 'n internetverbinding gebruik word. Die gebruiker kan enige tyd, op enige plek toegang tot die sagteware verkry, mits die sagteware op 'n bepaalde werkstasie geïnstalleer is en die gebruiker 'n betroubare internetverbinding gebruik. Die data word op 'n sentrale bediener gestoor en is die gebruiker nie verantwoordelik vir enige rugsteunkopieë nie. Internetgebaseerde sagteware bied aan 'n hele aantal gebruikers toegang tot 'n betrokke sagtewarepakket. Dié tipe sagteware word vir medium tot groot terminologiebestuursprogramme aanbeveel. Internetgebaseerde sagteware word in vergelyking met losstaande sagteware as meer gesofistikeerd beskou, aangesien internetgebaseerde sagteware dikwels meer gevorderde funksies en kenmerke insluit. Laasgenoemde tipe sagteware kan ook binne verskillende kontekste van terminologiebestuur gebruik word (Costa et al. 2017: 72).



Figuur 3: 'n Gedeelte van die sisteemagterkant van *tlTerm*

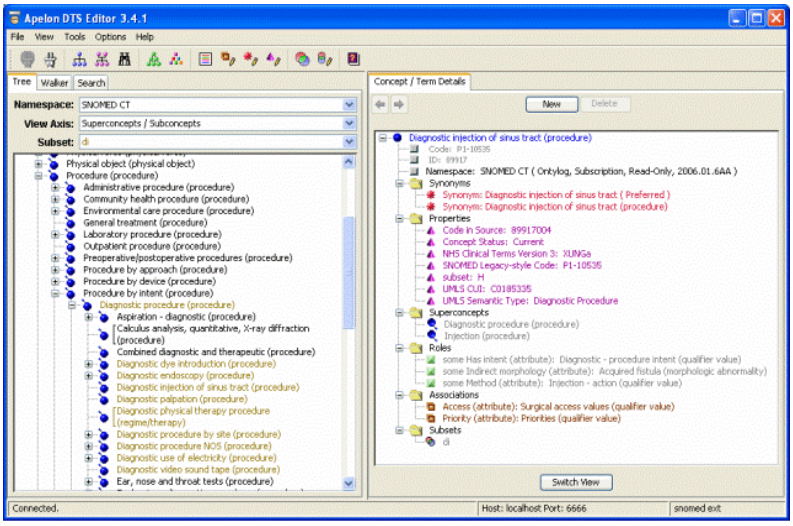
Mobiele terminologiese toepassings is tans besig om vorm binne terminologiebestuur aan te neem. Dié toepassings, soos byvoorbeeld *Glossaries2Go*, *Lokalise* en *The Interpreter's Wizard* (vgl. Costa et al. 2017: 73-80), word vir toestelle soos slimfone en tablette ontwikkel om op bedryfstelsels soos Android en/of iOS te funksioneer. In ander gevalle is daar mobiele terminologiese toepassings, soos byvoorbeeld *Glossary Assistant* (vgl. Costa et al. 2017: 73-80) en *Interplex*, wat ook op 'n rekenaar geïnstalleer kan word. In vergelyking met losstaande- en internetgebaseerde sagteware het mobiele terminologiese toepassings dikwels beperkte funksies. Sien byvoorbeeld die toepassing *The Interpreter's Wizard*. Hierdie toepassing laat die terminoloog toe om slegs tweetalige glossariums te bestuur en hierdie glossariums kan nie afgelaai word nie. Slegs een gebruiker het toegang tot die terminologiese inligting en die terminoloog is self verantwoordelik vir rugsteunkopieë van die data. Binne Suid-Afrika se veeltalige bestel sal hierdie tipe toepassing nie noodwendig 'n geskikte terminologiebestuursisteem wees nie.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)



Figuur 4: 'n Gedeelte van die sisteemagterkant van *Lokalise*

Sekere beskikbare terminologiebestuursisteme kan slegs binne spesifieke kontekste van terminologiebestuur gebruik word, byvoorbeeld *Apelon Distributed Terminology System* (*DTS*), wat uitsluitlik vir die bestuur van terminologie binne die gesondheidsektor gebruik word.



Figuur 5: 'n Gedeelte van die sisteemagterkant van *Apelon DTS*

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Terminologiese inligting kan ook binne kennisgebaseerde sisteme opgeneem word. Meyer et al. (1997) gebruik die term *kennisgebaseerd* om onderskeid te tref tussen twee verskillende dog komplementêre betekenisse, nl. (a) die algemene betekenis, d.i. vakspesifieke kennis en (b) die tegniese betekenis, d.i. enkodering as 'n basis van kennis. Elk van hierdie betekenisse word vervolgens kortliks bespreek.

Die eerste (algemene) betekenis van *kennisgebaseerd* behels die verkryging van vakspesifieke kennis. Hoewel 'n terminoloog nie 'n spesialis op elke vakgebied kan wees waarbinne hy/sy werk nie, vereis terminologieprosessering dat die terminoloog wel oor 'n mate van vakspesifieke kennis binne 'n bepaalde vakgebied beskik: "the acquisition of subject-field knowledge permeates virtually every aspect of terminology management, making it an activity that is highly knowledge-based" (Meyer et al. 1997: 98).

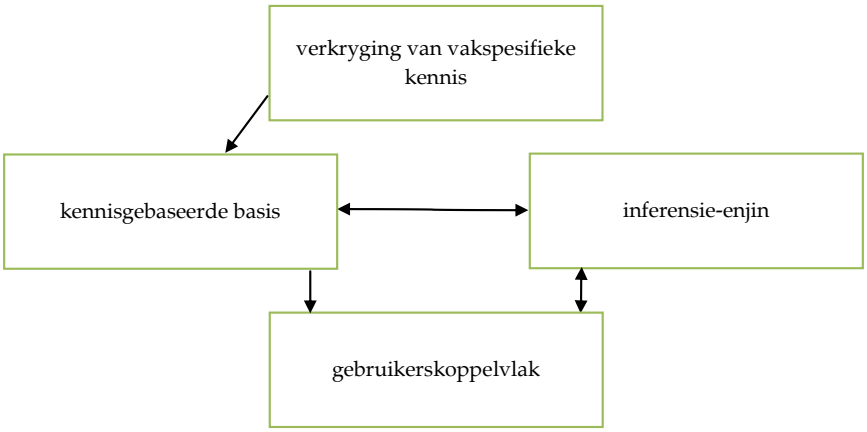
Afgesien daarvan dat die eindproduk van terminologiebestuur op kennis gebaseer is, is hierdie kennis — volgens die tweede (tegniese) betekenis van *kennisgebaseerd* — nie sistematies geënkodeer nie. Dit beteken dat 'n groot gedeelte van die vakkennis wat sorgvuldig deur die terminoloog bymekaar gemaak is, nié deur die teikengebruiker geraadpleeg kan word nie. Hierdie vakkennis vorm deel van die terminoloog se persoonlike geheuebank en vorm dus nie deel van enige reële kennisversameling nie. Dit is gevolglik nie toeganklik vir gebruikers van die uiteindelike terminologieproduk nie. Die teikengebruiker kan wel fragmente van hierdie vakkennis raadpleeg soos dit vervat is in verklarende inligting en gebruiksvoorbeelde. Addisionele vakkennis kan die kwaliteit van vakspesifieke tekste wat deur die teikengebruiker, soos byvoorbeeld 'n taalpraktisyn, geskep word, verhoog. Teikengebruikers, soos byvoorbeeld studente, kan die addisionele vakkennis as toegangspunte tot kennis gebruik (Meyer et al. 1997: 99). Binne die kennisgebaseerde benadering tot terminologie is kundiges dit eens dat teikengebruikers baat sal vind by die addisionele (vak)kennis wat tans nie binne 'n tradisionele gestruktureerde kennissisteem wat enige terminologieproduk onderlê, opgeneem is nie.

'n Kennisgebaseerde sisteem (Eng: *knowledge-based system*) wat deel vorm van kunsmatige intelligensie (Eng: *artificial intelligence*), is daarop gestel om (menslike) kennis vas te vang en as 'n bewaarplek (Eng: *repository*) van sodanige kennis te dien (Pratt 2022, Meyer et al. 1997: 99). Kennisgebaseerde sisteme beeld sowel die eienskappe van individuele konsepte as die verhoudings tussen konsepte uit.

Die tipiese argitektuur van 'n kennisgebaseerde sisteem sluit 'n kennisgebaseerde basis en 'n inferensie-enjin (Eng: *inference engine*) in. Die kennisgebaseerde basis bevat die vakkennis van 'n gegewe vakgebied. Dit is dan binne hierdie basis waar die (addisionele) vakkennis waaroor die terminoloog beskik, opgeneem kan word. Die inferensie-enjin maak op sy beurt afleidings uit die vakkennis wat in die kennisgebaseerde basis gestoor is. Kennisgebaseerde sisteme bevat ook gebruikerskoppelvlakke wat die gebruiker in staat stel om die sisteem te raadpleeg.

Die voorafgaande paragraaf word vervolgens skematies saamgevat.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)



Figuur 6: Skematiese voorstelling van 'n tipiese kennisgebaseerde sisteem

Dit is egter nie so eenvoudig om kennis binne 'n kennisgebaseerde sisteem op te neem nie. Hayes-Roth (1987: 293) stel dit as volg: "Knowledge does not come off the shelf, prepackaged, ready for use. On the contrary, knowledge is the word used to describe a variety of fragmentary bits of understanding …." In reaksie hierop stel 'n subkategorie van kunsmatige intelligensie, nl. kennismodellering (Eng: *knowledge engineering*), voor dat kennis sistematies geënkodeer word. Terwyl die kennismodelleerder gemoeid is met die enkodering van kennis wat uiteindelik deur 'n masjien gebruik gaan word, is die terminoloog gemoeid met die enkodering van kennis vir die gebruik deur 'n mens. Tog deel die terminoloog en die kennismodelleerder een kardinale belang: die sistematiese analise van gespesialiseerde konsepte (Meyer et al. 1997: 99).

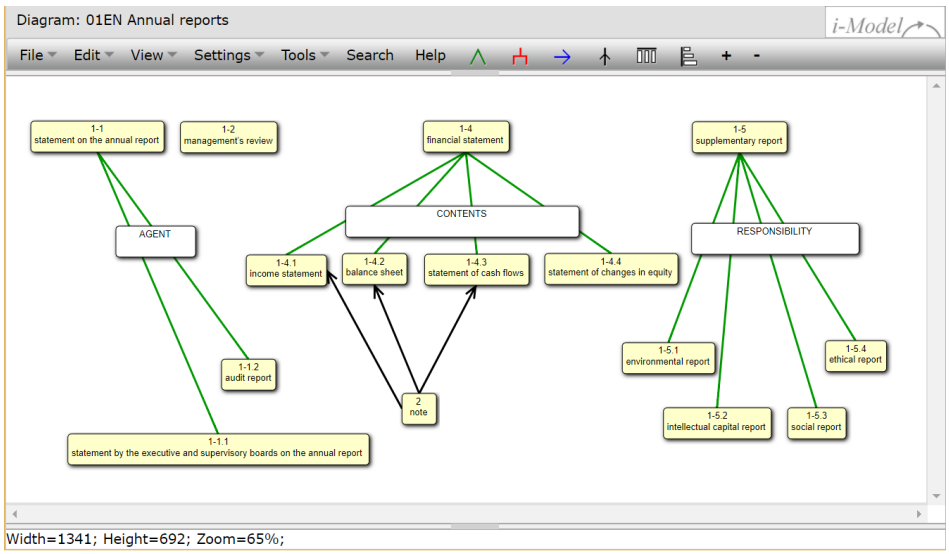
'n Wesenlike verskil tussen terminologiese databasisse en terminologiese kennisbasisse is dus die wyse waarop terminologiese inligting aangebied word. Binne terminologiese kennisbasisse word terminologiese inligting hoofsaaklik d.m.v. konseptuele sisteme aangebied. Daar word geargumenteer dat terminologiebestuur tot 'n groot mate gebaseer is op konseptuele sisteme en dat sodanige sisteme die skakels tussen verwante konsepte binne 'n bepaalde vakgebied uitbeeld (Wright en Budin 1997: 337). Binne 'n terminologiese kennisbasis kan daar byvoorbeeld ekwivalentrelasies, assosiatiewe relasies, generiese relasies, deelversamelingrelasies of ontologiese relasies opgeneem word.

