

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2108> (Volle uitgawe / Full volume)

Lexikos 35(2)

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2108> (Volle uitgawe / Full volume)

Lexikos 35(2)

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 35A:2025

AFRILEX SERIES 35A:2025



BURO VAN DIE WAT

STELLENBOSCH

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2108> (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 978-0-6398803-0-3
ISSN 2224-0039

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

Licensed under Creative Commons License CC BY 4.0.

Menings wat in artikels en resensies uitgespreek word, is nie noodwendig dié van AFRILEX of die Buro van die WAT nie.

Opinions expressed in the articles and reviews are not necessarily those of AFRILEX or of the Bureau of the WAT.

Lexikos is elektronies beskikbaar by <http://lexikos.journals.ac.za/>

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

Lexikos is elektronies beskikbaar by Sabinet, AJOL, Ebsco en Proquest

Lexikos is available online from Sabinet, AJOL, Ebsco and Proquest

Indekse Indexes

Arts and Humanities Citation Index® (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)

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)

Inhoud / Contents

Voorwoord	ix
Foreword	xii
<i>André H. du Plessis</i>	
'n Woord van AFRILEX	xv
A Few Words from AFRILEX	xvi
<i>Langa Khumalo</i>	
Redaksionele doelstellings	xvii
Editorial Objectives	xviii

Artikels / Articles

Die gebruik van kleur in aanlyn woordeboeke, met oorweging van drie Afrikaanse woordeboeke	1
<i>Maryna Adshade</i>	
Thematic Classification in Polish Electronic Dictionaries — From Tradition to New Tools and Opportunities	302
drie Afrikaanse woordeboeke	
<i>Barbara Batko-Tokarz</i>	
Generative Artificial Intelligence (GenAI) as Information Tool for Lexicographic Information Needs	276
<i>Theo J.D. Bothma and Rufus H. Gouws</i>	
Gender Bias in Computer-generated Thesauri: The Case of the Serbian Section of <i>Kontekst.io</i> , a Thesaurus of Synonyms and Semantically Related Terms	26
<i>Dragana Čarapić and Milica Vuković-Stamatović</i>	
Out-of-the-Box GPTs for Lexicography	362
<i>Gilles-Maurice de Schryver</i>	
Customised GPTs for Lexicography	397
<i>Gilles-Maurice de Schryver</i>	

'n Omvattender aanbod van semantiese data in aanlyn woordeboeke <i>Rufus H. Gouws</i>	46
Lexicographic Description of a Polysemous Word in a Learner's Dictionary Based on Its Lexical Prototype <i>Anton Karasev</i>	75
Exemplification of Sensitive Words for People with Disabilities in Monolingual English Learner's Dictionaries <i>Meimei Li</i>	94
Lexicography in Action: The Traversal from Coinage and Iconicity to Iconisation <i>Dorra Moalla Masmoudi</i>	117
Reconceptualising Contemporary Lexicography: Emotional Intelligence and the Sustainable Lexicographic Modelling of the Modal Verb <i>moći</i> <i>Milena Mrdak-Mićović and Nataša Jovović</i>	143
Aktuelle Aussprachewörterbücher des Deutschen und des Englischen im Vergleich <i>Krzysztof Nycz und Zygmunt Tęcza</i>	232
Spelling Variations in <i>A Dictionary of Setswana Personal Names</i> <i>Thapelo J. Otlogetswe</i>	324
Making Dictionary Content Accessible for People with Visual Impairments <i>Geraint Paul Rees</i>	165
One Name, Two Genres: The Curious Case of Polish Encyclopaedias of Law <i>Agnieszka Rzepkowska</i>	185

Projekte / Projects

Where Tools Are Few: Constructing Limited-Word Bilingual Learner's Dictionaries in a Low-Resourced Community in Malawi <i>Ian David Dicks</i>	256
--	-----

Lexicographic Post-processing and Refinement Issues in Wordnets: The Ukrainian Wordnet Experience <i>Maksym Vakulenko</i>	345
---	-----

Leksikonotas / Lexiconotes

A Systematic Literature Review of Over Three Decades of Dictionary Use Research in <i>Lexikos</i> <i>Balázs Fajt, Mátyás Bánhegyi and Katalin P. Márkus</i>	423
Noms de Marques dans la 9 ^e Edition du <i>Dictionnaire de l'Académie Française</i> <i>Jan Holeš et Zuzana Honová</i>	215

Leksikohuldeblyk / Lexicotribute

Frederik Johannes Lombard 1953 – 2025 <i>Alet Cloete, Phillip Louw en/and Gerhard van Wyk</i>	296
--	-----

Publikasieaankondigings / Publication Announcements

Voorskrifte aan Skrywers	444
Instructions to Authors	445

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 uitdaginge 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 verloop 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

met GLOBALEX en waardeur die redaksionele en koördineringswerk wat die gasredakteurs van hierdie afdeling, prof. Annette Klossa-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 meta-leksikografie 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 publiserings 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.

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

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.

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

'n Woord van AFRILEX

Na aanleiding van 'n goed georganiseerde 29ste kongres van die *African Association of Lexicography* (AFRILEX) wat by die Universiteit van 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 posities 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

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

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

Die gebruik van kleur in aanlyn woordeboeke, met oorweging van drie Afrikaanse woordeboeke

Maryna Adshade, *Buro van die WAT, Stellenbosch,
Suid-Afrika; Departement Afrikaans en Nederlands,
Universiteit Stellenbosch, Stellenbosch, Suid-Afrika*
(adshadem@sun.ac.za)

(<https://orcid.org/0009-0002-1491-7488>)

Opsomming: Wanneer die gebruik van kleur in sommige aanlyn Afrikaanse woordeboeke oorweeg word, wil dit voorkom asof kleur soms oneffektief of lukraak ingespan word. In ander gevalle kom dit voor asof kleure gekies word om vinnig met die handelsmerkidentiteit van die woordeboekuitgewer geassosieer te word eerder as ter wille van praktiese doeleindes. Dit kan moontlik weens abstrakte of vae voorstelle vanuit die metaleksikografie wees wat onbevredigende leiding aan praktiese leksikograwe bied, maar dalk ook 'n gevolg van die Suid-Afrikaanse, en Afrikaanse, leksikografie wat relatief stadig tot die Westerse wêreld ontwikkel. Die gebruik van kleur word in hierdie artikel bespreek met oorweging van 'n paar voorstelle wat reeds in metaleksikografiese navorsing gelug is. Hierdie voorstelle word aangevul met enkele riglyne ontleen aan die inligtingstechnologie. Die toepassing, al dan nie, van die teoretiese moontlikhede word bespreek ten opsigte van die aanlyn weergawe van die *Handwoordeboek van die Afrikaanse Taal*, *Aanlyn Woordeboek van die Afrikaanse Taal* en die *Woordeboek van Afrikaans Vandag*. Daar is ruimte vir verbetering in sekere aspekte van hierdie woordeboeke se kleurgebruik, en heelwat van die opwindende, meer gevorderde moontlikhede wat deur 'n beter kleurgebruik ingehou word, is nog toekomstmusiek binne die Afrikaanse leksikografie. Nietemin is dit opmerklik dat die gebruik van kleur belowend is en vorentoe 'n groter rol in toeganklikheid kan speel.

Sleutelwoorde: KLEURGEBRUIK, AANLYN WOORDEBOEKE, DATATOEGANKLIKHEID, WOORDEBOEKSTRUKTURE, ARTIKELSTRUKTUUR

Abstract: The Use of Colour in Online Dictionaries, with Consideration of Three Afrikaans Dictionaries. When considering the use of colour in some online Afrikaans dictionaries, it seems that colour is sometimes used ineffectively or haphazardly. In other cases, colours seem to be chosen mainly to align with the dictionary publisher's brand identity, rather than for practical purposes. This may result from abstract or vague metalexicographical suggestions that offer unsatisfactory guidance to practical lexicographers but could also be the result of South African, and Afrikaans, lexicography developing slower compared to the Western world. This article discusses the use of colour in online dictionaries, considering a few suggestions from metalexicographical research. These suggestions are then supplemented with some guidelines derived from

the field of information technology. The application of the theoretical possibilities, or lack thereof, is discussed with regard to the online version of the *Handwoordeboek van die Afrikaans Taal*, *Aanlyn Woordeboek van die Afrikaans Taal* and the *Woordeboek van Afrikaans Vandag*. These dictionaries show room for improvement in some aspects related to colour usage, with many exciting and more advanced possibilities still a future prospect for Afrikaans lexicography. Nevertheless, it is noteworthy that the use of colour is promising and may play a greater role in accessibility in the future.

Keywords: COLOUR USAGE, ONLINE DICTIONARIES, DATA ACCESSIBILITY, DICTIONARY STRUCTURES, ARTICLE STRUCTURE

1. Inleiding

Dit is algemeen bekend dat die leksikografiepraktyk deur die ontwikkeling en geskiedenis daarvan reeds verskeie uitdagings en paradigmaverskuiwings moes deurstaan (sien Gouws 2017: 430-438). Dit is ook goed bekend dat dit tans opnuut deur 'n enorme paradigmaverskuiwing uitgedaag word, naamlik die oorgang van die gedrukte na die elektroniese medium, veral na die aanlyn sfeer. In die leksikografiepraktyk se geskiedenis is hierdie oorgang onbetwisbaar die ingrypendste oorgang nóg (Gouws 2018c: 178) en dit gaan sodoende uiteraard met nuwe uitdagings gepaard. Dit is teweeggebring deur die koms van die era van elektroniese dataverwerking, en kan sedertdien toegeskryf word aan 'n bestendige stroom tegnologiese deurbreke wat gelei het tot radikaal nuwe benaderings in onder meer hoe leksikografiese produkte gefinansier, voorberei, saamgestel, gebruik en aangebied word (Gouws en Tarp 2024: 190).