Daar is sagteware, soos byvoorbeeld *Thinkmap*, beskikbaar waarbinne konseptuele sisteme geskep word en aan die teikengebruiker van die terminologieprodukt beskikbaar gestel word. Hierdie tipe sagteware kan ook by bestaande internetgebaseerde terminologiese sisteme geïntegreer word waar die teikengebruiker toegang tot sowel konseptuele sisteme as ander terminologiese inligting het. Nog sagteware wat gebruik kan word om konseptuele sisteme te bou, sluit in: *CMapTool*, *FreeMind* en *Visual Understanding Environment*. In die volgende paragraaf word daar gekyk na 'n konsepgebaseerde sisteem wat spesifiek vir terminologie ontwikkel is.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

i-Term

i-Term is internetgebaseerde sagteware wat platformonafhanklik is wat sowel 'n terminologiese as 'n konsepgebaseerde sisteem is. *i-Term* gebruik 'n byvoegingsmodule, *i-Model*, om sodanige sisteme te bou. Binne *i-Model* kan verskillende tipes relasies geskep word, byvoorbeeld assosiatiewe, deel-geheel-relasies en temporale relasies. Die terminoloog kan addisionele inligting soos kenmerkende eienskappe, onderverdelingskriteria (onderverdelingsdimensies) en notas met betrekking tot die konsepte by die konseptuele sisteme voeg. Gedeeltes van die terminologiese inskrywings se datavelde word outomaties gevul, gebaseer op die inligting van die visuele konseptuele sisteme wat met behulp van *i-Model* geskep is (Madsen 2005: 41, Steurs et al. 2015: 241). Binne 'n terminologiebestuursisteem is so 'n outovoltooifunksie nuttig, aangesien dit tyd en hulpbronne bespaar om terminologiese inskrywings aan te vul. Kundiges beskou *i-Term* as ultramodern, gebruikersvriendelik en konsekwent in die skep van konseptuele sisteme (Madsen 2005: 52, Costa et al. 2017: 72). In die onderstaande figuur word die term 'financial statement' binne die relevante konseptuele raamwerk geïllustreer.



Figuur 7: 'n Gedeelte van die sisteemagterkant van *i-Term se byvoegingsmodule, i-Model*

Behoeftes van gebruikers vir terminologiebestuursisteme

Gebruikerstudies (vgl. Bilgen 2009 en Costa et al. 2017) wat oor terminologiebestuursisteme uitgevoer is, het bevind dat gebruikers 'n behoefte het aan terminologiebestuursisteme wat maklik toeganklik is, vinnige (gevorderde) soektogte

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

kan uitvoer en gebruikersvriendelik is. Terminologiebestuursisteme moet verskillende datakategorieë hê na gelang van die behoeftes van die teikengebruiker en daar moet ook veeltalige terminologiese inskrywings binne sodanige sisteme opgeneem kan word. Laastens is daar 'n behoefte aan terminologiebestuursisteme wat koste-effektief is. Enkele gebruikerstudies wat oor terminologiese kennisbasisse uitgevoer is, fokus slegs op 'n klein groep gebruikers (López-Rodríguez et al. 2012). In een van hierdie studies wat deur Prieto Velasco en López Rodríguez (2007) uitgevoer is, is daar bevind dat gebruikers terminologiese inligting wat visueel aangebied word, meer bruikbaar vind en dus meer tevredenheid toon met sodanige terminologieprodukte. In López-Rodríguez et al. 2012 se studie word gemeld dat enige tipe visuele hulpbronne wat binne 'n terminologiese kennisbasis gebruik word, gemik moet wees op die profiel van die teikengebruiker. Hierdie studies is dit eens dat navorsing oor terminologiese kennisbasisse uitgebrei moet word na groter groepe gebruikers, asook na verskillende terminologiese kennisbasisse. Met betrekking tot die Suid-Afrikaanse konteks moet die gebruik van terminologiebestuursisteme en die implementering van terminologiese databasisse 'n gevestigde praktyk word, alvorens enige gebruikerstudies oor terminologiese kennisbasisse en konsepgebaseerde sisteme vir die Suid-Afrikaanse gebruiker uitgevoer kan word. Aanvanklike gebruikerstudies moet dus fokus op terminologiebestuursisteme en terminologiese databasisse.

'n Keuse oor die mees geskikte terminologiebestuursisteem of terminologiese konsepgebaseerde sisteem vir 'n gegewe terminologieproduk moet gebaseer wees op navorsing wat uitgevoer is. Navorsing moet fokus op aspekte soos die getal datakategorieë wat binne die terminologiebestuursisteem beskikbaar is, die getal veelvuldige — en veeltalige — glossariums en termlyste wat binne 'n terminologiebestuursisteem opgeneem kan word, die *housekeeping*-funksie van die terminologiebestuursisteem, statistiese funksies en die koste-implikasies van so 'n sisteem. Daar word vervolgens na hierdie aspekte gekyk.

Die tipe terminologiese inligting wat vir die teikengebruiker van die terminologieproduk beskikbaar gestel word, bepaal die datakategorieë wat binne 'n terminologiebestuursisteem of terminologiese kennisbasis opgeneem word. Die *International Organization for Standardization* (ISO) 12620 (2009) lys byna 200 moontlike datakategorieë vir 'n enkele terminologiese inskrywing waarvan slegs drie datakategorieë, d.i. *term, bron* en *datum*, as verpligte velde gelys word (Muegge 2010: 18). Sekere terminologiebestuursisteme, soos byvoorbeeld *Evoterm*, het vooropgestelde datamodelle wat nie deur die gebruiker aangepas kan word nie. Die gebruiker het dus net toegang tot 'n spesifieke stel datakategorieë wat nie noodwendig relevant vir 'n bepaalde teikengebruiker van die terminologieproduk is nie. Terminologiebestuursisteme soos *Flashterm 2018* en *Autshumato TMS* het 'n vooropgestelde datamodel, maar dié model kan wel aangepas word. Laastens is daar terminologiebestuursisteme, soos byvoorbeeld *ODABA*, beskikbaar wat die terminoloog toelaat om sy/haar eie datamodel binne die terminologiebestuursisteem te bou. Datakategorieë wat binne 'n terminologiebestuursisteem

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

of terminologiese kennisbasis geskep moet word en nie deel vorm van die moontlike datakategorieë wat deur die ISO 12620 gelys word nie, moet voldoen aan die ontwerp van datakategoriespesifikasies soos omskryf deur die ISO 12620 (ISO 26162-1 2019).

Vir dokumenteringsdoeleindes maak die sogenaamde *housekeeping*-funksie (onderhoudfunksie) 'n integrale deel van 'n terminologiebestuursisteem en terminologiese kennisbasis uit. Hierdie funksie dui gewoonlik inligting soos die rekordnommer en datum van 'n inskrywing aan, wanneer 'n inskrywing geskep of aangepas is, wie aan 'n bepaalde inskrywing gewerk het, die vakgebied of projek waaronder 'n inskrywing geklassifiseer moet word en die status van elke inskrywing.

Statistiese funksies binne 'n terminologiebestuursisteem is 'n nuttige funksie om die getal suksesvolle en onsuksesvolle soektogte wat binne 'n terminologiese databasis of terminologiese kennisbasis gedurende 'n gegewe tydperk uitgevoer is, te kontroleer. Hierdie inligting help die terminoloog om die terminologiese inligting wat binne so 'n databasis of kennisbasis opgeneem is, aan te vul (Bauer 2015: 337).

Sommige terminologiebestuursisteme kan gratis gebruik word (ook bekend as *oopbron*) terwyl ander terminologiebestuursisteme teen 'n eenmalige bedrag aangekoop kan word. Sagteware wat gratis beskikbaar is, het dikwels beperkte funksies wat 'n remskoen in die proses van terminologiebestuur kan wees. Sekere pakkette laat ook die gebruiker toe om die sagteware teen 'n maandelikse of jaarlikse subskripsiefooi te gebruik.

5. Terminologiebestuursdiagramme, datavloeidiagramme vir terminologie en terminologiewerkvloeidiagramme

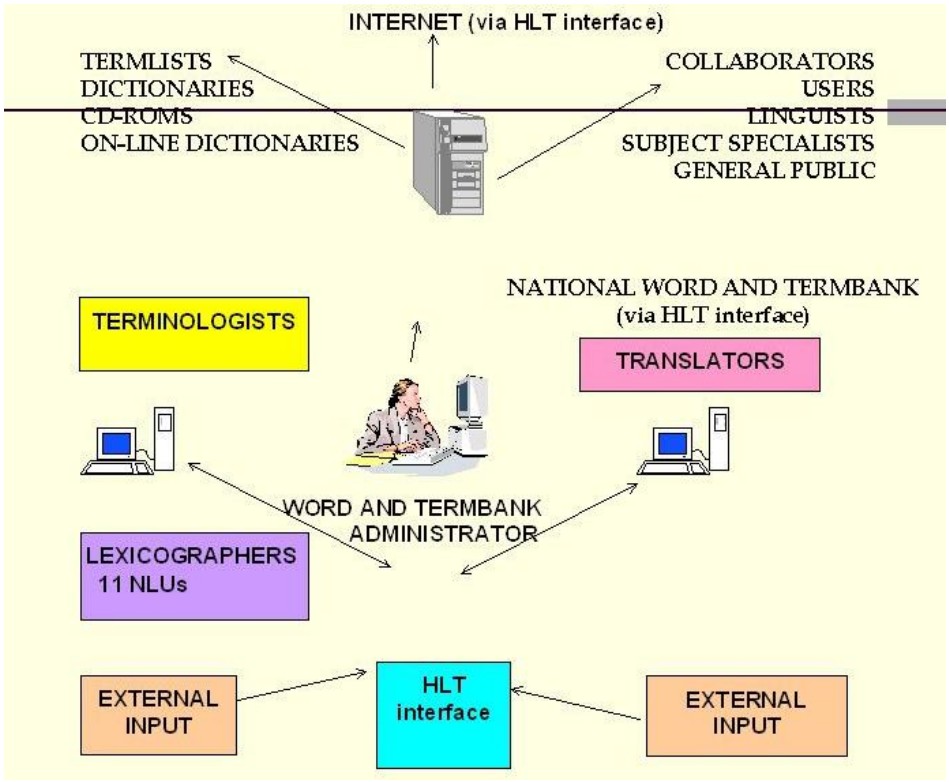
Binne terminologiebestuur word daar onderskeid getref tussen 'n terminologiebestuursdiagram, 'n datavloeidiagram vir terminologie en 'n terminologiewerkvloeidiagram. Hoewel hierdie grafiese voorstellings uniek is en saamgestel en aangepas word na gelang van die aard van 'n gegewe terminologiebestuursprogram, is daar sekere aspekte binne elk van hierdie diagramme wat binne enige terminologiebestuursproses in ag geneem moet word. Daar word vervolgens kortliks verwys na terminologiebestuursdiagramme en datavloeidiagramme vir terminologie gevolg deur 'n bespreking oor terminologiewerkvloeidiagramme.

Terminologiebestuursdiagramme

'n Terminologiebestuursdiagram kan beskryf word as die oorhoofse diagram wat die groter konteks gee waarbinne terminologiebestuur uitgevoer word. Hierdie diagram kan in drie dele opgedeel word, nl.: die terminologiebestuursisteem, die sisteemvoorkant en die sisteemagterkant. Binne die terminologiebestuursisteem word die terminologiese inligting opgeneem en gestoor. Aan

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

die sisteemvoorkant word die terminologiese inligting deur die teikengebruiker van die terminologieproduk geraadpleeg. Die sisteemagterkant sluit in wie by die proses van terminologiebestuur betrokke is, byvoorbeeld terminoloë, vakspesialiste, vertalers en databankadministrateurs. Ter illustrasie word een van die leksikografie- en terminologiebestuursdiagramme van die Terminologiekoördineringsafdeling van die Nasionale Taaldiens gegee (Alberts 2006).



Figuur 8: 'n Terminografie- en terminologiebestuursdiagram van die Terminologiekoördineringsafdeling, Nasionale Taaldiens (Alberts 2006)

In die bostaande figuur word die terminologiebestuursisteem as die *Human Language Technology (HLT) interface* (koppelvlak) gemerk. Aan die voorkant word die terminologiese (en leksikografiese) inligting d.m.v. 'n standaard webblaaier deur die teikengebruikers, soos byvoorbeeld vakspesialiste, linguiste en die algemene publiek geraadpleeg. Aan die agterkant is kundiges soos terminoloë en vertalers by die proses van terminologiebestuur betrokke.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Datavloeidiagramme vir terminologie

In die breë sin kan 'n datavloeidiagram omskryf word as 'n (visuele) voorstelling van hoe inligting binne 'n bepaalde proses of sisteem vloei. Meer spesifiek dan, hierdie diagram dui aan hoe data hanteer word, wie die data hanteer, waar die data gestoor word en wat met die data gedoen word. 'n Datavloeidiagram identifiseer potensiële probleme in 'n proses of 'n sisteem, bevorder doeltreffendheid t.o.v. werkvloei en vermy die duplisering van data (Brändle 2013, Chi 2021). Aspekte wat in 'n datavloeidiagram vir terminologie opgeneem moet word, sluit in die versameling van data, die prosessering van data, die integrering van data in 'n terminologiebestuursisteem of terminologiese kennisbasis en die beskikbaarstelling van terminologiese inligting aan die teikengebruiker van die terminologieproduk.

Terminologiewerkvloeidiagramme

'n Terminologiewerkvloeidiagram stel 'n herhalende patroon voor van hoe take aan die agterkant van terminologiebestuur van die begin tot die einde uitgevoer moet word en wie vir sodanige take verantwoordelik is. Alvorens 'n terminologiewerkvloeidiagram saamgestel kan word, moet daar 'n terminologiewerkvloei-analise gedoen word om die knelpunte in die terminologiebestuursproses te identifiseer en op te los om moontlike vertragings te voorkom. Hierdie knelpunte kan 'n (direkte) uitwerking hê op die opbrengs op belegging wat deur terminologiebestuur gegenereer kan word. Indien daar nie in 'n terminologiewerkvloei voorsiening gemaak vir die deurlopende bywerking van die status van terminologiese inskrywings wat beskikbaar gestel kan word nie, kan dit onder andere die bekendstellingsproses van 'n bepaalde terminologieproduk vertraag. Die sogenaamde *delayed time-to-market* van 'n produk lei tot 'n verlies aan inkomste, omdat dit nie binne 'n gegewe tydperk bekendgestel kon word nie. Die vervaardiger van sodanige produk kan ook nie noodwendig 'n mededingende voordeel uit die mark trek nie. Hoewel die konsep van 'n terminologieproduk as kommersiële produk wat potensieel 'n opbrengs kan lewer, nog nie werklik binne die Suid-Afrikaanse konteks posgevat het nie, kan hierdie aspek nie buite rekening gelaat word in die beplanning van 'n behoorlike terminologiebestuursprogram (of -programme) vir die Suid-Afrikaanse situasie nie.

Berns (2020) meld dat enige terminologiewerkvloei geoptimaliseer kan word binne die raamwerk van slegs teksskepping, proeflees, vertaling en die publisering van data. Daar is egter addisionele aspekte wat binne elke terminologiewerkvloei in ag geneem moet word. Om al die aspekte saam te vat wat vir 'n terminologiewerkvloei van belang is — en wie daarvoor verantwoordelik is — word die terminologiewerkvloeidiagram binne terminologiebestuur in die addendum voorgestel. Hierdie diagram is spesifiek ontwerp vir die veeltalige Suid-Afrikaanse konteks, en kan dien as 'n riglyn vir die samestelling van 'n terminologiewerkvloeidiagram.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Die eerste aktiwiteit in hierdie terminologiewerkvloei is navorsing. Hierdie aktiwiteit kan opgedeel word in navorsing op makrovlak en navorsing op mikrovlak. Navorsing op makrovlak sluit die algemene navorsing in wat oor die tegniese aspekte van terminologiebestuursisteme of terminologiese kennisbasisse gedoen moet word ten einde 'n geskikte terminologiebestuursisteem of terminologiese kennisbasis vir 'n spesifieke konteks en vir 'n spesifieke teikengebruiker te kies of te ontwerp. Op dié vlak word daar die rekenaargeletterdheid van die teikengebruiker van die terminologieproduk bepaal om te verseker dat daar voorsiening vir verskillende vlakke van bedrewenheid gemaak kan word. Makrovlaknavorsing word deur die terminoloog in samewerking met die IT-spesialis of kennismodelleerder uitgevoer. Navorsing op mikrovlak fokus in die eerste plek op die profiel van die teikengebruiker van die terminologieproduk. In die tweede plek word mikrovlaknavorsing gedoen vir die identifisering en versameling van geskikte terminologiese hulpbronne binne 'n spesifieke vakgebied vir 'n spesifieke teikengebruiker waar hierdie hulpbronne vir die onttrekking van terminologiese inligting gebruik sal word. Die terminoloog en die vakspesialis is verantwoordelik vir navorsing op mikrovlak. Die tweede aktiwiteit behels die versameling van terminologiese inligting, byvoorbeeld terme, konseptuele relasies, verklarende inligting, gebruiksvoorbeelde en linguistiese inligting. Terminologiese inligting kan handmatig, semi-outomaties of outomaties uit die geskikte terminologiese hulpbronne deur die terminoloog onttrek word. Die verifiëring van die terminologiese inligting wat onttrek is, is die derde aktiwiteit wat deur die terminoloog en vakspesialis gedoen moet word. Met die profiel van die teikengebruiker van die terminologieproduk in gedagte, is dit hier belangrik om die "geraas" (ongeldige potensiële terme) en "stilte" (ontbrekende geldige terme) wat in termlyste mag voorkom, onderskeidelik te verwyder en by te voeg. Binne hierdie aktiwiteit word verklarende inligting en gebruiksvoorbeelde ook saamgestel. Die volgende drie aktiwiteite, d.i. aktiwiteite vier, vyf en ses, hou verband met die voorbereiding van die terminologiese data vir die terminologiebestuursisteem of terminologiese kennisbasis. Die terminologiese inligting in die brontaal word deur die taalpraktisyn geproeflees en geredigeer waarna die vertaler dié data na die teikentaal toe vertaal. Om die akkuraatheid van die vertalings te verseker moet die vertaalde terminologiese data deur die vakspesialis nagegaan word. Dit is verkieslik dat die teikentaal die vakspesialis se sterkste taal is. Binne 'n veeltalige bestel soos Suid-Afrika is vakspesialiste se sterkste taal nie noodwendig die teikentaal nie. In sulke gevalle word die vertalings deur die vakspesialis én 'n spreker wie se sterkste taal die teikentaal is, gekontroleer. Die vertaalde terminologiese inligting word ook deur 'n taalpraktisyn geproeflees en geredigeer. Die volgende aspek binne hierdie aktiwiteite is die verifiëring en standaardisering van die vertaalde terminologiese inligting. Binne die Suid-Afrikaanse konteks berus hierdie verantwoordelikheid by die gesaghebbende liggaam, die Pan-Suid-Afrikaanse Taalraad (PanSAT). Aktiwiteite vier, vyf en ses word deur die terminoloog

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

gekoördineer. Die vertaalde terminologiese inligting word in die sewende aktiwiteit binne die terminologiewerkvloei deur die databankadministrateur, terminoloog, IT-spesialis of kennismodelleerder sistematies na die terminologiebestuursisteem ingetrek of ingevoer. Die voorlaaste aktiwiteit word deur die databankadministrateur of terminoloog behartig waar die status van inskrywings wat ingetrek of ingevoer is, bygewerk word en volledige inskrywings aan die teikengebruiker van die terminologieproduk beskikbaar gestel word. Om voorsiening te maak vir die terugvoer wat van die teikengebruiker van die terminologieproduk ontvang word, soos byvoorbeeld om die huidige terminologiese inligting te hersien of om nuwe relevante terminologiese inligting by te voeg, word daar in die laaste aktiwiteit in die terminologiewerkvloei vereis dat die terminologiese inligting deurlopend bygewerk word. Die terminoloog, die vakspesialis, die vertaler én die taalpraktisyn is by hierdie aktiwiteit betrokke.

6. Samevatting

Terminologiebestuur is 'n multidissiplinêre aktiwiteit en word nie alleenlik deur die terminoloog uitgevoer nie, maar in samewerking met ander kundiges en gesaghebbende liggame. In hierdie artikel word aangedui dat terminologiebestuur gestruktureer kan word en op makro-, meso- en mikrovlak uitgevoer kan word. Terminologiebestuur op sowel meso- as mikrovlak kan ook Suid-Afrikaanse taalgemeenskappe se terminologiese behoeftes bevredig en word nie uitsluitlik deur gesaghebbende liggame op nasionale vlak (makrovlak) uitgevoer nie. Hoewel daar verskillende terminologiebestuursisteme vir verskillende kontekste van terminologiebestuur beskikbaar is, moet sodanige sisteme eerder vir 'n spesifieke konteks van terminologiebestuur wat aan die behoeftes van 'n spesifieke teikengebruiker van 'n terminologieproduk voldoen, ontwerp word. Die voorgestelde praktiese terminologiewerkvloeidiagram dien as 'n riglyn om terminologiebestuur in die veeltalige Suid-Afrikaanse omgewing meer doeltreffend en gebruikersgeoriënteerd te maak.

Eindnote

1. Gebaseer op die inligting wat geredelik beskikbaar is wanneer 'n internetsoektog uitgevoer word.
2. Vgl. http://www.dac.gov.za/sites/default/files/LPD_Implem%20Plan_10%20April%202003.pdf
3. *tlTerm* kan wel as losstaande sagteware funskioneer, maar dit is meer doeltreffend wanneer veelvuldige gebruikers aan 'n projek werk om dit so op te stel dat dit via 'n netwerk of die internet funksioneer.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Bronnelys

Alberts, M. 2006. National Language and Terminology Policies — A South African Perspective. *TAMA Workshop, 15 November 2006, Lessius University College, Antwerp.*

Alberts, M. 2010. National Language and Terminology Policies — A South African Perspective. *Lexikos* 20: 599-620.

Allard, M.G.P. 2012. *Managing Terminology for Translation Using Translation Environment Tools: Towards a Definition of Best Practices.* Ongepubliseerde Ph.D.-proefskrif. Ottowa: Universiteit van Ottowa.

Antia, B.E. 2015. Language Policy and Terminology in South Africa. Kockaert, H.J. en F. Steurs (Reds.). 2015. *Handbook of Terminology. Volume 1:* 467-488. Amsterdam/Philadelphia: John Benjamins.

Bauer, S.C. 2015. Managing Terminology Projects. Kockaert, H.J. en F. Steurs (Reds.). 2015. *Handbook of Terminology. Volume 1*: 324-340. Amsterdam/Philadelphia: John Benjamins.

Bergenholtz, H. en S. Tarp (Reds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.

Bergenholtz, H., S. Nielsen en S. Tarp (Reds.). 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow.* Bern/New York: Peter Lang.

Berns, K. 2020. The Perfect Terminology Workflow. https://www.linkedin.com/pulse/perfect-terminology-workflow-kerstin-berns/

Bilgen, B. 2009. *Investigating Terminology Management for Conference Interpreters*. Ongepubliseerde proefskrif. Ottowa: Universiteit van Ottowa.

Bothma, T.J.D., R.H. Gouws en D.J. Prinsloo. 2016. The Role of e-Lexicography in the Confirmation of Lexicography as an Independent and Multidisciplinary Field. Margalitadze, T. and G. Meladze (Reds.). 2016. *Proceedings of the XVII EURALEX International Congress, Lexicography and Linguistic Diversity, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia, 6–10 September, 2016:* 109-116. Tbilisi: Ivane Javakhishvili Tbilisi State University.

Bowker, L. 2020. Terminology Management. Angelone, E., M. Ehrensberger-Dow en G. Massey (Reds.). 2020. *The Bloomsbury Companion to Language Industry Studies*: 261-283. Londen/New York/Oxford/New Delhi/Sydney: Bloomsbury Academic.

Brändle, D. 2013. Terminology Workflows and Interfaces. *International Terminology Summer School 2013, Cologne, Germany, 15–19 July 2013.*

Chi, C. 2021. A Beginner's Guide to Data Flow Diagrams. https://blog.hubspot.com/marketing/data-flow-diagram

Costa, H., G.C. Pastor en I. Durán-Muñoz. 2017. Assessing Terminology Management Systems for Interpreters. Pastor, G.C. en I. Durán-Muñoz (Reds.). 2017. *Trends in e-Tools and Resources for Translators and Interpreters*: 57-84. Leiden/Boston: Brill Rodopi.