Dit is juis op 'n aspek van die laasgenoemde vlak van vernuwing waarop hierdie artikel toegespits is. Soos Almind (2005: 37) aandui, is daar ongelukkig uitgewers (en dalk selfs leksikograwe) wat glo dat hulle 'n behoorlike aanlyn woordeboek kan publiseer deur gewoon 'n gedrukte woordeboek te digitiseer en 'n soekenjin op die gebruikerskoppelvlak van daardie digitale weergawe aan te bring. Hy (ibid.) meen dat dié uitgewers verkeerdelik veronderstel dat potensiële woordeboekgebruikers daarmee tevrede sal wees om 'n gedrukte ontwerp op 'n skerm te sien vertoon. Dit lei tot twee probleme: Daar word nie daarmee tred gehou dat 'n aanlyn woordeboek 'n produk in eie reg is nie, en as bemarkingstrategie sal die benadering misluk bloot omdat dit woordeboekgebruikers sal teleurstel (Almind 2005: 37). Hy is daarom oortuig daarvan dat leksikograwe by die beplanning van aanlyn woordeboeke dit eerder van meet af as sodanig moet beplan en nie slegs 'n gedrukte woordeboek aanlyn beskikbaar moet stel nie (Almind 2005: 37). Die grootste rede hiervoor kan toegeskryf word aan een kernbegrip: toeganklikheid (Almind 2005: 38).

Hoewel reeds twee dekades gelede gelug, geld dié siening dat baie aanlyn woordeboeke gewoon gedigitiseerde gedrukte woordeboeke is, vandag steeds heelwat aanlyn woordeboeke. Boonop bly dit, ondanks meer onlangse vordering, besonders relevant in die Suid-Afrikaanse (en Afrikaanse) leksikografielandskap, waar die ontwikkeling op vele terreine steeds stadiger as verwag vorder

en die aanlyn leksikografie veral agter dié van Westerse lande bly (Beyer en Louw 2022: 41). Met inagneming hiervan het hierdie artikel ten doel om een ontwerpelement van aanlyn woordeboeke, naamlik die gebruik van kleur, te bespreek. Die doel is ook om te wys op 'n paar opwindende moontlikhede en om enkele voorstelle omtrent suksesvolle aanwending van kleur te maak.

Om die bespreking hiervan te kontekstualiseer, moet die gebruik van kleur eers kortliks relatief tot die voorkoms van aanlyn woordeboeke en woordeboekstruktuur oorweeg word. Ná die teoretiese bespreking word daar gekyk hoe kleur in drie Afrikaanse woordeboeke aangewend word, naamlik die aanlyn weergawe van die sesde uitgawe van die *Handwoordeboek van die Afrikaanse Taal* (voortaan HAT Aanlyn), die *Aanlyn Woordeboek van die Afrikaanse Taal* (voortaan Aanlyn WAT)¹, en die *Woordeboek van Afrikaans Vandag* (voortaan WAV).

2. Teoretiese beskouinge oor die gebruik van kleur in woordeboeke

Voordat die gebruik van kleur in woordeboeke bespreek kan word, is dit nodig om daarop te wys dat kleurgebruik nie slegs 'n kosmetiese aanvulling tot die data-aanbieding in woordeboeke behoort te wees nie. Kleur moet eerder as nuttige hulpmiddel in 'n woordeboek se strukturele aanbod ingespan word om te verseker dat aanlyn woordeboeke as funksionele en praktiese gebruiksinstrumente kan dien. Kleur moet dus beskou word as 'n praktiese middel tot 'n doel en 'n funksionele element in woordeboekartikels.

2.1 Die verband tussen woordeboekstrukture en kleurgebruik

Woordeboeke, ook aanlyn woordeboeke, kan beskou word as draers van tekssoorte, oftewel groot tekste opgebou uit kleiner, gerangskikte tekste (Wiegand 1996: 135-136). Hierdie tekste moet só aangebied word dat die struktuur daarvan die doel van die tersaaklike tekste kan vervul (Wiegand 1983: 104). Dit beteken dat dit só gerangskik moet word dat woordeboeke as praktiese gebruiksinstrumente kan dien om woordeboekgebruikers se leksikografiese behoeftes te bevredig (Wiegand 1998: VII, 62).

Vir woordeboeke om nutsprodukte te wees, moet daar volgens Gouws (2018b: 217-218) by die beplanning van enige woordeboek, ongeag die medium, met drie integrale komponente tred gehou word: leksikografiese funksie(s), woordeboekinhoud en woordeboekstrukture. Die behoeftes en naslaanvaardighede van die beplande woordeboek se beoogde teikengebruikers behoort eerstens die funksie(s) te bepaal, waarna die inhoud op grond van die funksie(s) vasgestel moet word (sien Gouws 2018b: 217-218). Uiteindelik moet die inhoud op 'n gepaste wyse aangebied word deur die inspanning van verskillende relevante woordeboekstrukture (Gouws 2018b: 217). Gepaste woordeboekstrukture moet dus in oorleg met funksies en inhoud bepaal word.

Uit hierdie komponente is die eerste twee redelik geyk in die metaleksikografie, en in 'n hoë mate tans nog vanselfsprekend omdat dit grotendeels onver-

anderd in aanlyn woordeboeke geld. Die derde komponent kan egter te midde van die huidige paradigmaverskuiwing onsekerheid en probleme lewer, want gedrukte woordeboekstrukture kan nie altyd min of meer onveranderd in aanlyn woordeboeke behou word nie (Gouws 2018b: 218). 'n Besinning hieroor, en die faktore wat dit kan beïnvloed, is daarom nodig omdat dié strukture voorsiening maak vir leksikografiese data (sowel binne of buite woordeboekartikels) sodat woordeboekgebruikers data kan opspoor (Gouws 2023: 383). Hieruit is woordeboekstrukture duidelik onontbeerlik — daarsonder sou woordeboeke, ongeag die gehalte van die woordeboekinhoud, onbevredigende produkte wees omdat woordeboekgebruikers nie vinnig en ondubbelsinnig hulle verlangde inligting daaraan sou kon onttrek nie (De Schryver 2003: 173).

Hier is die fokus nie op 'n omvattende bespreking van woordeboekstrukture nie. Daar word ook aanvaar dat heelwat van die hieropvolgende terminologie algemeen gangbaar is sodat alles nie in diepte bespreek word nie. Om die rol en die aanbod van kleurgebruik in woordeboekstrukture te verstaan, is dit nietemin nuttig om kortliks na kernstrukture en gidsstrukture te verwys, asook en veral na die interne soekstruktuur as tipe gidsstruktuur.

Alhoewel die onderskeid nie 'n verabsoluttering is nie en oorvleuelings tussen en binne die kategorieë kan bestaan, kan kernstrukture as die hoofstrukture van woordeboeke beskou word terwyl gidsstrukture die sekondêre vlak uitmaak. Kernstrukture hou verband met die beskikbaarheid en posisionering van leksikografiese data (Mihindou 2006: 163) en sluit onder meer die makrostruktuur, mikrostruktuur en dataverspreidingstruktuur in (vgl. Gouws 2005: 63). Hierteenoor geld gidsstrukture byvoorbeeld die toegangstruktuur, mediostruktuur en sekere buitetekste (vgl. Gouws 2001: 104; Gouws 2005: 63), oftewel daardie strukture wat woordeboekgebruikers help om data op te spoor en van die woordeboekraadplegingsproses sin te maak (Mihindou 2006: 163).

Dit is spesifiek by gidsstrukture, soos die toegangstruktuur, waar kleur in aanlyn woordeboeke 'n groter rol kan en moet speel, en waar leksikograwe doelgerig te werk moet gaan om dit te bereik. Die toegangstruktuur bied die soekroete wat woordeboekgebruikers volg wanneer hulle woordeboeke raadpleeg om by die verlangde woordeboekartikel uit te kom (Gouws en Prinsloo 2005: 165). Dit funksioneer in samewerking met ander strukture — veral die dataverspreidingstruktuur, mikrostruktuur, mediostruktuur en artikelstruktuur — en het die belangrike doel om woordeboekgebruikers met so min as moontlik moeite die benodigde data te laat bereik (Gouws en Prinsloo 2005: 172). Daar kan onderskei word tussen 'n eksterne en interne soekstruktuur, maar daar word vervolgens slegs op die laasgenoemde gefokus. Vir meer oor die eksterne soekstruktuur kan Gouws en Prinsloo (2005), asook Wiegand en Beer (2013) geraadpleeg word.

By die interne soekstruktuur gaan dit oor die soekroete wat woordeboekgebruikers binne woordeboekartikels volg om die verlangde datatipe of datakategorie te bereik (Gouws en Prinsloo 2005: 170) sodat hulle aan hulle onmiddellike inligtingsbehoefte kan voldoen (vgl. Tarp 2012: 328). Soos Wiegand en Beer (2013: 139) dit stel, ondersteun die interne toegangstruktuur woordeboekgebruikers om hulleself binne artikels te oriënteer nadat hulle eksterne tekstuele toegangs-

elemente gevolg het om die tersaaklike artikel te bereik.

Hier is dit nodig om weer te verwys na De Schryver (2003: 173) se punt dat woordeboeke onbevredigende produkte sal wees as woordeboekgebruikers nie vinnige, ondubbelsinnige toegang tot data kan verkry nie. Gouws en Prinsloo (2005: 170, 172) se opmerking moet hiermee in verband gebring word wanneer hulle tereg noem dat woordeboekgebruikers selde die hele woordeboekartikel wil lees om die verlangde inligting te onttrek, en daarom meer waardering het vir woordeboeke waarin hulle vinniger toegang tot die verlangde data kan kry. Hieruit is dit duidelik dat die interne toegangstruktuur noodsaaklik is om vinnige toegang tot relevante data te verseker, en soos Wiegand en Beer (2013: 139) aandui, des te meer in langer gekondenseerde woordeboekartikels. Dit is ook hier waar kleur, indien dit effektief en funksioneel ingespan word, 'n waardevolle rol in vinniger, ondubbelsinnige toegang kan speel.