Departement van Sport, Kuns en Kultuur. https://www.dsac.gov.za/ [10 November 2022]

De Schryver, G.-M. 2009. State-of-the-art Software to Support Intelligent Lexicography. Zhu, R. (Red.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary and Lexicology. Volume 2*: 584-599. Beijing: Beijing Normal University.

Durán-Muñoz, I. 2010. Specialized Lexicographical Resources: A Survey of Translators' Needs. Granger, S. en M. Paquot (Reds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of ELEX2009*: 55-66. Louvain-la-Neuve: Presses universitaires de Louvain.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Evans, J. 2024. Put Your Money where Your Morals Are — Where SA's Top 5 Banks Rank in Climate Policy. https://www.dailymaverick.co.za/article/2024-11-15-south-africas-banks-struggle-to-align-climate-strategies-with-urgent-decarbonisation-goals/

Gouws, R.H. en D.J. Prinsloo. 2005. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. *International Journal of Lexicography* 18: 25-46.

Hayes-Roth, F. 1987. Expert Systems. Shapiro, Stuart C. (Red.). 1987. *Encyclopedia of Artificial Intelligence. Volume 1:* 287-298. New York: Wiley-Interscience.

International Organization for Standardization. 2019. ISO 26162-1:2019(E). *Management of Terminology Resources — Terminology Databases*.

López-Rodríguez, C.I., M.B. Castro en A.G. Aragón. 2012. User Needs to the Test: Evaluating a Terminological Knowledge Base on the Environment by Trainee Translators. *The Journal of Specialised Translation* 18: 57-76.

Madsen, B.N. 2005. Implementation of Research Results in the Development of an Internet Terminology and Knowledge Management System. *Terminology Science & Research / Terminology: Science et Recherche* 16: 39-53.

Meyer, I., K. Eck en D. Skuce. 1997. Systematic Concept Analysis within a Knowledge-based Approach to Terminology. Wright, S.E. en G. Budin (Reds.). 1997. *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*: 98-118. Amsterdam/Philadelphia: John Benjamins.

Muegge, U. 2010. Terminology Management. https://bioprocessintl.com/business/regulatory-affairs/terminology-management-237288/

National Language Policy Framework Implementation Plan. http://www.dac.gov.za/sites/default/files/LPD_Implem%20Plan_10%20April%202003.pdf [14 Augustus 2021]

Pratt, M.K. 2022. Data Curation. https://www.techtarget.com/searchbusinessanalytics/definition/data-curation

Prieto Velasco, J.A., M.I. Tercedor Sánchez en C.I. López Rodríguez. 2007. Using Multimedia Materials in the Teaching of Scientific and Technical Translation. *Linguistica antverpiensia* 6: 115-134.

Sauberer, G. 2018. Terminology with Ease, a 5-step Action Plan for Your Professional Terminology Management. http://termnet.eu/wp-content/uploads/2018/03/5-Steps-Action-Plan_Terminology-with-Ease-1.pdf

Steurs, F., K. de Wachter en E. de Malsche. 2015. Terminology Tools. Kockaert, H.J. en F. Steurs (Reds.). 2015. *Handbook of Terminology. Volume 1:* 222-249. Amsterdam/Philadelphia: John Benjamins.

Tarp, S. 2004. Basic Problems of Learner's Lexicography. *Lexikos* 14*:* 222-252.

Warburton, K. 2014. Terminology Management. Sin-wai, C. (Red.). 2014. *The Routledge Encyclopedia of Translation Technology*: 682-699. Londen: Routledge.

Warburton, K. 2015. Managing Terminology in Commercial Environments. Kockaert, H.J. en F. Steurs (Reds.). 2015. *Handbook of Terminology. Volume 1:* 360-392. Amsterdam/Philadelphia: John Benjamins.

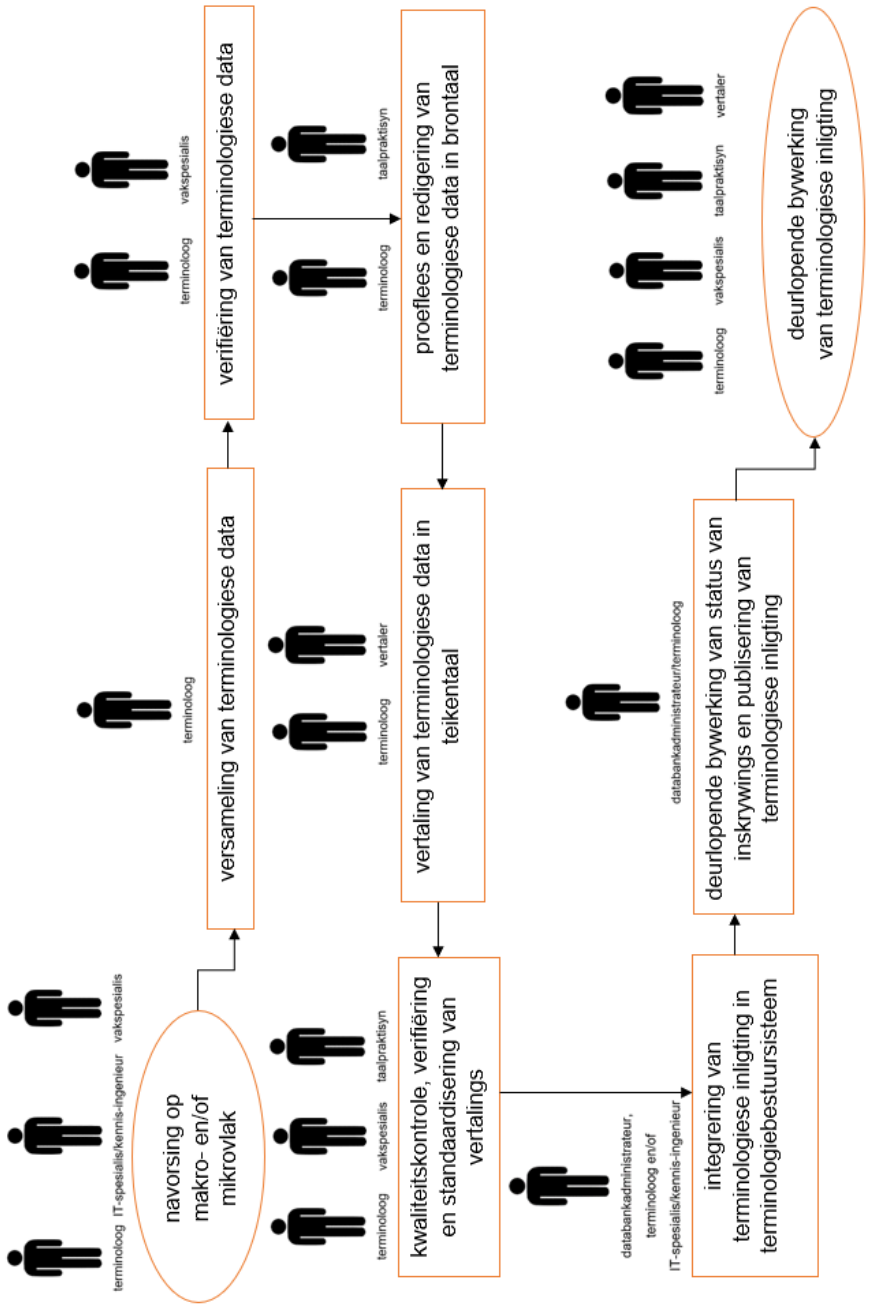
Warburton, K. 2024. Developing a Business Case for Managing Terminology. https://termologic.com/wp-content/uploads/2024/05/roi-article-warburton.pdf

Wright, S.E. en G. Budin. 1997. *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management.* Amsterdam/Philadelphia: John Benjamins.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2049 (Article)

Addendum

Voorgestelde terminologiewerkvloeidiagram vir die Suid-Afrikaanse konteks



http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

Endemann's *Wörterbuch der Sotho Sprache* (1911): A Worthy Candidate for Digitisation

Elsabé Taljard, *Department of African Languages, University of Pretoria, Pretoria, South Africa (elsabe.taljard@up.ac.za) (https://orcid.org/0000-0002-4507-1633)*

Gertrud Faaß, *Institute for Information Science and Language Technology, University of Hildesheim, Hildesheim, Germany (gertrud.faass@uni-hildesheim.de) (https://orcid.org/0000-0002-8130-617X)*

Danie Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, South Africa (danie.prinsloo@up.ac.za) (https://orcid.org/0000-0003-0054-4676)*
and
Sonja Bosch, *Department of African Languages, UNISA, Pretoria, South Africa (seb@hbosch.com) (https://orcid.org/0000-0002-9800-5971)*

Abstract: This article re-evaluates *Wörterbuch der Sotho Sprache*, a historically significant, yet neglected Sotho–German dictionary, published in 1911 by Berlin missionary Karl Endemann. Its marginalisation stems from its choice of German as target language, outdated orthography, missionary orientation, and deviation from modern lexicographic principles. Rather than a conventional comparison with modern Sepedi dictionaries, this study positions Endemann's work within its historical and cultural context. Key lexicographic elements such as grammatical formatives, alphabetical categories, high-frequency lemmas, semantically related paradigms, and culturally significant entries are analysed in detail. The findings often reveal strengths that match or even surpass those of later Sepedi dictionaries. Despite its value, user access remains limited due to linguistic complexity and unavailability. With digitisation now permitted by the publisher, this study outlines a multi-phase strategy to enhance usability, including the use of OCR4all, an open-source tool for text recognition. While the digitisation process is not without challenges, the application of OCR4all has yielded impressive accuracy. However, despite the high-quality output, this margin of error necessitates manual verification to ensure the integrity of the digitised content as a reliable and accessible resource for modern users.

Keywords: ENDEMANN, SOTHO–GERMAN, CULTURAL HERITAGE, GRAMMATICAL FORMATIVES, HISTORICAL CONTEXT, LEXICOGRAPHIC PRINCIPLES, DIGITISATION, OCR4ALL, MODERN DAY USERS, ACCESSIBILITY

Lexikos 35 (AFRILEX-reeks/series 35: 2025): 462-481

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

Opsomming: Endemann se *Wörterbuch der Sotho Sprache* (1911): 'n waardige kandidaat vir digitalisering. Hierdie artikel herevalueer *Wörterbuch der Sotho Sprache*, 'n histories betekenisvolle dog verwaarloosde Sotho–Duitse woordeboek, wat in 1911 deur die Berlynse sendeling Karl Endemann gepubliseer is. Die marginalisering daarvan spruit uit sy keuse van Duits as doeltaal, verouderde ortografie, sendingoriëntasie en afwyking van moderne leksikografiese beginsels. Eerder as 'n konvensionele vergelyking met moderne Sepedi-woordeboeke, plaas hierdie studie Endemann se werk binne sy historiese en kulturele konteks. Sleutelleksikografiese elemente soos die hantering van grammatikale formatiewe, alfabetiese kategorieë, hoëfrekwensielemmas, semanties-verwante paradigmas en kultureel betekenisvolle inskrywings word in detail ontleed. Die bevindinge openbaar dikwels sterkpunte wat ooreenstem met of selfs dié van latere Sepedi-woordeboeke oortref. Ten spyte van die waarde daarvan, bly gebruikerstoegang beperk weens linguistiese kompleksiteit en onbeskikbaarheid. Met digitisering wat nou deur die uitgewer toegelaat word, sit hierdie studie 'n multifase strategie uiteen om bruikbaarheid te verbeter, insluitend die gebruik van OCR4all, oopbronsagteware vir teksherkenning. Alhoewel die digitiseringsproses nie sonder uitdagings is nie, het die toepassing van OCR4all indrukwekkende akkuraatheid opgelewer. Ten spyte van die hoë kwaliteit uitset, noodsaak 'n foutmarge egter handmatige verifikasie om die integriteit van die gedigitiseerde inhoud te verseker as 'n betroubare en toeganklike hulpbron vir moderne gebruikers.