2.2 Moontlikhede vir die gebruik van kleur as toegangselement

Die vinniger en ondubbelsinnige toegang tot data in woordeboekartikels hou naas die posisionering van data verband met hoe data aangebied word, oftewel die voorkoms van 'n woordeboekartikel. Die leksikografie is nie geïsoleerd nie, soos duidelik blyk uit die oorweging van tegnologie en die ontwikkeling daarvan in die huidige paradigmaverskuiwing, soos genoem in die inleiding. In die vinnig ontwikkelende digitale wêreld waar mense se digitale geletterdheid toenemend verbeter, is dit daarom 'n wesenlike kwessie dat woordeboeke ontwikkel moet word om steeds relevant en bruikbaar te wees, ook ten opsigte van hoe dit visueel by ander eietydse digitale bronne aansluit.

Soos Gouws (2018c: 179-180) aandui, is dit nodig om te besef dat heelwat potensiele toekomstige woordeboekgebruikers tot nuwe geslagte, soos Generasie Z, behoort en hierdie geslagte volwaardige digitale burgers is vir wie die internet 'n gegewe is. Die vanselfsprekendheid van die internet in hulle lewe, asook in dié van baie lede van ander generasies, beïnvloed hoe hulle met digitale hulpbronne werk en wat hulle verwagtinge daarvan (en van die voorkoms daarvan) is. Dit is dus vir lewensvatbaarheidsredes noodsaaklik om tred te hou met gebruikers se vorderende of gevorderde digitale geletterdheid, asook hulle verwagtinge van digitale bronne. In die huidige digitale era waartydens digitale burgers dikwels bederf is met vinnige, moeitelose toegang tot 'n oorvloed inligting (vgl. Gouws en Tarp 2017: 394), sal woordeboeke wat outyds lyk of grotendeels 'n gedrukte woordeboek op 'n skerm is, onbevredigend wees en dalk in onbruik verval. Met die merkwaardige tempo waarteen kunsmatige intelligensie ontwikkel en die gebruikers daarvan van pasgemaakte oplossings kan verskaf, is dit nou dalk meer toepaslik as ooit vantevore.

Aanlyn woordeboeke moet daarom, soos reeds genoem, as 'n produk in eie reg beskou word en meer wees as 'n elektronies beskikbare gedrukte woordeboek (vgl. Almind 2005: 37). As aanlyn produkte kan daar tekens van vernuwing en ooreenkomste met bepaalde internetpraktyke wees, ook ten opsigte

van die ontwerp en algemene voorkoms van woordeboeke. Rundell (2015: 320) stel byvoorbeeld dat leksikograwe kan aanvaar dat woordeboekgebruikers vertrouwd is met internetbeginsels, soos soekstrategieë en 'n sekere "woordeskat", maar dit geld myns insiens ook vaardighede om internetbronne te kan navigeer, waarin kleur 'n rol kan speel. Hierdie veronderstelling word bevestig deur Müller-Spitzer, Koplenig en Töpel (2012: 441) wat daarop wys dat woordeboekgebruikers 'n behoefte het aan aanlyn woordeboeke wat bekendheid toon met ander toepassings en webformateringsbeginsels.

Een vorm hiervan is die gebruik van kleur volgens bepaalde formateringsbeginsels. Swasty en Adriyanto (2017: 17-18) wys byvoorbeeld vyf ontwerpbegin-sels van goeie gebruikerskoppelvlakke uit, waarvan een die goeie gebruik van kleur is. Kleur kan 'n aanwys in gebruikerskoppelvlakke se ontwerp wees (Jamil en Denes 2024: 53) deurdat dit 'n wesenlike rol in estetiese waarde, bruikbaarheid en toeganklikheid kan speel (WCAG 2024). Alhoewel die hoofdoel van aanlyn woordeboeke uiteraard nie is om estetiese waarde te hê nie, bly estetika 'n faktor omdat gebruikers se persepsie van gebruikerskoppelvlakke se estetiese waarde hulle persepsie van die bruikbaarheid daarvan kan beïnvloed. In hierdie verband verwys Alharoon en Gillan (2020: 1876) na studies wat 'n positiewe korrelasie toon tussen skynbare en inherente bruikbaarheid, dit wil sê hoe gebruikers 'n koppelvlak se gebruiksgemak skat soos gebaseer op voorkoms (nog voordat hulle dit gebruik het), teenoor die vasstelling van gebruiksgemak nadat hulle met die koppelvlak gewerk het. Dus kan aanlyn woordeboeke wat onafgerond, oudmodies of andersins onesteties is gebruikers ontmoedig om die woordeboek te gebruik. Dit sluit weer aan by Almind (2005: 37) se siening dat 'n slaafse navolging van gedruktewoordeboekbeginsels as bemarkingstrategie kan misluk omdat woordeboekgebruikers teleurgesteld kan wees in die produk.

Die kwessie van estetika en kleur as 'n belangrike ontwerpbegin-sel is dus ook op woordeboekontwerp van toepassing. Soos Mihindou (2006: 165) aantoon, verkies woordeboekgebruikers daardie woordeboeke waarin kleur gebruik word omdat kleur aantrekliker en meer ekspressief is. As 'n aanwys behoort kleurgebruik egter nie net kosmetiese waarde te hê nie en dit moet funksioneel wees. Hier is Müller-Spitzer en Koplenig (2014: 144) se siening dat vorm die funksie in 'n woordeboek moet volg, relevant. Hierbenewens moet kleur in samewerking met ander elemente gebruik word en kleur kan nie die enigste manier wees waarop sekere inligting oorgedra word nie.

In die interne toegangstruktuur kan verskillende kleure woordeboekgebruikers toelaat om vinniger en makliker toegang tot die verlangde data te verkry (Mihindou 2006: 165). Eweneens wys Kilgarriff en Kosem (2012: 49) daarop dat kleurgebruik navigasie kan vergemaklik deur woordeboekgebruikers te help om tussen verskillende datatipes te onderskei. Hulle noem nie spesifieke voorbeelde nie, maar verskillende kleure kan byvoorbeeld as struktuurmerkers gebruik word om soeksones in aanlyn woordeboeke te ekspliseer en duidelik van ander af te baken. Kleur kan sodoende tot 'n meer gebruikersvriendelike mikro-argitektuur bydra deurdat sekere elemente visueel van mekaar onderskei en in spesifieke kleure geblok word. Kundige woordeboekgebruikers met 'n goeie woor-

deboekkultuur kan sodoende vinniger data opspoor deur die oog gewoon na die verlangde deel te laat gly. Die voorwaarde is egter dat 'n verduideliking van die betekenis van spesifieke kleure in die gebruikersgids, 'n sleutel onderaan die artikels, of iets dergeliks ingesluit moet word.

Benewens die skeiding en duidelike afbakening van verskillende soeksones, kan kleur woordeboekgebruikers help om op data af te kom wat hulle andersins dalk nie (so maklik) sou vind nie. As woordeboekgebruikers met 'n woordeboek se benadering tot kleur bekend is, kan die gebruik van verskillende kleure hulle aandag trek na soeksones wat in eenkleurige artikels tussen ander soeksones sou kon verdwyn (vgl. Mihindou 2006: 165). Hierbenewens kan kleur gebruik word om mediostrukturele verwysings duideliker sigbaar te maak en van omliggende data te onderskei, wat ook gebruikers na meer data as net die onmiddellik verlangde data kan lei. Soos tipies in internetbronne, kan mediostrukturele verwysings wat hiperskakels is as sodanig met kleur aangedui word sodat woordeboekgebruikers met 'n oogopslag kan sien waar kits-toegang tot ander verwante artikels en data moontlik is. Dit kan veral by implisiete kruisverwysings nuttig wees wat andersins maklik misgekyk kan word. Verder kan verskillende kleure gebruik word om woordeboekinterne en woordeboek-eksterne kruisverwysings te onderskei. Hierdie opsies kan in verskillende mates tot politoeganklike en meer gebruikersvriendelike woordeboekartikels bydra.

Kleur kan dus gebruik word om verbande tussen verskillende dele van 'n woordeboek te lê. Mihindou (2006: 166) is byvoorbeeld ten gunste van die gebruik van bypassende kleure sodat woordeboekgebruikers verskillende tekste binne die woordeboek kan identifiseer — spesifiek deur die sentrale teks van buitetekste te onderskei. Kleur kan só as 'n tipe indeks dien omdat gebruikers vinnig kan sien waar verskillende dele van die woordeboek begin. Hierdie voorstel is uiteraard minder ter sake in aanlyn woordeboeke waarin hiperskakelroetes na verskillende woordeboekdele (wat dikwels op 'n navigasiebalk vertoon word) moontlik is. 'n Aanpassing van Mihindou se voorstel kan wel oorweeg word: Bypassende kleure kan binne woordeboekartikels en buitetekste gebruik word om woordeboekgebruikers bewus te maak van waar hulle addisionele data kan vind wat ook nuttig kan wees. Indien woordeboekgebruikers binne 'n bepaalde artikel op 'n spesifieke kleur afkom, kan dit hulle byvoorbeeld na geïntegreerde buitetekste verwys, oftewel daardie buitetekste wat tot 'n woordeboek se werklike doel bydra (Klosa en Gouws 2015: 144). Andersins kan dit die verhouding tussen dele van die sentrale teks en spesifieke buitetekste, of spesifieke dele van bepaalde buitetekste, aandui wanneer byvoorbeeld sekere struktuurmerkers, teksdele, rame om teksdele, of die skermkootagtergrond in spesifieke kleure gemerk word.

In meer gevorderde aanlyn woordeboeke kan kleur ook in artikels met 'n dinamiese veelvlaklike artikelstruktuur 'n waardevolle rol aanneem. Dié artikels kan omvattend of beperk wees: Omvattende artikels is volle artikels waarvan die geheelbeeld nie in 'n enkele skermkoot sigbaar is nie terwyl beperkte artikels dele van omvattende artikels is wat in aparte skermkote verskyn (Gouws 2018a: 52). Binne veelvlaklike artikels kan 'n seektonnel voorkom wat woorde-

boekgebruikers toelaat om tot die beperkte artikels toegang te verkry. Gouws en Bothma (2024: 362) beskryf 'n soektonnel as 'n struktuur waardeur 'n gebruiker stap vir stap deur verskillende artikelvlakke (die beperkte artikels) kan beweeg. Hulle beweeg dan horisontaal of sywaarts deur 'n vlak om verdere kommentaar of subkommentaar binne bepaalde vakke te bereik (Gouws 2023: 386; Gouws en Bothma 2024: 362). Hier sou 'n vertikale beweging tussen verskillende vlakke, via 'n struktuur soos 'n soekskag, ook moontlik wees.

Om deur soektonnels, of ook soekskagte, te beweeg, moet gebruikers eers op die behoefte aan addisionele data dui deur byvoorbeeld op knoppies of ander struktuurmerkers te klik (Gouws 2018a: 50). Vir hierdie beweging, hetsy na verskillende vakke of vlakke, sal 'n meer komplekse koppelvlakontwerp egter benodig word (Gouws en Bothma 2024: 361). Dit is juis by hierdie aanduiding van gebruikers se behoefte aan nog data en in meer komplekse koppelvlakontwerpe waar kleur nuttig kan wees. Dit is omdat bepaalde kleure gebruik kan word om daardie knoppies of struktuurmerkers duideliker te maak. So kan kleur dit duideliker maak waar gebruikers op versoek beperkte artikels kan vertoon, oftewel waar addisionele data beskikbaar is indien die gebruiker dit sou wou sien. Hier kan kleur ook indirek 'n rol speel in gebruikers se gemak en vermoë om verrigte en uitgebreide data te bekom.

By 'n veelvlakkige artikelstruktuur, soektonnels en soekskagte kan kleur nogmaals 'n rol speel om verbande tussen verskillende dele van 'n woordeboek te illustreer. Die addisionele data wat via sodanige soektonnels of soekskagte bekom kan word, word in 'n nuwe skermkoot vertoon wat aan die tersaaklike data-tipe (soos voorbeeldmateriaal, verwante leksikale items, grammatika of illustrasies) toegewy is (Gouws 2018a: 50). Hier kan kleur veral in beperkte artikels binne soektonnels help om byvoorbeeld dele van die sentrale teks aan ander relevante dele te merk, of die sentrale teks aan spesifieke geïntegreerde buitetekste te verbind, soos bo genoem. Daarbenewens kan dit 'n rol in niegeïntegreerde buitetekste speel, dit wil sê daardie buitetekste waaraan woordeboekgebruikers nie data kan onttrek wat tot 'n woordeboek se doel bydra nie (Gouws en Prinsloo 2005: 59), soos 'n woordeboek se beleid of gebruikersgids. In laasgenoemde kan kleur byvoorbeeld wys hoe en waar spesifieke kleure in artikels voorkom om aanduiders en struktuurmerkers te merk. Tydens hulle woordeboekraadpleging kan dit woordeboekgebruikers weer help om struktuurmerkers te sien wat hulle byvoorbeeld via 'n hiperskakelroete of opskietvenster vanuit 'n woordeboek-artikel kan neem na 'n niegeïntegreerde buiteteks of 'n gedeelte daarvan, soos 'n hanteringsbeleid.

2.3 Oorwegingspunte ten opsigte van doeltreffende kleurgebruik

Ofskoon kleur op verskeie maniere geïntegreer kan word, moet dit nie op 'n lukrake wyse geskied en bloot ingesluit word omdat die aanlyn wêreld die opsie bied nie. Dit beteken eerstens dat kleur nie oordadig gebruik moet word nie. Volgens Almind (2005: 39) moet kleur slegs gebruik word om belangrike data

uit te lig, en kleurgebruik moet tot "goeie" (kontrasterende) kleure beperk word. Andersins kan oormatige kleurgebruik vinnig tot 'n sirkusontwerp lei (Almind 2005: 52), wat beteken dat dit byvoorbeeld nes die oorgebruik van tipografiese struktuurmerkers, soos kursief, vinnig steurend kan word. Dit kan lei tot 'n onaantreklike en deurmekaar voorkoms, maar kan ook teksverdigting veroorsaak wat tot 'n stadiger, potensieel dubbelsinnige inligtingonttrekking lei. Die uitdaging by die beplanning van die gebruikerskoppelvlak is dus die fyn balans tussen te veel en te min kleur.

Met verwysing na Almind (2005) se eise aan aanlyn woordeboeke, merk Fuertes-Olivera en Tarp (2014: 196-197) op dat kleur tot 'n elegante, aangename uitleg asook tot woordeboekartikels se leesbaarheid moet bydra deurdat maklik leesbare en kontrasterende kleure gekies moet word. Deur laasgenoemde toe te pas en elke datatipe in 'n nuwe reël te laat begin, kan die leesbaarheid van woordeboekartikels verhoog word (Fuertes-Olivera en Tarp 2014: 197). Dit kan ook 'n gebruikersvriendelike mikro-argitektuur bevorder deur woordeboekgebruikers toe te laat om vinnig by die verlangde data uit te kom.

Betreffende 'n elegante, aangename uitleg kan die keuse van kleure gebaseer wees op die gebruikerskoppelvlakontwerpers se persepsie van estetiese waarde (Jamil en Denes 2024: 54). Dit kan andersins gebaseer word op kleure wat pas by die tersaaklike woordeboek se kleurskema en dit wat daarmee geassosieer word (vgl. Almind 2005: 51). 'n Beperkte kleurpalet as verkose benadering kan om kosmetiese redes aanloklik wees, onder andere as dit pas by 'n woordeboek se beeld en die kleure wat daarmee geassosieer word. In die WAT se geval is die kleuridentiteit byvoorbeeld kenmerkend blou, soos prominent op die WAT se korporatiewe blad (<https://www.wat.co.za/>). 'n Woordeboek se kleuridentiteit kan vir handelsmerkbestuurdoeleindes belangrik wees (vgl. Swasty en Adriyanto 2017: 22), maar by produkte soos woordeboeke waarvan die primêre doel praktiese bruikbaarheid is, moet die funksionaliteit van kleurgebruik ook oorweeg word.

Ten spyte van die moontlike estetiese voordeel, kan 'n beperkte kleurpalet vir sommige woordeboekgebruikers niefunksioneel wees. Dit geld byvoorbeeld woordeboekgebruikers wat sukkel om tussen kleure te onderskei, byvoorbeeld vanweë kleurblindheid, vir wie skakeringe van hoofsaaklik een kleur (anders as swart op wit) of spesifieke kleurkombinasies niefunksioneel of moontlik selfs verwarrend of hinderlik kan wees. Die bedoelde betekenis van die kleure as tipe struktuurmerkers gaan dus verlore — bowenal in die afwesigheid van 'n sleutel of kort verduideliking van die kleure se betekenis op die artikelskermkoot. Daarbenewens kan lae kontrasterende kleure tot 'n aansienlik verminderde estetiese ervaring lei vir gebruikers wat nie so maklik tussen kleure differensieer nie (Jamil en Denes 2024: 54). Dus kan 'n beperkte kleurgebruik vir hierdie woordeboekgebruikers nóg funksioneel, nóg esteties wees.

Die gebruik van kontrasterende kleure is derhalwe belangrik. Nietemin gee nie Almind (2005) of Fuertes-Olivera en Tarp (2014) 'n bevredigende, konkrete verduideliking van wat in woordeboeke relatief tot wat moet kontrasteer nie. Buiten die tradisionele voorbeeld van wit op swart as goeie kleurpaar gee hulle

ook nie verdere voorbeelde van spesifiek goeie kontrasterende kleure sodat dit duidelik is wat dit prakties beteken nie. Dié enkele voorbeeld is onvoldoende, want soos Kilgarrieff en Kosem (2012: 49) aandui, is wit en swart lank reeds nie meer die enigste kleuropsies in moderne woordeboeke nie. Dit kan dalk selfs onbedoeld lei tot die voorkoms van 'n aanlyn woordeboek wat niks meer as 'n gedrukte ontwerp op 'n skerm is nie, waarteen Almind (2005: 37) waarsku. Die vroeër genoemde kwessie van estetika en die skynbare en inherente bruikbaarheid van 'n koppelvlak is ook weer ter sake. In die plek van hierdie tradisionele kombinasie verwys ander na rooi en groen as alternatiewe kleurpaar wat dikwels in gebruikerskoppelvlakke voorkom (vgl. Kilgarrieff en Kosem 2012: 49; Jamil en Denes 2024: 53). Hiermee is daar egter ook 'n probleem, want rooi en groen is juis een van die algemeenste kleurpare wat vir veral kleurblinde gebruikers probleme kan veroorsaak (vgl. Soegoto, Fadhlurrahman en Hermawan 2019: 2).

Daar kan egter voorgestel word dat die kontras wat in aanlyn woordeboeke se kleurgebruik gehandhaaf moet word, op twee vlakke kan geld: eerstens tussen die voorgrond (al die tekstuele dele tesame) en die agtergrond, en tweedens tussen die tekstuele dele op die voorgrond. Betreffende die voorgrond-agtergrond-kontras moet kleure uiteraard maklik onderskeibaar wees sodat die data nie met die agtergrond versmelt nie. Dit blyk dat donker teks op 'n ligte agtergrond vir die meeste mense meer leesbaar is as ligte teks op 'n donker agtergrond — sien in hierdie verband Dobres, Chahine en Reimer (2017: 68-69) se verwysing na verskeie studies waarin dit bevind is. Verder kom dit voor asof swart of donkerblou teks op 'n wit agtergrond makliker lees en meer esteties is as ander kleurpare (Shieh en Lai 2008: 345). Beide hierdie toepassings behoort vir alle woordeboekgebruikers toeganklik te wees, want ook diegene wat soms sukkel om tussen kleurskakeringe te onderskei, kan kontras opmerk (Bigman, in Berisso 2018: 96).