Sleutelwoorde: ENDEMANN, SOTHO–DUITS, KULTURELE ERFENIS, GRAMMATIKALE FORMATIEWE, HISTORIESE KONTEKS, LEKSIKOGRAFIESE BEGINSELS, DIGITISERING, OCR4ALL, MODERNE GEBRUIKERS, TOEGANKLIKHEID

Introduction

In 1911, Karl Heinrich Julius Endemann, a Berlin missionary, published his dictionary of the Sotho language *Wörterbuch der Sotho Sprache*, 1911. Very little information is available regarding the actual process of data collection — in the *Vorwort* to the dictionary, Endemann (1911: VII) indicates that he collected the data in the period 1861–1873 while being active as a missionary in the then Transvaal, specifically on the missionary stations Gerlachshoop, Phatametsane, Botschabelo and Malokong. It can only be assumed that he made use of oral accounts as sources of data, since no textual material was available at that stage. Shrouded in mystery, however, is the source of his data on Setswana and Sesotho, which the dictionary also caters for. There is no evidence that speakers of these languages were present in the area in which he worked. Kosch (2012: 226) mentions that the dictionary has since its publication remained at the fringes of scholarly investigation — in their discussion of Bantu language dictionaries utilising a left-expanded article structure, Gouws and Prinsloo (2005) make no mention of Endemann's dictionary, and apart from two articles focusing on the dictionary by Kosch (2011, 2012) and references to the *Wörterbuch* in Kosch (1993) and Lombard (1970), there is no serious engagement with Endemann's *Wörterbuch* as an exceptional lexicographic endeavour. One reason is probably because the target language is German, and it is therefore not readily accessible to scholars

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

of the African languages. It was also compiled with a specific function in mind, i.e., to assist German missionaries in their missionary field work — a function that may be difficult to reconcile with the information needs of the modern-day user of a Sepedi dictionary. The dictionary furthermore places a high demand on the dictionary using skills and grammatical knowledge of the user, and it makes use of an orthography that is long since obsolete. Nevertheless, we agree with Kosch (2011) that this dictionary contains a "huge repository of information that awaits linguistic and cultural scrutiny and insights", and to which we would like to add, a lexicographic reappraisal. In our reappraisal, we take our cue from the following quote from Tarp (2013: 294): "[H]istory should not be viewed as an ever-growing progress, but as a process with its ups and downs. In this respect, some old lexicographical works … are in some aspects extremely advanced even compared with present-day dictionaries."

The aim of this article is therefore to reappraise the *Wörterbuch* within the context of existing bilingual Sepedi dictionaries, rather than to evaluate it against modern day lexicographic theory and practice. Based on our findings, we conclude that the dictionary warrants wider exposure, something that can only be attained by making it — or at least parts of it — available in digital format. Digitisation is also an important step in the preservation of this canonical work.

This article consists of three parts: an investigation of selected aspects of the macro- and microstructures of Endemann (1911), followed by a discussion of the (in)accessibility of the dictionary to modern day target users. In the third section of the article, the digitisation strategy and the necessary steps for making the dictionary available and useful for modern day users are described. We envisage a multiphase digitisation approach with different levels of accessibility to the dictionary, and different levels of true electronic features.

Contextualisation

In her article *Innovation and compromise in K. Endemann's dictionary of the Sotho language* Kosch (2011) concentrated on shortcomings and weaknesses of the dictionary, indicating that Endemann violated some good lexicographic principles in the compilation of the dictionary. In our reappraisal of the dictionary within the context of other bilingual Sepedi dictionaries, we address the following issues: treatment of grammatical formatives, over and/or under treatment of alphabetical categories, treatment of high frequency lemmas, completing semantically related paradigms and treatment of lemmas with cultural significance. The dictionaries that have been selected to provide a contextualised assessment of Endemann's dictionary cover a timespan of almost five decades, with publication dates ranging from 1967 to 2015. These dictionaries have been, and still are, regarded as standard Sepedi reference works, and we believe that comparing Endemann's dictionary with these dictionaries will give a more balanced view of the dictionary as a lexicographic artefact and its value as a utility instrument. The dictionaries selected for this investigation are the following:

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

De Schryver, G.-M. 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English*, henceforth OBSD;

Kriel, T.J. 1967. *The New English–Northern Sotho Dictionary*, henceforth NENSD;

Kriel, T.J. 1983. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*, henceforth PWNSA83;

Kriel, T.J., E.B. van Wyk and S.A. Makopo. 1989. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*, henceforth PWNSA89;

Mojela, V.M., M.C. Mphahlele, M.R. Selokela and W.M. Mojapelo. 2015. *Sesotho sa Leboa–English Bilingual Dictionary*, henceforth SsLEBD, and

Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho Woordeboek*, henceforth GNSW.

Treatment of grammatical formatives

Grammatical formatives are usually notoriously undertreated in bilingual dictionaries in which the source language is an African language, since these formatives are typically not carriers of lexical meaning. It is therefore difficult to provide a translation equivalent in a target language which does not distinguish the same rich morphological structure.

In Tables 1 and 2, a summary of the treatment of two grammatical formatives, i.e., *a* and *sa* in six Sepedi paper dictionaries is provided. These two formatives have a high degree of homography and therefore represent a variety of different grammatical functions. Being a grammatical formative, the equivalent of which does not exist in either Afrikaans or English, it follows that instead of a translation equivalent, a paraphrase or a full description, akin to a definition, is the only option available to the lexicographer to provide the user with the kind of information that will ensure a successful lookup. Rather than refer in this regard to sense distinctions (since grammatical formatives generally do not express a 'sense' in the usual sense of the term), the focus in this discussion is on the identification of the various grammatical functions carried by *a* and *sa* in the various dictionaries.

In Table 1, the different functions of *a* as distinguished by various Sepedi paper dictionaries are listed. Where the tick mark is in brackets, it indicates that the specific function is not identified as such, but treated by means of an example.

As is evident from Table 1, Endemann succeeds in identifying most of the grammatical functions of the formative *a*, as distinguished in modern grammars and collectively treated in the other dictionaries. When compared to the other dictionaries, the only obvious omissions are the past tense function and possibly the function as an ideophone. The function of *a* as formative of the consecutive is not mentioned, but neither does any of the other dictionaries refer to this function, except Ziervogel and Mokgokong (1975: 1). The same pertains to *a* as relative pronoun of class 1 and a phonetic description of *a*. Two

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

functions that are only treated by Endemann are those of *a* as verbal ending and as copula. The reference to *a* as negative particle is however rather dubious: it is described as a negative particle in the negation of the expression *a-e!* 'No!' (pronounced in a staccato manner) (Endemann 1911: 39).

Treatment of the formative *-a*						
	Endemann 1911	**De Schryver 2007**	**Kriel 1967**	**Kriel 1983**	**Kriel et al. 1989**	**Ziervogel & Mokgokong 1975**
s.c. class 1	√	√	x	√	√	√
s.c. class 6	√	√	√	x	√	√
o.c. class 6	√	√	x	x	√	x
p.c. class 6	(√)	√	x	√	√	√
demonstrative class 6	√	√	√	√	√	√
present tense morpheme	√	√	x	√	√	√
question particle	√	√	x	x	√	√
hortative particle	√	√	x	√	√	√
past tense morpheme	x	√	x	x	√	x
exclamation / interjection	√	x	√	√	√	√
ideophone	x	x	√	√	x	(√)
relative pronoun class 1	x	x	x	x	x	√
relative pronoun class 6	√	x	x	√	x	√
possessive formative	√	x	x	x	x	√
phonetic description	x	x	x	x	x	√
formative of consecutive	x	x	x	x	x	√
-a as verbal ending	√	x	x	x	x	x
copula	√	x	x	x	x	x
negative morpheme	√	x	x	x	x	x

Table 1: Treatment of the grammatical formative a across six paper dictionaries

The treatment of the formative *a* far exceeds its treatment in the other dictionaries, not only in terms of the distinction between the different functions of this formative, but also in terms of dictionary space allocated to the treatment of *a* as lemma. Each page in this dictionary contains two columns, and three and a quarter column, spread over two pages, are dedicated to the treatment of *-a*. The only comparable treatment is that of Ziervogel and Mokgokong (1975), who use roughly one and a half column for their treatment of *a*. The encyclopaedic information that is provided by Endemann is extremely detailed, with copious examples being provided, cf. in this regard the description of *-a* as possessive formative (translated from the original German; the examples have been transcribed to follow the standard orthography):

> Genitive particle, more correctly possessive particle, since in Sotho, and in the Bantu languages in general, the particular relationship is always a possessive relationship and indicates belonging, for which reason Sotho and Bantu generally do not form a genitive object (Genetivus objectivus); where it (the term) is currently used, it is a language corruption of which Europeans are guilty which must be rooted out — The subject pronoun of the logical governing noun ("regens") is prefixed to the possessive particle. This situation leads to the possessive particle being identical to the verbal root *a* and indeed in it lies the idea of direction towards

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

the following governed noun ("rectum"). Its high tone correlates with the emphasis of the connection with the governed noun. Here, the prefixed pronoun acquires a relative character — The following may serve as examples of the possessive construction according to the different noun classes:

mohlanka wa morena "servant of the Lord"
batho ba (= ba-a) kgoši "people of the chief"
modumo wa lewatle "roaring of the sea"
melao ya mmuši "commandments of the ruler"
botho bja (= bo-a) tate "humanity of (i.e., my) father"
leina la (= le-a) ngwana "name of the child"
mala a kgomo (= a-a) "intestines of the cow"
seedi sa (= se-a) letšatši "light of the sun"
nku ya (= e-a) mohumi "sheep of the rich person"
diphoofolo tša (= di-a) naga "wild animals of the veld"

(Endemann 1911: 39)

The second formative *sa* was selected because it appears towards the end of the alphabet. Endemann's dictionary was compiled according to the so-called traditional method, which means that compilation was based on the lexicographer's intuition without proper attention being paid to lexicographic planning and the formulation of a strategy for inclusion and/or exclusions of lemmas, cf. Kosch (2011: 112-113) and Prinsloo and De Schryver (2007: 179). Entering lemmas as the lexicographer comes across them leads to a number of macrostructural inconsistencies, such as over-treating the initial sections or alphabetical categories and under-treating categories dealing with the letters appearing towards the end of the alphabet (Prinsloo and De Schryver 2007: 186). Apart from checking the extent of the treatment of *sa* as grammatical formative, focusing on *sa* can therefore also provide initial insight to which extent Endemann succumbed to lexicographer's fatigue when compared to other dictionaries compiled using the same compilation strategy. Compare the table below, and once again, tick marks appearing in brackets indicate that the specific function is not identified, but treated by means of an example:

Treatment of the formative -*sa*						
	Endemann 1911	**De Schryver 2007**	**Kriel 1967**	**Kriel 1983**	**Kriel et al. 1989**	**Ziervogel & Mokgokong 1975**
s.c. class 7	x	√	x	√	x	√
p.c. class 7	x	√	√	√	√	√
aspectual prefix	√	√	√	√	√	√
negative morpheme	√	√	x	(√)	√	√
verb stem	√	√	√	√	√	√
(adverb)	x	x	√	x	x	x
(formative of past tense)	x	x	x	√	x	x
ideophone / interjection	x	x	x	√	√	√

Table 2: Treatment of the grammatical formative sa across six paper dictionaries

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

The first obvious seeming omission is the non-treatment of *sa* as (consecutive) subject concord of class 7. This is an unfortunate omission, since this function is probably the most prominent one of this formative. Although Endemann (1911: 456) does not distinguish *sa* as possessive particle, it does figure in his treatment of *-a* as base for possessive concords, where an illustrative example is also provided — see the excerpt above. It is however highly unlikely that a user will look up the lemma *-a* when looking for the meaning of *sa*. He does not treat *sa* as ideophone, but neither do De Schryver (2007) or Kriel (1967). Kriel's (1967: 355) distinction of *sa* as adverb is debatable, as is Kriel's (1983: 267) distinction of *sa* as a formative of the past tense. As a whole, Endemann's treatment of the formative *sa* compares well with that of the other dictionaries. There does not seem to be any evidence of lexicographer's fatigue in his treatment of this particular formative — this preliminary conclusion is further explored in the next paragraph.