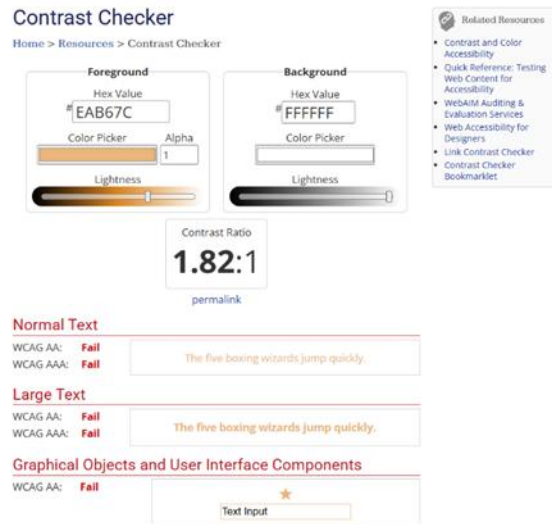
'n Oorweging van die kontras tussen die voorgrond en agtergrond alleenlik verseker egter nie dat alle woordeboekgebruikers die kleurgebruik funksioneel sal ag nie. Daar moet ook kontras tussen komponente van die gebruikerskoppelvlak op die voorgrond wees (Jamil en Denes 2024: 55), oftewel teks en komponente soos struktuurmerkers, aftrekkieslyste, hiperskakels en knoppies om data te versteek of te onthul. Die maklikste manier om te verseker dat alle woordeboekgebruikers tussen voorgrondkleure kan onderskei, is om sekere kleurkombinasies in die geheel te vermy (Berisso 2018: 96) en eerder 'n sogenaamde kleurblinde kleurpalet te gebruik. 'n Kleurblinde kleurpalet maak tegelyk vir alle woordeboekgebruikers voorsiening, dus sowel gebruikers wat sukkel om tussen bepaalde kleure te onderskei as diegene wat nie probleme daarmee ervaar nie (vgl. Soegoto et al. 2019: 5). Daar is dus geen party wat deur die toepassing daarvan benadeel word nie. Prakties behels dit die vermyding van rooi-groen-kombinasies, of geel en blou saam omdat dit algemeen bekend is dat veral kleurblinde mense kan sukkel om tussen hierdie kleurkombinasies te onderskei (Berisso 2018: 94). Verder word helder en donker kleure in 'n kleurblinde palet in kombinasie gebruik sodat mense makliker tussen verskillende kleure kan differensieer (Soegoto et al. 2019: 5); die verskil tussen dié kleure is met die eerste

oogopslag meer opmerklik, en helder kleure is boonop makliker onderskeibaar as dowwer kleure, wat geneig is om te vermeng (Bigman, in Berisso 2018: 96).

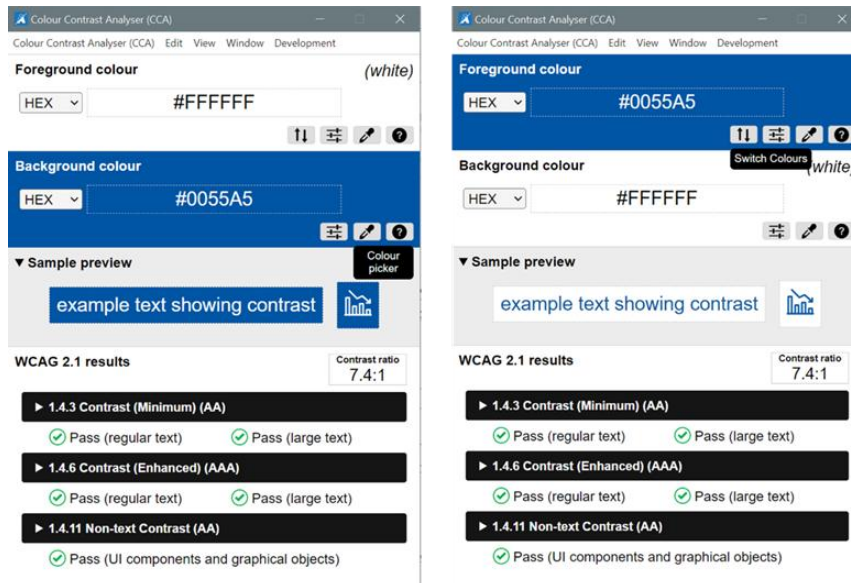
Met inagneming van die vroeër verwysing na aanlyn woordeboeke se aansluiting by webformateringsbeginsels, kan riglyne vir hoër inhoudstoeganklikheid op gebruikerskoppelvlakke geraadpleeg word om die gebruik van kontrasterende kleure meer konkreet te maak. Die Webinhoudtoeganklikheidsriglyne (Web Content Accessibility Guidelines in Engels, oftewel WCAG 2024) gee byvoorbeeld riglyne om toegankliker, meer gebruikersvriendelike webblaaie moontlik te maak en is gerig op kwessies wat met waarneembare en gebruikaspekte van webblaaie verband hou. Een tegniek vanuit die riglyne (sien riglyn 1.4.3 in WCAG 2024) is die gebruik van minstens 'n 4,5:1-kontrasverhouding tussen die teks en agtergrond by gewone teks (dit wil sê, teks kleiner as 18 pt indien dit nie vetgedruk is nie, of 14 pt indien vetgedruk), of 'n 3:1-verhouding by groter teks (groter as 18 pt, soos in opskrifte). Hierdie 3:1-verhouding behoort volgens die riglyne (1.4.11 in WCAG 2024) ook van toepassing te wees op komponente in die voorgrond, soos die teks en knoppies van die koppelvlak. Dit is slegs onnodig om hierdie voorgrondkontras te handhaaf wanneer komponente soos knoppies onaktief is (dus nie interaksie van die gebruikerskant af vereis is) of waar dit deur die gebruiker gewysig is om anders te lyk as wat die webontwerper in gedagte gehad het (WCAG 2024).

Dit is hier nodig om daarop te wys dat riglyn 1.4.3 en 1.4.11 tipies in samehang met 1.4.6 beskou word. Dié riglyn is 'n strenger toepassing van 1.4.3 en gerig op kontrastoeganklikheid vir gebruikers met geweldige sigprobleme. Toepassing daarvan sal vereis dat gewone teks 'n 7:1-kontras (in plaas van 'n 4,5:1-kontras, soos by riglyn 1.4.3) het, en groter teks oor 'n 4,5:1-kontras (teenoor die 3:1-kontras van 1.4.3) beskik. In woordeboeke waarin teks dikwels gekondenseerd en dig is, kan hierdie riglyn egter tans nog redelik ingewikkeld wees om te handhaaf. Daar word daarom hier voorgestel dat woordeboekopkoppelvlakontwerp aanvanklik op riglyn 1.4.3 en 1.4.11 gerig is, met latere toepassing van riglyn 1.4.6 indien nodig. Laasgenoemde word dus nie geïgnoreer nie, maar is dalk iets wat eerder mettertyd ingewerk kan word wanneer woordeboekplatforms werklik volwaardige aanlyn produkte word. Dit is dalk ook iets wat eerder met 'n skakelaar of skuifknoppie geaktiveer kan word sodat die gebruiker self kan bepaal of hulle teks en elemente volgens 1.4.11 geformateer wil hê.

Om hierdie riglyne se voorgestelde kontrasverhoudings prakties en maklik te kan meet, kan verskeie hulpbronne geraadpleeg word waarop 'n mens kleurcodes kan invoer om te kyk of dit aan die Webinhoudtoeganklikheidsriglyne sal voldoen. Dié platforms of toepassings gee dan vinnig terugvoer om te wys of die kleurpaar geskik kan wees of nie (sien Figuur 1 en 2 hieronder). Enkele voorbeelde sluit in TPGi se Colour Contrast Analyser (2025), WebAIM (2025), of Luminosity Colour Contrast Ratio Analyser deur Juicy Studio (2025), wat almal gratis is om te gebruik. Ander hulpbronne, ook gratis, gee weer breedvoerige voorbeelde en palette en stelle bypassende en webveilige kleure wat sigbaar behoort te wees — sien in hierdie verband byvoorbeeld die Index of Color Contrast Samples (s.a.).



Figuur 1: Die kleurpaar slaag nie die kontrastoets nie, soos bepaal deur WebAIM (2025) se aanlyn hulpbron



Figuur 2: Hier slaag die kleurpaar die kontrasriglyne, soos getoets op TPGi se Colour Contrast Analyser (2025) wat gebruikers toelaat om met 'n druppertjie die kleurpaar te toets, en maklik met die druk van 'n ander knoppie te ruil

Hierdie kwessie moet in verband gebring word met Müller-Spitzer en Koplenig (2014: 164) se siening dat woordeboeke sover moontlik volgens 'n hindernisvrye ontwerp saamgestel moet word. Soveel moontlik gebruikers moet die koppelvlak se ontwerp dus toeganklik vind (Jamil en Denes 2024: 53), aangesien toeganklikheid direk tot funksionaliteit bydra. Dit is veral belangrik dat woordeboeke as gebruiksinstrumente vir alle gebruikers voorsiening moet maak indien moontlike aanpassings geen party benadeel nie.

3. Kleur in die HAT Aanlyn, Aanlyn WAT en WAV

In hierdie afdeling word enkele artikels uit drie aanlyn Afrikaanse woordeboeke — die HAT Aanlyn, Aanlyn WAT en WAV — ten opsigte van kleurgebruik bespreek, met enkele opmerkinge wat ook oor die strukture of uitleg gemaak word. Daar word primêr gefokus op hoe kleur reeds in die artikels gebruik word, maar waar gepas, word voorstelle vanuit die vorige afdeling betrek om dit duideliker te maak.

Die bespreking word nie gestruktureer ten opsigte van die genoemde woordeboeke nie, maar eerder volgens kategorieë: afkortings, uitdrukkings, en geëtiketteerde lemmas. Dit is so gekies sodat die kleurgebruik by verskillende datatipes oorweeg kan word. Aandag word nie aan die datatipes of inhoud geskenk nie, maar eerder hoe die datatipes aan die woordeboekgebruiker vertoon word.

3.1 Afkortings

Indien 'n afkorting soos QR in die WAV nageslaan word (Figuur 3), is die algemene indruk van die gebruik van kleur dat meestal duidelike kleure gebruik word. Verskillende kleure word aangewend om verskillende datatipes te merk. Dit sal vaardige woordeboekgebruikers toelaat om vinnig op die verlangde data te kan afkom indien hulle die woordeboek se kleurkode leer ken.

QR afk. [Engelse verkorting van **quick response**]

› **QR-kode** | QR-kodes | soort tweedimensionele strepieskode wat bestaan uit swart blokkies van verskillende groottes op 'n wit agtergrond, soos afgedruk op papier of gegenereer op 'n skerm, van waar 'n skandeerder of slimfoon dit vinnig kan lees en dekodeer: › *By enige deur of hek tot enige gebou en by 'n hele paar ad hoc-beheerpunte wat oornag opspring rondom woonbuurte, moes almal weer 'n QR-kode (soms twee) skandeer met hul fone en hul temperatuur laat neem met 'n meetgewertjie* (Etienne Fourie, *Rapport*, 6 Februarie 2022). › *Skandeer die QR-kode vir 'n aanlyn toer van die eiendom* (Netwerk24, 4 Februarie 2022). › *Al hoe meer mense het oorgeskakel na digitale betaalmodes – hetsy tik-en-betaal, e-beursies of deur QR-kodes te skandeer* (Liesl Peyper, 19 Desember 2021).

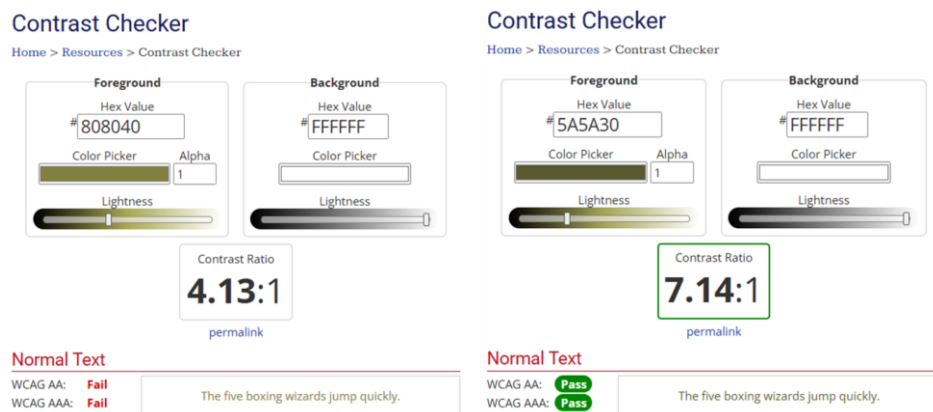
Figuur 3: QR soos in die WAV

Dit is opmerklik dat dieselfde donkerblou (en boonop vetgedrukte teks as pluspunt) gebruik word om die lemma as gidselement te merk, asook die volvorm van die leksikale item wat deur die lemma verteenwoordig word. Dit is nuttig en gerieflik, want woordeboekgebruikers wat slegs die volvorm wil naslaan, sal dit vinnig vind deur die oog na die bypassende blou in dieselfde reël te laat gly. Die lemma en volvorm word ook duidelik van die woordsoortaanduiding en aanduiding van die Engelse oorsprong geskei, wat elk in hulle eie kleure gemerk is. In die betekenisparafraseveld word 'n nuwe kleur nogmaals gebruik, wat die betekenisparafrase as nog 'n mikrostrukturele element duidelik van die omliggende datatipes differensieer. Ná die betekenisparafrase word drie sitate aangebied, met die sitaat en gepaardgaande bronverwysing wat in twee afsonderlike kleure gemerk word. Dit is baie nuttig, want die kleurkombinasie van die sitaat en bronverwysing is duidelik onderskeibaar en 'n gebruiker kan vinnig sien waar die sitate vandaan kom en hoe oud die sitaat is. Dit kan byvoorbeeld help om die gebruikskonteks te bepaal, of die geloofwaardigheid en betroubaarheid van die sitate te probeer beoordeel. Met inagneming van die sitaat se ouderdom kan die gebruiker weer iets oor die ontwikkeling of eietydse gebruiklikheid van die leksikale item aflei.

Deur telkens die verskillende datatipes in verskillende kleure te merk, word kleur as nietipografiese struktuurmerker ingespan. Dit beteken dat kleur as formele element in die artikel gebruik word, en 'n sterker identifiseringswaarde moontlik maak deur die soeksone duideliker te merk. Dit versterk die punt van vroeër dat kleur tot 'n meer gebruikersvriendelike mikro-argitektuur kan bydra deur sekere elemente visueel van mekaar te onderskei en te blok. Soos genoem, kan kundige woordeboekgebruikers met 'n goeie woordeboekkultuur data só vinniger opspoor deur na spesifieke kleure te soek, met die voorwaarde dat 'n verduideliking van die betekenis van spesifieke kleure iewers vermeld moet word. Dit geld weer eens hier, want as kleur as struktuurmerker gebruik word, is dit nodig om dit byvoorbeeld in die gebruikersgids toe te lig wanneer die ander struktuurmerkers verklaar word.

Betreffende die sigbaarheid van die kleure is die meeste kleure donker, wat voordelig is met inagneming van die wit agtergrond. Daarbenewens is dit nuttig dat verskillende kleure, soos blou, grys, purper en 'n olyfgroen, gebruik word, asook kleure wat redelik maklik van mekaar onderskei kan word. By die gebruik van die ligter kleure kan daar egter dalk ietwat aangepas word. As die grys waarin die woordsoortaanduiding gemerk is met WebAIM (2025) teen 'n wit agtergrond getoets word, lewer dit 'n 6,79:1-kontras, wat bevredigend is met inagneming van die vroeër genoemde riglyne. Die olyfgroenkleur lewer weer 'n 4,13:1-kontras, wat nie aan die 4,5:1-mikpunt vir gewone teks voldoen nie. Dit beteken nie dat die kleur totaal onduidelik is nie, maar dit is eerder 'n aanduiding dat 'n effense aanpassing daarvan dit vir meer gebruikers duideliker sal maak. Sou die huidige olyfgroen byvoorbeeld effens donkerder gemaak word, soos met die skuifknoppie op WebAIM (2025), sal dit kan slaag sonder om die totale voorkoms drasties te moet verander (Figuur 4). Soortgelyk kan kleure op

die voorgrond, oftewel die verskillende tekstuele elemente se kleure, getoets word ten einde die kontraswaarde daartussen te bepaal en aan te pas, indien nodig.



Figuur 4: Die oorspronklike olyfgroen (links) kan maklik effens donkerder gemaak word om duideliker te wees (regs)

Wanneer die afkorting *URL* in die WAV nageslaan word (Figuur 5), is dit duidelik dat die kleurgebruik konsekwent toegepas word.

URL *afk., s.nw.* | URL'e, URL's | [Engelse verkorting van **Uniform Resource Locator**] (*internet*) adres wat 'n webblaaiër 'n spesifieke webblad of lêer op die internet help vind: › *As jy aanlyn koop, wees op die uitkyk vir verdagte URL's* (Huisgenoot, 20 November 2020). › *Die webleser gebruik die adres, vertaal die domeinnaam na die bediener se IP-adres en die res van die URL dui die pad na die spesifieke lêer op die gevindde bediener aan.*

Figuur 5: *URL* in die WAV

Die afkorting en volvorm is weer in dieselfde skakering blou, en ook dieselfde skakering as in die artikel van *QR*. Soortgelyk is die woordsoort, aanduiding van die afkorting se oorsprong, betekenisparafrase, en sitate en gepaardgaande bronverwysings konsekwent gemerk. Dit is bevredigend, want soos genoem, sal dit die woordeboekgebruiker met 'n goeie kennis van die WAV help om makliker die verlangde inligting te onttrek. Wat egter effens onduidelik is, is waarom die etiket en aanduiding van die afkorting se Engelse oorsprong in dieselfde purper gemerk is wanneer alle ander datatipes afsonderlike kleure kry.

Wanneer dieselfde lemma in die Aanlyn WAT nageslaan word (Figuur 6),

is die voorkoms van die kleur met die eerste oogopslag ook redelik geslaag. Soos in die WAV, word die lemma, woordsoort, meervoudsvorm, volvorm en sitate byna elk in afsonderlike kleure gemerk.

URL afk., **URL**'s. uniform resource locator (*Eng.*), uniforme hulpbronspoorder: *Die adres van die webbladsy ... is waar die URL ingetik word* (Burg., 1 Apr. 2001, E). *Die URL sal begin met https://* (R. Venter, vert.: Regmaakgids, 2003, 77). *Noudat jy URL's 'n bietjie beter verstaan, kan jy die probleem vinnig oplos deur self die protokol-spesifiseerder in te tik http://ftp.is.co.za/* (Beeld, 14 Des. 1999, E).

Figuur 6: Die Aanlyn WAT se artikel vir *URL*

Hier word die volvorm egter as betekenisparafrase beskou en in swart gegee, wat teenoor die blou, vetgedrukte lemma en meervoudsvorm staan. Relatief tot die WAV sal die benadering effens meer moeite van die woordeboekgebruiker vereis, want die gebruiker kan nie in dieselfde reël net aanhou soek na die blou waarin die lemma gegee is om die volvorm te kry nie. Die gebruiker moet eerder die besonderhede lees om dit te bekom. Nietemin is die woordsoort wat in pienkpers gemerk is duidelik onderskeibaar van die lemma en meervoudsvorm. In die betekenisparafrase is 'n ander voordeel dat die presiseringkruisverwysings wat deur hiperskakels gevolg kan word, gemerk is in die blou kenmerkend van hiperskakels. In dié wyse voldoen die Aanlyn WAT aan Müller-Spitzer et al. (2012: 441) se reeds genoemde punt dat woordeboekgebruikers wil hê aanlyn woordeboeke moet 'n bekendheid met ander toepassings en webformateringsbeginsels toon. Woordeboekgebruikers behoort die hiperskakels danksy die tipiese blou onmiddellik as sodanig te kan eien, wat gerieflik is en toeganklikheid verhoog.