Over and/or under treatment of alphabetical categories

In order to test the preliminary conclusion regarding over and/or under treatment of alphabetical categories in Endemann's dictionary further, a random selection of pages from both Endemann (1911) and Ziervogel and Mokgokong (1975) was made, and the number of lemmas treated on each page was recorded. The results are reflected in Figure 1:

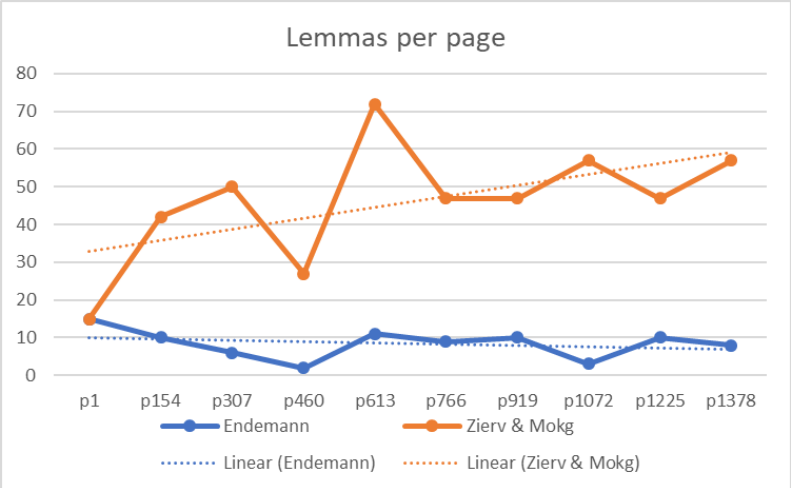


Figure 1: Lemmas per page in Endemann (1911) and Ziervogel and Mokgokong (1975)

As can be seen from Figure 1, the number of lemmas per page in Endemann's dictionary stays relatively stable as one moves through the alphabetical catego-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

ries. The number of lemmas treated in Ziervogel and Mokgokong however, follows a steadily increasing trajectory. The stability in the number of lemmas per page in Endemann (1911) is in sharp contrast with Prinsloo and De Schryver's (2007: 186) findings with regard to, for example, Kriel's (1983) dictionary. They indicate that the number of articles on page 2 of the dictionary is 22, whereas this number is 75 on page 281, thus indicating an extremely reduced treatment of lemmas as the lexicographer moved towards the end of the alphabet.

The results obtained from this experiment are quite remarkable, considering the resources that were at the lexicographer's disposal during compilation of the dictionary, and also, the fact that he was simultaneously learning the language as he was collecting the data. Endemann collected the data used in his dictionary during the 12 years that he spent in South Africa — from 1861 to 1873 — doing missionary work in Sekhukhuniland and Botshabelo near Middelburg (Kosch 2011: 112). To say that he had no access to the electronic and technological resources available to modern-day lexicographers would be stating the obvious. Furthermore, Kosch (2011: 113) points out that the collection of the data was not carried out with the express purpose of compiling a dictionary, that Endemann had no formal lexicographic training, and that his data was probably mostly based on orally collected data, since very few written texts were available at that stage. Even so — as is clear from Figure 1 above — he managed to avoid one of the pitfalls to which even modern-day lexicographers are no strangers, cf. Prinsloo and De Schryver (2002) for the over treatment of the alphabetical category **K** in the WAT.

Treatment of high frequency lemmas

Inclusion of lemmas based on their frequency of use is a feature of modern corpus-based lexicography. When the lemma lists of corpus-based dictionaries are compared to frequency lists drawn from a corpus, these dictionaries usually fall seriously short, in that high frequency lemmas are not included. Compare De Schryver and Prinsloo (2000: 294-297) for a detailed discussion of macrostructural deficiencies regarding the omission of high frequency lemmas in selected Sepedi dictionaries. Including lemmas based on frequency considerations rests on the assumption that users will look up high frequency words, an assumption which has been substantiated by De Schryver and Joffe (2004: 190). Including high frequency items therefore increases the probability that dictionary users' look-up needs are met. For the purpose of this discussion, a frequency list was extracted from a 7.5 million-word Sepedi corpus. The assumption was that a comparison of the topmost frequent items on the frequency list with the lemma list of the dictionary would not deliver significant results, since (a) many of the top frequency items are homographic grammatical formatives, and (b) being so frequent, it is highly unlikely that these items would not cross the lexicographer's path. The probability of entering these lemmas into the dictionary is therefore high, even when the lemma list is compiled based on the intuition of the lexicographer. This assumption is borne out by the results of a small experiment in

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

which the top 100 nouns and verbs were extracted from the frequency list. On this list 67 items are nouns, 33 are verbs. Since the frequency list is a raw one, i.e., an unlemmatised list, it also contains plural nouns. For the purpose of the experiment, plural nouns were counted as being treated if they appear in the article of the corresponding singular noun. Of the 67 nouns, only four found on the frequency list are not treated in Endemann's dictionary, i.e., *baithuti* 'learners' (frequency rank 150), *Afrika* 'Africa' (frequency rank 188), *sekolo* 'school' (frequency rank 199) and *tšhelete* 'money' (frequency rank 202). Of the 33 verbs, only *-rile* 'said' is not treated, although it is referred to in the encyclopaedic information provided for the lemma *-re* 'say'.

Consequently, the top 100 derived verb stems were manually isolated from the frequency list. The lowest frequency rank included in the list is 1034, thus drilling down much deeper than isolating the first 100 most frequent nouns and verbs. Since the corpus is a raw corpus, i.e., not annotated with part-of-speech, only items that are non-ambiguous were included in the list. Three categories of treatment are distinguished: items that are treated, items that are not treated, but are referred to within the article by means of an example or as part of the grammatical information provided, and items that are not treated in the dictionary at all. Compare Figure 2 in this regard:

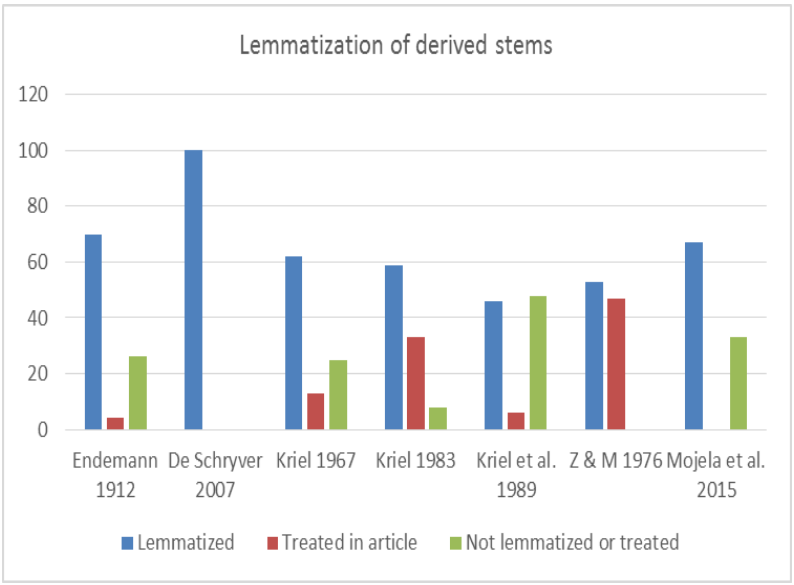


Figure 2: Treatment of derived verbs stems across seven dictionaries

From Figure 2 above, it is clear that Endemann's dictionary outperforms five of the other dictionaries with regard to the treatment of high frequency items. The full score obtained by De Schryver's dictionary is due to the fact that this is a

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

corpus-based dictionary, and the selected derived verb stems all made the cut-off point, frequency wise, for inclusion in the dictionary. In terms of lemmatisation of high frequency items, the dictionary of Mojela et al. (2015) seems to compare well with De Schryver (2007) and Endemann (1911). Endemann (1911) treats 70 of the most frequent items; Mojela et al. (2015) treat 67; however, the actual treatment of the lemmas in these two dictionaries is vastly different, and this is mostly due to the over-simplified microstructure of the dictionary of Mojela et al. (2015). For the lemma *-makatša* 'surprise' for example, the latter provides the following information: part of speech, which is incorrectly indicated as a noun, and a translation equivalent 'surprise', incorrectly spelled as 'suprise'. Endemann provides grammatical information by indicating that it is the causative form of the base form *-makala*, with 'amaze' (*in Erstaunen setzen*) as translation equivalent. With regard to the lemma *-amogela*, Mojela et al. once again provide the part of speech, and also three translation equivalents, i.e., 'receive' (incorrectly spelled as 'receice'), 'welcome' and 'accept'. Endemann again provides grammatical information by indicating that it is an applicative form of the base form *amoga*, followed by four translation equivalents. i.e., 'relieve someone of a load', 'receive hospitably', 'catch (something which is thrown)' and 'echo', followed by a usage example and its idiomatic translation, followed by a literal translation of the example. Generally speaking, Endemann's dictionary therefore compares very well with existing Sepedi dictionaries with regard to treatment of high frequency items.

Completion of semantically related paradigms / lexical sets

Rundell (2015: 302) points out that it is common practice for dictionaries to cover all members of any clearly defined set of lexical items, citing days of the week and signs of the Zodiac as examples of such lexical sets. We are aware of the questions raised by Swanepoel (2010: 429 et seq.) with regard to the identification of such sets, the first being how they are to be identified. Swanepoel's discussion is aimed at the identification and utilisation of lexical sets in the crafting of better definitions, but since this is not the focus of our investigation, we follow Atkins and Rundell (2008: 123) who define a lexical set as "groups of words that share a common element of meaning such as days of the week, or months of the year, or birds, trees, flowers, and metals." By implication, lexical sets are based on numerous sense relations such as synonymy and hyponymy, or the fact that they belong to the same semantic domain or field, see Swanepoel (2010: 429). The days of the week is a prototypical example of a lexical set, therefore we used that as a first topic to ascertain to what extent Endemann was aware of the standard lexicographic practice of completing semantically related paradigms. Once again, we used the selected Sepedi dictionaries listed above as a benchmark. The outcome of this first small investigation was rather surprising. All dictionaries lemmatise the names of all seven weekdays, except Endemann (1911). None of the weekdays appear in his dictionary. However, the time frame dur-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

ing which his data collection took place needs to be considered — it could be that the Western concept of a week as a time unit consisting of seven days had at that stage not yet been conceptually established in the Sepedi-speaking community. It is also significant that the Sepedi word for 'week' does not appear in the dictionary. As Kosch (2011: 114) remarks: "the dictionary was compiled … to capture the uncorrupted forms of the language from a time before contact with Europeans". Furthermore, it is unlikely that the names of the weekdays had at the time of the compilation of the dictionary been standardised, considering that the first official attempt at standardising the orthography, terminology and spelling rules of Sepedi was only undertaken in 1930 (Kosch 1993: 23).

A second lexical set, i.e., the four seasons was consequently investigated, also with interesting results. The four official terms as provided in the Terminology and Orthography no 4 of 1988 are *marega* 'winter', *seruthwana* 'spring', *selemo* 'summer' and *lehlabula* 'autumn'. Apart from Endemann (1911), all six the other dictionaries treat these four terms; Endemann only treats three of them and provides the following translation equivalents: *marega* 'winter', *selemo, lehlofo* 'spring, early summer' and *lehlabula, hwetla* 'late summer, autumn'.[1] There are however, considerable differences with regard to the translation equivalents provided by the other dictionaries. The only term that all dictionaries agree on is the one for 'winter' Kriel (1967), Kriel (1983) and Kriel et al. (1989) provide 'autumn' as a translation equivalent of *seruthwana*, whereas the other dictionaries give 'spring' as an equivalent. There seems to be considerable overlap between *selemo* and *seruthwana*, with both having 'spring', 'early summer' and 'ploughing time' (the latter in the case of Kriel et al. (1989) and Ziervogel and Mokgokong (1975)) as translation equivalents. It may very well be that the distinction of four seasons is a case of cultural projection, where the Western conceptualisation of four seasons is projected onto the Sepedi linguistic system, leading to a mismatch between what could be described as an indigenous knowledge system and the treatment of its members in modern Sepedi dictionaries. In indigenous knowledge systems, seasons are based on weather patterns, and in the case of agriculture-based cultures on agricultural activities in a given area, and therefore do not necessarily correspond to the Western, meteorologically based notion of four seasons. In Nyungar, an aboriginal language spoken in Southwest Australia, 6 seasons are distinguished: season of the young *birak* (dry and hot, burning time), of adolescence *bunuru* (hottest part of the year), of adulthood *djeran* (cooler weather begins), of fertility *makuru* (coldest and wettest season of the year), of conception *djilba* (mixture of wet days with increasing number of clear, cold nights and pleasant warm days) and of birth *kambarang* (longer dry periods) (http://www.bom.gov.au/iwk/calendars/nyoongar.shtml); the ancient Hindu calendar also distinguishes six seasons: spring, summer, monsoon, autumn, pre-winter and winter (https://www.javatpoint.com/seasons-in-india), and in countries close to the equator, a distinction is made between the wet season and the dry season. The possibility that Endemann's treatment of terms referring to only three seasons (winter, spring and early summer, and late summer and autumn) could in actual fact be a better reflection of the African conceptual

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

system regarding seasonal distinctions should therefore not be summarily dismissed. As a matter of interest, Colenso's *Zulu–English Dictionary* (1905), compiled around the same time as Endemann's, only includes entries for summer (*ihlobo*) and winter (*ubusika*). However, under the entry for *ihlobo* (summer), the season is further subdivided into five distinct parts. Like contemporary Sepedi dictionaries, modern Zulu dictionaries also adopt the Western concept of four distinct seasons.