Iets wat nie in die WAV gedoen is nie, maar hier 'n aanwins is, is die groen blok om die leksikale item wat nageslaan is. Dit laat woordeboekgebruikers toe om vinnig deur die artikel te kan gaan en die gesoekte item te kan raaksien, ongeag waar dit in die artikel staan. In die geval van 'n naamwoord soos *URL* is dit nuttig as 'n gebruiker iets oor die sintaktiese optrede wil kan aflei deur byvoorbeeld te kyk watter lidwoord(e) voor die item kan staan, of in ander items se geval dalk watter setsels daarmee optree, of 'n adjektief attributief of predikatief optree, ensovoorts. Daar is hier egter twee beperkinge. Die woordeboekgebruiker het eerstens nie die opsie om die blok te verwyder nie. Dit is egter wenslik om byvoorbeeld 'n knoppie te kan druk wat die blok afhaal of die opsie geheel afsit sodat 'n skoner weergawe van die artikel telkens vertoon word. Die tweede beperking is dat die groen blok slegs om die gesoekte vorm van die lemma verskyn. In die geval van byvoorbeeld 'n werkwoord sal slegs die teenwoordige tyd gemerk word as die woordeboekgebruiker dit so gesoek het; enige ander vorm van dieselfde item (soos die verledetydsvorm of die leksikale item as deelwoord) bly ongemerk. As die gebruiker spesifiek net daardie vorm soek, is dit nuttig. Dit sou wel nut-

tiger wees om moontlik 'n tweede soekmoontlikheid in te sluit wat alle vorme van die item merk om ook daardie vorme wat nie ingetik is nie, in te sluit.

Wat die WAT-redaksie verder kan oorweeg, is om soos in die WAV die bronverwysings in ander kleure as die sitate te merk. Die WAT is 'n omvattende woordeboek, en omvattendheid behoort ook op mikrostrukturele vlak en binne elke mikrostrukturele element te geld (Gouws 2017: 442). Daarom kan dit nuttig wees om bronverwysings in ander kleure te merk omdat dit as visuele skeiding tussen sitate kan dien en gebruikers sal toelaat om vinnig die ouderdom en bron van elke sitaat vas te stel. Hoewel die WAT 'n sinchroniese woordeboek is, beskik dit as langtermynprojek ook in 'n mate oor diachroniese kenmerke. Dit is tipies van omvattende woordeboeke dat daar 'n mate van 'n chronologiese aanduiding van vorm- en betekenisontwikkeling is (Gouws en Prinsloo 2005: 49). Deur dan bronverwysings duideliker van die sitate te onderskei, kan woordeboekgebruikers met 'n oorsig die diachroniese gegewens raaksien, wat behulpsaam kan wees.

3.2 Uitdrukkings

Die artikels wat hierbo gewys is, is telkens kort. In sodanige artikels is dit waarskynlik vir woordeboekgebruikers makliker om die verlangde data te vind, ongeag die kleurgebruik, omdat daar gewoon minder data is wat hulle moet deurwerk. In langer artikels wat veral algemeen is in 'n omvattende woordeboek soos die Aanlyn WAT kan dit egter ingewikkelder wees, en 'n gebruikersvriendelike mikroargitektuur (wat sinvolle kleurgebruik insluit) is dan juis noodsaaklik. Dit is omdat langer artikels uiteraard meer data bevat en 'n hoë kwantiteit data wat swak bestuur word, kan daartoe lei dat woordeboekgebruikers data-oorlading ervaar (Gouws en Tarp 2017: 392). Dit vind plaas wanneer iemand se eiesoortige kapasiteit om inligting te verwerk en in te neem, oorskry word (Speier, Valacich en Vessey, in Gouws en Tarp 2017: 395). Die gevolge is dat gebruikers ans of verwarring ervaar, en hulle onttrek dan nie die benodigde inligting nie omdat dit waarvan hulle probeer sin maak ongeorganiseerd, onbeheersd en inkoherent voorkom (Wurman en Bradford, in Gouws en Tarp 2017: 395). Die artikels kom dan opgeblaas voor en raak weens 'n oorvloed data onprakties (Gouws en Tarp 2017: 392). Dit verhoed direk dat 'n woordeboek aan sy doel as praktiese gebruiksinstrument kan voldoen.

Indien *soveel* byvoorbeeld in die Aanlyn WAT nageslaan word (Figuur 7), behoort dit uit die uitdrukkingsrubriek duidelik te wees waarom die kwessie van data-oorlading belangrik is om te oorweeg in teksdigte, gekondenseerde artikels. Hier is dieselfde lof oor duidelike kleurgebruik wat vroeër genoem is nie sonder meer van toepassing nie. Die uitdrukkinghofie (asook die twee subhofies "gespesialiseerd" en "idiomaties") is duidelik in swart gemerk en kan maklik gedifferensieer word van die kernkomponent as soekwoord, maar van daar af sal die data onder die subhofies nie vir alle woordeboekgebruikers ewe duidelik onderskeibaar wees nie.

UITDR.
A GESPESTALISEERD
moontlik: soveel as moontlik Ook, geselstaal, suffel as moontlik. Die maksimum getal of hoeveelheid, of die maksimum waartoe jy in staat is of wat uitvoerbaar is: 'n Kind oriënteer hom ... in die wêreld deur vertrou te raak met **soveel** as moontlik van die voorwerpe ... wat hy ... teëkom (H.J. Harmse in Harmse – Kirstein: Pedag. 2, 1980, 233). Eerkeër, toe hy my in sy tuin betrap waar ek in my vuil néppie sit en tuinslakke kou, het hy my aangeraal om **soveel** as moontlik slakke te eet, want ... die goed is 'n pes" (K. Kombuis in Insig, Des. 2001, 61). Die Hooftbestuur het ... aan sy verteenwoordigers in die Provinsiale Adviserende Komitee Insake Onderwysdienste (PADKO) opdrag gegee om te beding om **soveel** as moontlik onderwysposte te behou (Unie, Sept. 1991, 54).
soveel: soveel as (n.d. Eng. as much as)
1 Ook, geselstaal, suffel as. Die maksimum of tot 'n maksimum van: **Soveel** as wat 'n man op 'n leerwa kon laai, is vir een pond (R2) verkoop (J.D. Burger in F.A. Venter: Water, s.j., 430). Dit (die elektroniese mikroskoop) kan voorwerpe net 'n paar maal of **soveel** as 1,600,000 maal vergroot (Burg., 30 Jun. 1958, 5 --- opskr.).
2 Ook, geselstaal, suffel as. Eens of enigiens: Daar is nie **soveel** as 'n groen uielof vir slaai nie (E. Kotze: Wind, 2007, 73). Hulle kon hom nie 'n tree versit nie --- hulle kon hom nie **soveel** as roer nie (C.J. Langenhoven: Werke II, 1956, 25).
3 Ook, geselstaal, suffel as. Net (NET 1) of bloot (BLOOT 5): As sy Lordskap **soveel** as 'n ooglid sien beeft solank as hy hierdie skepsel galg toe dons, sal hy julle laat slaan met die kate (C.J. Langenhoven: Weg [2], 1950, 103). Die Liewenheer het my gesien met die hardste baard denkbaar, ... want as ek **soveel** as een enkele dag oorslaan, lê die skeermesem se snykant eers om en klim dan kou-kou bo-op die stappels op sonder om een af te kry (P.H. Nortje: Spook, 1995, 7).
4 Ook, geselstaal, suffel as. In dieselfde mate as: Erna is deel van myself. **Soveel** as wat ek haar handewerk is, **soveel** is sy myne (A.H. de Vries: Kruispad, 1966, 9). Intussen moes die reis voortgaan, en moes sy voorgee om iets wat alle pit verloor het, nog net **soveel** as voorheen te geniet (S. Troskie: Rapsodie, 1962, 186).
5 Ook, geselstaal, suffel as. So baie as: "Alto, oom Gert --- **soveel** as ek van 'n advokaat dink --- hoop ek tog van harte dat ek hom nooit in my lewe nog nodig kry nie" (E. Elff-v.d. Walt: Fontein, 1956, 77). **Soveel** as wat sy van babas hou, kan sy nie eers begin om haar voor te stel hoe dit moet wees om twee pasgeborenes gelyk te versorg nie (E. Steyn, vert.: Skeuring, 2011, 1).
soveel: soveel so (n.d. Eng. so much (so) that) Ook, geselstaal, suffel so. In so 'n mate: Van die kerkvaste Afrikaners in Rhodesië is dié geslag wat in die land gebore is --- **soveel** so dat hulle dikwels nie hul kerk verlaat wanneer hulle met Engelsprekendes in die huwelik tree nie (Huisg., 21 Jan. 1966, 14). Fiat ... se aandeel in die Europese notemark het sedert 1984 standhoudend gegroei. **Soveel** so dat hy Volkswagen verlate jaar as die markleier in daardie mark oortref het (Tegniek, 21 Jul. 1989, bylne, 12).
Vgl. SOSEER 1.
soveel: soveel te ... Ook, geselstaal, suffel te ... In 'n hoë mate: Almal (wat wil gaan boer) moet in die geleentheid wees om eers so ver te kom in die skool as moontlik. As almal eers kan matrikuleer, **soveel** te beter (M.-L. Theron: Reënboop, 1940, 114). Hy wil vir eens en vir altyd ontslae raak van die ander man se telg in sy huis. En as dié telg nog 'n plaas en 'n flukse skoonseun by Hagelkraal se buite kan voeg, **soveel** te beter (S.D. Fourie: Lanternswaier, 2011, 149). Die oprigting van 'n kerk was **soveel** te meer noodsaaklik vir diegene wat op die afgeleë grens gewoon het (C.G. Botha: Ons S.A., s.j., 32).
soveel: soveel vir (n.d. Eng. so much for) Ook, geselstaal, suffel vir. Tot daarnatoe (DAARNATOE) met: 'n Groot dankie aan mnr. J.P. vir sy daadwerklke bydrae om die Sionskerk te restoreer. Dit sou hom seker heelwat minder gekos het as hierdie gebou ... nie deur vandale bygekom is nie --- **soveel** vir die bewaringsgedagte wat sommige Paarloute het (Burg., 23 Mrt. 1983, 22). Toe die verkiesing verby is, skop hulle die AWB in die tande en later selfs vir die KP. **Soveel** vir die samewerking aan die regterkant (K. Smith: AWB, 1996, 87).
B IDIOMATIESE
soveel: soveel en soveel (soms suffel en suffel) Met 'n eienskap, hoeveelheid of graad wat nie gespesifiseer word nie: Dis 'n vloedjaar --- amper weeklks word berig van mense wat dakloos gelaat is, oor **soveel** en **soveel** miljoene rand se skade (T. Holland: Nog 'n casanova, 2016, 1).
soveel: soveel (soms suffel) te se (n.d. Eng. so much as to say) Asof om te se: As ons dit (die parlement) die praatuis noem, is dit gewoonlik met 'n spotlaggie, **soveel** te se: die lede verkies hul tyd nie te praat in plaas van hul werk te doen en wette te maak (Herman – Goldblatt: Janus, 1962, 54).