A third lexical set that was perused was that of body parts, specifically facial features. There is no fixed definition as to what exactly constitutes facial features, so for the purpose of this study we selected the following: *sefahlogo/sefahlego* 'face', *phatla* 'forehead', *leihlo* 'eye', *nko* 'nose', *molomo* 'mouth', *thama/lerama* 'cheek', *ntšhi* 'eyelashes, eyebrows' and *seledu* 'chin'. All dictionaries, except De Schryver (2007), treat all these items; De Schryver misses out on the last three.[2] This is again due to the fact that his dictionary is frequency-based; however, including these lemmas even though they did not make the cut-off point frequency-wise, should have been considered. Once again, Endemann holds his own when compared to other Sepedi dictionaries with regard to the completion of lexical sets, despite the lacunae discussed above.

Treatment of culturally significant lemmas

Endemann's (1911) treatment of lemmas is characterised by an extremely rich sense distinction and detailed definitions of lemmas, the latter especially with regard to culturally-bound lemmas. Apart from providing a translation equivalent, an extended paraphrase of meaning is provided in some cases of lemmas where a cultural element is present. Since Ziervogel and Mokgokong (1975) is the only other dictionary that is more or less comparable to Endemann (1911) with regard to comprehensiveness, the treatment of the lemmas as provided in the former is provided by way of comparison. A few illustrative examples will suffice. (The examples from Endemann (1911) have been translated from the original German and transcribed to follow the standard orthography; only translation equivalents and paraphrases of meaning are provided):

1. *lesiba*

   name of a musical instrument; it is a flat bow strung with a string, at one end a long quill is attached, this end of the bow is put in the mouth and the sound is produced by means of the quill, by inhaling and exhaling through the mouth; one hears two tones (Endemann 1911: 473).

   kind of musical instrument on which one blows (Ziervogel and Mokgokong 1975: 1177).

2. *mongala*

   a protester, someone who resists laws and customs, an uncircumcised person (who resists circumcision), a deserter (from circumcision), who is forever

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

banned (as a result of his resistance against customs), an instigator (Endemann 1911: 339).

Not treated in Ziervogel and Mokgokong (1975).

3. *sedimo*

oracle; ghostly being that can be heard, but not seen. An animal that is slaughtered so that its entrails can be used for divination, is also called thus (Endemann 1911: 273).

offering, sacrifice (to spirits) (Ziervogel and Mokgokong 1975: 156).

4. *sekgapa*

name of a musical instrument, consisting of and bow and string, the resonance is created by means of a calabash that is fastened in the middle of the bow, on which the string rests, and it gives two tones (Endemann 1911: 209).

calabash, stringed musical instrument, † bass. (Ziervogel and Mokgokong 1975: 604).

5. *tšilo*

millstone; (i.e., a fist stone, a hewn, fist-sized, rounded stone, with which grain is ground on a grinding stone) (Endemann 1911: 595).

grinding stone, mill (Ziervogel and Mokgokong 1975: 1218).

Even though not extensive, these examples do give a sense of the cultural richness to be found in Endemann (1911), making the dictionary a valuable cultural artefact, and as such, worth not only of preserving, but also of making it accessible for modern day users.

Inaccessibility of Endemann (1911)

Despite comparing favourably with existing Sepedi dictionaries, Endemann (1911) is not accessible to modern day users. There are several reasons for this, the least not being that it has been out of hard-copy print for more than a century. The publication rights currently reside with De Gruyter Mouton (Verlag) and an e-version of the dictionary was published in 2012 and is available at a cost of €229[3], which makes it quite prohibitively expensive for South African users. The format of the dictionary is an unsearchable pdf-version, which is not particularly useful. A second reason for the inaccessibility of the dictionary is the language pair (Northern) Sotho–German, which is a rather unusual one, and it narrows the pool of potential target users substantially. The biggest stumbling block for the modern-day dictionary user is however the orthography used by Endemann and, as a result, the unusual ordering of alphabetical categories. In the introduction of the dictionary (Endemann 1911: 1-8), a detailed description of the orthography is provided in which the use of specific orthographic sym-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

bols to represent certain sounds is explained, but this information is provided in German only and requires an in-depth knowledge of phonetics. Without having read the introduction, the hapless user would seek in vain for the word *-bala* in an alphabetical stretch B following A; the bilabial fricative [β] is represented by the symbol v, and is therefore to be found between the alphabetical stretches U and X (the semi-vowel [w] is represented as ŏa and is to be found under O), where X is used to represent the velar fricative [ɣ], in the modern orthography represented by g. These challenges are exacerbated by the fact that no index appears in the front matter. Lastly, the left-expanded article structure is problematic, especially from a user-perspective, and in the case of disjunctively written languages, more so in the case of nouns than of verbs. In a left-expanded lemmatisation procedure, the stem is the alphabetical point of reference, similar to a purely stem-based lemmatisation procedure. The main difference between left-expanded and stem-based lemmatisation strategies is that in the case of the former, the prefixal element is included as part of the lemma sign, albeit in a slot preceding the stem. Compare the excerpt from the alphabetical stretch E in Figure 3 by way of illustration:

le-ėma, Pl. *ma-*, „Horde, Stamm, Parteilichkeit (S.-S.); Zaun vom *kχoro* (N.)“ Der Sinn des W. ist der eines Dinges, das individuell (für sich) Stellung einnimmt.

se-ėma, Pl *li-*, „Einfriedigung, Dornzaun vom *kχoro*, Pfahlhütte mit offenen Wänden (Laube) im *kχoro;* Spruch, Sprichwort (also „stehende Redensart“), Wort (einzelnes), Meinung; Versammlung von stehenden Menschen (S.-S.)“. *χo sava ligma* „witzeln“, *χo sava ligma kà* „bewitzeln“.

se-ęma (S.-S.) „Dickleibigkeit, Wassersucht“. Vergl. *ima*.

ėmàėma, Iterat. v. *ema*, „immer stehen bleiben, zögern, säumen“.

se-emahale (S.-S.) „Ding, welches unbeweglich feststeht“ (z. B. Grenzstein, Denkmal) Von *ema* + *hale* (= *χale* „lange“)

Figure 3: Excerpt from the alphabetical stretch E (Endemann 1911: 60).

Note that there is vertical alignment on *e*, which is a feature of a true left-expanded article structure (Gouws and Prinsloo 2005: 43). In order to look up

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

any noun, the user first needs to identify the stem of the noun, and as Gouws and Prinsloo (2005) and Van Wyk (1995: 89) point out, stem identification is problematic in itself, and requires sound morphophonological knowledge of the user, since the prefix morphology of nouns is not only irregular, but is also subject to fairly complex morphological rules. As a result, "it is simply not possible for either the user or the lexicographer to determine unambiguously what the form of the isolated stem is" (Gouws and Prinsloo 2005: 33). This can lead to inconsistency and deviation from the stated editorial policy with regard to lemmatisation. In Endemann (1911) for example, verb stems to which the object concord of the first-person singular *N-* has been prefixed, are not treated at all, whereas reflexive verb stems with the prefix *i-* are treated under the letter I, which does not represent a left-expanded article structure, where lemmas are treated under the first letter of the stem. Nouns in class 9 are particularly problematic: nouns which display a nasal prefix *N-* are treated under the relevant nasal, e.g. *nku* 'sheep' under N, *mpho* 'gift' under M and *ngaka* 'doctor' under NG, whereas nouns that do not display a visible class prefix are lemmatised under the first letter of the noun, e.g. *tau* 'lion' under T and *kgomo* 'cow' under KG, despite the fact that the stem of the latter is actually *-gomo*. These inconsistencies impede the successful retrieval of information from the dictionary.

Digitisation of the dictionary

In the last section of our article, we map out the steps we follow for the digitisation of the dictionary. The ideal option would be to transform the complete paper dictionary to a completely digital dictionary, displaying the features of a true electronic dictionary. However, the researchers have been granted permission by the publisher, De Gruyter Brill, to publish only parts of the dictionary electronically, without any cost. Most of the challenges outlined above can potentially be solved by transforming the original dictionary to a fully-fledged electronic dictionary. Digitisation is the first step in this transformation process.

There is currently no digital or searchable version of the dictionary, which requires the use of OCR (Optical Character Recognition) to extract all the dictionary entries. Furthermore, the dictionary features an obsolete, rather idiosyncratic, phonetically-based orthography, which needs to be converted to the modern orthography. We decided to use OCR4all (https://www.ocr4all.org/, Reul et al. 2019), an open source tool that is free to use. OCR4all offers the following functions:

— Segmentation: this process involves dividing a scanned document into its constituent parts, such as separating text from images or distinguishing individual characters, words, and lines. Segmentation ensures that the OCR system accurately identifies the structure of the document.

— Recognition: after segmentation, recognition is the core OCR function where the tool analyses the segmented parts and converts the text from image format into machine-readable text. This is where characters and words are identified based on patterns.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

— Ground Truth Production (GTP): this refers to creating a manually corrected version of the recognised text to serve as a reference or "ground truth." This step ensures high accuracy by comparing the OCR output with the verified, corrected version, enabling further improvements in OCR performance.

For the digitisation of the Endemann dictionary, the following steps were followed:

Step 1: Segmentation for OCR processing. The scanned pages of the dictionary are divided into individual sections (characters, words, lines) in preparation of the OCR process. During segmentation each page is divided into two columns, and the texts to be recognised are marked in advance for several pages. Information such as page numbers and headings which form part of the access structure, are ignored. Segmentation is initially performed on 10 to 20 pages, and after several rounds of training, the system suggests the parts of the text to be recognised with high accuracy.

Step 2: Training of a Custom Recognition Model. In this phase, a specialised OCR model is created, adapted to the unique font used in the dictionary. We start with a pre-trained OCR model from OCR4all, i.e., the "default/deep3_fraktur19/4," which is suitable for recognising Fraktur fonts commonly used in historical texts. This is followed by incremental improvement by means of Ground Truth Production (GTP). First, the pre-trained model is used to produce the initial OCR output. This is followed by manual correction of the initial 10 to 20 pages to correct recognition errors, specifically those related to the specialised phonetic symbols and unique fonts used in the dictionary. The OCR model is then retrained, using the corrected output, refining its accuracy incrementally with each cycle of corrections. This iterative process, also called 'bootstrapping' assists the model to adapt to the specific characteristics of the dictionary, leading to improved accuracy over time.

Step 3: Generation of a full-text output. A .txt file is generated, containing the complete digital format of the dictionary. Segmentation information is also exported — segmentation details such as the structure of pages, words and characters are exported in .xml format, thus preserving the layout and organization of the dictionary for future use.

Step 4: Post-processing. This entails the application of shell scripting to clean and refine the OCR4all-produced output text, ensuring accuracy and formatting consistency. Shell scripting refers to writing a series of commands for the shell to execute. A shell script is essentially a file that contains multiple instructions, which can automate tasks that would otherwise be performed manually by typing commands one by one. In the context of post-processing for dictionary digitisation, shell scripting can be used to:

— Automate corrections of formatting issues.
— Perform batch replacements of specific characters or patterns (e.g., correct-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

ing OCR errors).
— Organise and structure the output text.
— Extract specific data or manipulate files (like splitting or merging).
— Add page numbers that had to be ignored in the OCR-process.
— Separate the lemmas from the rest of the article (see 1 in Figure 5)
— Re-join words that are separated by a hyphen (see 2 in Figure 5)

Shell scripting is particularly useful for processing large amounts of text data efficiently and repetitively, making it ideal for refining the dictionary's digital text after completion of the OCR process. Figure 4 presents an illustration of the way in which entries are organised while Figure 5 displays the output resulting from the OCR process.

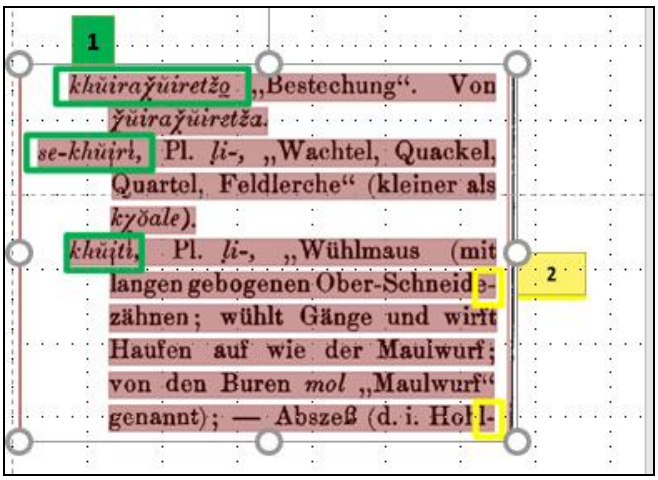


Figure 4: Organizing the entries

khiirahuiretšo "Bestechung". Von huirahuiretša.

se-khiiri , Pl. di-, "Wachtel, Ouackel, Ouartel, Feldlerche" (kleiner als kgoale).

khiiti , Pl. di-, "Wühlmaus (mit langen gebogenen Ober-Schneidezähnen; wühlt Gänge und wirft Haufen auf wie der Maulwurf; von den Buren mol "Maulwurf" genannt); - Abszeß (d. i. Hohlgang); - Viehweide (Feld, wo die Wühlmaus baut?)".

Figure 5: Output resulting from OCR

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

The digitisation process is not flawless, but the overall results are surprisingly good. There is for example an overall error rate of 0.2% in the OCR results, therefore each article should be checked before copying. This error rate is calculated by OCR4all and would probably need to be further evaluated. Also, the model does not recognise curved brackets and articles grouped together by means of these brackets are erroneously split into multiple articles. Compare the following example shown in Figure 6, and the OCR-results in Figure 7:

khuinoñ = *khunoñ*, } N., „männliches
khuinoù = *khunoù*, } rotes oder rot-
braunes Stück Vieh". Vergl.
-χuveḷu.

Figure 6: Articles grouped together by means of curved brackets by Endemann

khuinon = khunon, N., "männlicheskhuinou =
khunou, rotes oder rotbraunes Stück Vieh". Vergl. -
hubedu.

Figure 7: Resulting OCR output

After completion of the OCR process a number of working groups were formed. The German–English working group consists of German speakers who collect data from the OCR files and who also add English translations. The English–Sotho working group is composed of speakers of the Sotho languages and dialects and they are responsible for quality assurance. This working group can comment on Endemann's interpretation of the Sotho data. The third working group, the Sotho–English–Afrikaans group is tasked with translation of the English text into Afrikaans. All data is stored in pre-database tables to ensure an automated transfer into the database which will be utilised by the planned web-interface.

Conclusion and future work

The wealth of lexical, cultural and linguistic information contained in Endemann's *Wörterbuch der Sotho Sprache* (1911) makes it an invaluable resource for resource scarce languages, and it is therefore imperative that this resource should not only be preserved, but that it should also be made accessible to a wider tar-

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

get audience, especially to the speakers of these languages. There are still some unresolved questions with regard to the dictionary itself, due to the lack of historical information regarding its actual compilation. The use of the term 'Sotho' is one such issue. It is not clear what the exact reference of this term is, whether it refers to what is traditionally known as Northern Sotho (Sesotho sa Leboa), encompassing a number of dialects spoken in the southern and central parts of the Limpopo province, or rather to Sepedi, a dialect spoken in Endemann's area of missionary work.

Digitisation is but the first step towards producing a fully-fledged, electronic dictionary containing selected articles from the dictionary. By utilising full electronic features of such dictionaries, many of the challenges outline above, e.g. restricted access and difficult access structure can be addressed. The next step is the creation of a database, containing the proposed microstructure of the electronic version. The digitised data in Endemann's dictionary will be used as a basis with the addition of further data categories, e.g. English and Afrikaans translation equivalents, and selected grammatical information.

Endnotes

1. Examples from Endemann are transcribed into the standard orthography and are also provided with translation equivalents in English.
2. However, even though these three items are not treated in the central list, they are treated in the back matter of the dictionary by way of illustrations and bilingual captions.
3. See https://www.lehmanns.de/shop/geisteswissenschaften/32384232-9783111418513-woerterbuch-der-sotho-sprache-sued-afrika

References

Dictionaries

Colenso, J.W. 1905. *Zulu–English Dictionary.* Second Edition. Pietermaritzburg: Shuter & Shooter.

De Schryver, G.-M. 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English*. Cape Town: Oxford University Press Southern Africa.

Endemann, K. 1911. *Wörterbuch der Sothosprache*. Abhandlungen des Hamburgischen Kolonialinstituts, Band VII. (Reihe B. Völkerkunde, Kulturgeschichte und Sprachen Band 4). Hamburg: L. Friedrichsen & Co.

Kriel, T.J. 1967. *The New English–Northern Sotho Dictionary*. King William's Town: Educum.

Kriel, T.J. 1983. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho.* Third edition. Pretoria: J.L. van Schaik.

Kriel, T.J., E.B. van Wyk and S.A. Makopo. 1989. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Fourth, revised and expanded edition. Pretoria: J.L. van Schaik.

Mojela, V.M., M.C. Mphahlele, M.R. Selokela and W.M. Mojapelo. 2015. *Sesotho sa Leboa–English Bilingual Dictionary*. Cape Town: Phumelela Books.

Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho Woordeboek*. Pretoria: J.L. van Schaik.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
https://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2050 (Article)

Other sources

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

De Schryver, G.-M. and D. Joffe. 2004. On How Electronic Dictionaries are Really Used. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 187-196. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

De Schryver, G.-M. and D.J. Prinsloo. 2000. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 1: The Macrostructure. *South African Journal of African Languages* 20(4): 291-309.

Gouws, R.H and D.J. Prinsloo. 2005. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.

Kosch, I. 1993. German-speaking Pioneers in African Linguistics and Literature with Special Reference to Northern Sotho. *South African Journal of African Languages* 13 (Supplement 2): 2-5.

Kosch, I. 2011. Innovation and Compromise in K. Endemann's Dictionary of the Sotho Language (1911). *South African Journal of African Languages* 31(1): 110-120.

Kosch, I. 2012. Challenges of Predictability and Consistency in the First Comprehensive Sotho Dictionary. *Lexikos* 22: 226-242.

Lombard, D.P. 1970. Bantoetaalstudie: Dictionaries in Northern Sotho. *Bantu Education Journal* 16(3): 12-13.

Prinsloo, D.J. and G.-M. de Schryver. 2002. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with Special Reference to Afrikaans and English. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 483-494. Copenhagen: Center for Sprogteknologi, Københavns Universitet.

Prinsloo, D.J. and G.-M. de Schryver. 2007. Crafting a Multidimensional Ruler for the Compilation of Sesotho sa Leboa Dictionaries. Mojalefa, M.J. (Ed.). 2007. *Rabadia Ratšhatšha: Studies in African Language Literature, Linguistics, Translation and Lexicography:* 177-201. Stellenbosch: SUN PReSS.

Reul, C., D. Christ, A. Hartelt, N. Balbach, M. Wehner, U. Springmann, C. Wick, C. Grundig, A. Büttner and F. Puppe. 2019. OCR4all — An Open-source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings. *Applied Sciences* 9(22): 4853.

Rundell, M. 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.

Swanepoel, P. 2010. Improving the Functionality of Dictionary Definitions for Lexical Sets: The Role of Definitional Templates, Definitional Consistency, Definitional Coherence and the Incorporation of Lexical Conceptual Models. *Lexikos* 20: 425-449.

Tarp, S. 2013. Old Wisdom: The Highly Relevant Lexicographical Knowledge Obtainable from a Specialized Dictionary from 1774. *Lexikos* 23: 394-413.

Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)
http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2065

Publikasieaankondigings / Publication Announcements

Edward Finegan and Michael Adams. *The Cambridge Handbook of the Dictionary*. 2024, 801 pp. ISBN 978-1-108-83624-1 (Hardback), ISBN 978-1-108-81887-2 (Paperback), ISBN 978-1-108-86443-5 (eBook). Cambridge/New York, NY: Cambridge University Press. Price: US$ 170.00 (Hardback), US$ 170.00 (eBook). (Review in this issue.)

Pedro A. Fuertes-Olivera. *A Guide to Practical Online Lexicography.* 2025, 334 pp. ISBN: 9781032731544 (Hardback), 9781003426998 (eBook). London/New York: Routledge. Price: US$ 180.00 (Hardback).

Yongwei Gao. *A Dictionary of Contemporary English Neologisms and Trendy Words.* 2024, 406 pp. ISBN: 978-7-5213-5810-0 (Paperback). Beijing: Foreign Language Teaching and Research Press. Price $9.32. (Review in this issue.)

Du Ruiqing and Jiang Yajun. *New Century Chinese–English Dictionary*. Third Edition. 2025, XXXVIII + 2490 pp. ISBN 978-7-5213-5418-8 (Hardback). Beijing: Foreign Language Teaching and Research Press. Price $54.43. (Review in this issue.)

Stephen Turton. *Before the Word was Queer: Sexuality and the English Dictionary, 1600–1930.* 2024, xv + 333 pages. ISBN 978-1-316-51873-1 (Hardback). Cambridge: Cambridge University Press. Price: US$ 94.67 (Hardback). (Review in this issue.)

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

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 enigeen 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, Posbus 245, 7599 STELLENBOSCH, Republiek van Suid-Afrika.

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat versoenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie **8 000 woorde** oorskry nie.

Elke artikel moet voorsien wees van 'n **opsomming** van ongeveer 200 woorde en ongeveer 10 **sleutelwoorde** in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutelwoorde **in Engels**. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigeen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n **vertaling van die oorspronklike titel** bevat.

2. Grafika

Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting. Buiten in spesiale gevalle moet verwysings na *Lexikos*-artikels tot twee of drie per artikel beperk word. Uitsonderings moet met die redakteur van *Lexikos* uitgeklaar word. Dít word gedoen om die status van *Lexikos* in verskeie internasionale indekse te behou.

4. Aantekeninge/voetnote/eindnote

Aantekeninge moet deurlopend in die vorm van boskrifte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

http://lexikos.journals.ac.za; https://doi.org/10.5788/35-1-2067 (Volle uitgawe / Full volume)

INSTRUCTIONS TO AUTHORS

(*For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website:* http://lexikos.journals.ac.za/)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
(4) **Projects:** Discussions of lexicographical projects.
(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
(6) **Lexicovaria:** Any of a large variety of articles containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
(7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index* (ISI).

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free, however, to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

6. Free offprints and copies

Lexikos is only available electronically on http://lexikos.journals.ac.za from volume 28 onward. No offprints or copies are available.

7. Invitation and editorial address

All interested authors are invited to submit contributions, preferably in electronic format, for publication in *Lexikos* to: lexikos@sun.ac.za, or

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH, Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Format and length of articles

Manuscript should preferably be submitted in electronic format by email or on a disk, in software compatible with MS Word. The typeface used should preferably be 10-point Palatino or Times Roman. Contributions should not exceed **8 000 words**.

Each article must be accompanied by **abstracts** of approximately 200 words and approximately 10 **keywords** in the language in which it is written, as well as **in English**. English articles of South African origin should carry an abstract and keywords in Afrikaans, whilst English articles of foreign origin should carry a second abstract and keywords in any of the other languages mentioned. In cases where this is not done, the editors will provide an Afrikaans version. Ensure that the abstract in the second language also contains a **translation of the original title**.

2. Graphics

Figures such as tables, graphs, diagrams and illustrations should be in an appropriate size to be well accommodated within the page size of *Lexikos*, namely 18 cm high by 12 cm wide. The locations of figures within the text must be clearly indicated. If orthographic marks or graphics used in the text prove problematic, a printout of the manuscript or an email in .pdf format may be requested.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details. Self-references to *Lexikos* should be limited to two or three per article, except in exceptional circumstances. Exceptions should be cleared with the editor of *Lexikos*. This is done to preserve the status of *Lexikos* in various international indices.

4. Notes/footnotes/endnotes

Notes must be numbered consecutively by superscript numbers and grouped together at the end of the manuscript under the heading **Endnotes**.