Figuur 7: Die Aanlyn WAT se hantering van uitdrukkings onder *soveel*

Soos hierbo te sien, word drie kleure gebruik om onderskeidelik die volgende te merk: 'n mosgroen kleur vir die uitdrukking en betekenisparafrase, 'n asuurblou vir die wisselvorm, en 'n blougroen vir die sitate. Relatief tot mekaar lyk die kontrasverhoudings soos volg indien die kleurcodes op WebAIM (2025) ingevoer word: 1,49:1 vir die mosgroen teenoor asuurblou; 1,36:1 wanneer die mosgroen teenoor die blougroen gemeet word; en 1,09:1 vir die asuurblou relatief tot die blougroen. Hierdie groenblou palet kan esteties wees deurdat dit pas by die WAT se kenmerkende blou palet, waarna vroeër ook verwys is, maar dit misluk redelik in die 4,5:1-mikpunt vir gewone teks. Dié mikpunt is wel net 'n riglyn, maar die lae kontras tussen die gebruikte kleure dui daarop dat minstens sommige gebruikers, veral kleurblinde of siggestremde gebruikers, dit uitdaggend kan vind om gemaklik en met sekerheid tussen hierdie kleure te onderskei. Dit lei weer tot 'n minder effektiewe en praktiese gebruik van kleur wat veroorsaak dat dit 'n woordeboekgebruiker langer kan neem om deur die data te werk omdat hulle nie met dieselfde gemak as voorheen net direk die tersaaklike deel kan identifiseer deur 'n spesifieke kleur te soek nie.

Dit kan uiteindelik in artikels so lank soos dié van *soveel* (en veral in langer artikels omdat die Aanlyn WAT ook baie daarvan bevat) moontlik tot data-oormoed lading lei. Dit is omdat baie data in die artikel gehuisves moet word, maar nie noodwendig goed bestuur word in terme van hoe dit uitgelê is en van omliggende data geskei word nie. Die vroeër punt oor 'n dinamiese veelvlakkige artikelstruktuur geld weer hier, want in lang artikels soos in die Aanlyn WAT kan 'n soektonnel en die gebruik van beperkte artikels help om data-oormoed lading te vermy. Daarbenewens kan dit woordeboekgebruikers self laat besluit in watter mate hulle verrykte en uitgebreide data wil bekom. Soos genoem, kan kleur so ingespan word om woordeboekgebruikers te wys waar hulle die behoefte op

meer data moet aandui om sodoende deur die soektonnel te beweeg. Dit kan ook gebruik word om data binne die soektonnel met ander dele van die woordeboek in verband te bring.

Indien die wisselvorm *suffel* nageslaan word (Figuur 8), is die situasie ietwat beter. Hier is die risiko van data-oorlading effens minder en dit behoort makliker te wees om tussen datatipes te onderskei.

suffel (suf-fel) (geselstaal)

I telv.

1 Sien SOVEEL I 1: *Suffel* reën soos nou het ek in '61 laas gesien (Huisg., 19 Des. 1985, 111). Dit is *suffel* (beduie met duim en voorvinger) swakker as laasweek (Carl Boplaas: Knikkertjie, 1998, 39). Vir my was dit 'n allerverskriklike bedrag. Waar sou ek *suffel* geld vandaan kry? (V. Schalkwyk – Hiemstra: Duineboerdery, 2013, 249).

2 Sien SOVEEL I 2: Ek het jou nie *suffel* kere kraal toe sien loop nie (A.A.J. v. Niekerk: Bittergousblom, 1974, 8). Almal hier in Frik du Preezstraat werk mos alles uit volgens 1994. Dis altyd *suffel* jaar voor 1994 of *suffel* jaar daarna (J. Goosen: Pawpaw, 2002, 12). Die Hollands wat hulle daai se dae gepraat het, was al van *suffel* skroeiwarm somers gelouter, dat dit toe al lekker winddroog was (J. le Roux: Engel, 2014, 186).

II bw.

1 Sien SOVEEL II 1: Na 'n ruk voel ek *suffel* beter lat ek opstaan en met my twak afsit stal toe (Carl Boplaas: Knikkertjie, 1998, 23). O'Neill het hulle ... aangeryde brood bekendgestel as 'n voorbeeld van hoe geloof soos die onsigbare suurdeeg *suffel* meer van die koringkorrel maak (J. le Roux: Engel, 2014, E).

2 Sien SOVEEL II 2: Hy het in die distrik geboer. *Suffel* kan ek onthou. "Hy's tog seker nie verlief op haar nie, dus hoef die gedagte jou nie te ontstel nie." "My dier, ek kan nie *suffel* omgee nie" (J. Meintjes: Mallemeule, 1964, 60).

UITDR.

A GESPESTALISEERD

moontlik: suffel as moontlik (geselstaal) Sien SOVEEL AS MOONTLIK (SOVEEL II, uitdr.): 'n Mens moet *suffel* as moontlik spaar vir jou oudag.

suffel: suffel as (geselstaal) (n.d. Eng. as much as)

1 Sien SOVEEL AS 1 (SOVEEL II, uitdr.): Wildesls-poeier. *Suffel* as wat op 'n halfkroon kan gaan, met water (D. Richard: Ysterhout, 1974, 5).

2 Sien SOVEEL AS 2 (SOVEEL II, uitdr.): Daar mag nie *suffel* as 'n stoffie of 'n strepie op wees nie (J. Spies: Profeet, 1984, 15).

3 Sien SOVEEL AS 3 (SOVEEL II, uitdr.): As jy *suffel* as jou oog knip, stuur ek jou uit die klaskamer uit.

4 Sien SOVEEL AS 4 (SOVEEL II, uitdr.): Jy skinder nog net *suffel* as altyd.

5 Sien SOVEEL AS 5 (SOVEEL II, uitdr.): *Suffel* as wat ek hom jannet kry, kan ek hom nie nou help nie.

suffel: suffel so (n.d. Eng. so much (so) that) (geselstaal) Sien SOVEEL SO (SOVEEL, uitdr.): My kop het so gepyn --- *suffel* so dat ek skoon dubbelvisie gehad het.

suffel: suffel te ... (geselstaal) Sien SOVEEL TE ... (SOVEEL II, uitdr.): Ek is alleen; *suffel* te meer noudat dit grensdeltyd is.

suffel: suffel vir (n.d. Eng. so much for) (geselstaal) Sien SOVEEL VIR (SOVEEL II, uitdr.): My buurman leen nie sy grassnyer uit nie. *Suffel* vir buurmanskap!

B IDIOMATIES

suffel: suffel en suffel (meer dikw. soveel en soveel) (geselstaal) Met 'n eienskap, hoeveelheid of graad wat nie gespesifiseer word nie: Ag oom, ek sê ook altyd ... die rokkie kos *suffel* en *suffel* maar dis nie waar nie (Mikro: Koperkanon, s.j., 78). Daai galjoentjie is net *suffel* en *suffel* werd (A.H. de Vries: Dubbeldoor, 1963, 79). Onopvallend op die tweede laaste bladsy 'n berig oor die aandklokkeiling ingevolge artikel *suffel* en *suffel* (J. Goosen: Kat, 1986, 58).

sê: suffel (meer dikw. soveel) te sê (geselstaal) (n.d. Eng. so much as to say) Asof om te sê: *Suffel* te sê: Hel, ek begeer dit nou diep in my hart ... (Carl Boplaas: Knikkertjie, 1998, 199). My ma en ant Mietjie kyk hom kwaad aan, *suffel* te sê: "Sê net iets, Koos Brits!" (J. Goosen: Ons, 1990, 142).

Figuur 8: Die Aanlyn WAT se hantering van *suffel*

Hier word die etiket "geselstaal" telkens in purper aangedui, wat dit goed onderskei van die omliggende teks en wat ook heelwat sigbaarder is. Om seker te maak dat dit sluit, word die etiket in *soveel* se artikel ingesluit, maar daar word dit in die betekenisparafrase as parentese gedoen. In terme van kleur word dit weer in mosgroen gedoen, wat dit minder sigbaar maak. Laasgenoemde maak wel sin met oorweging van die Aanlyn WAT se benadering, maar die gebruik van 'n ander, maklik onderskeibare kleur om 'n etiket in die minder gebruiklike wisselvorm "suffel" se artikel aan te dui, is baie nuttig.

Die verhouding tussen die artikel vir *soveel* en *suffel* word verder in die uitdrukkingsrubriek gesluit deurdat daar na "soveel" gepresiseer word, wat gepas is wanneer 'n woordeboekgebruiker na die meer gebruiklike wisselvorm verwys moet word. Wat dit egter vir sommige gebruikers ingewikkeld kan maak, is dat hierdie presisering in 'n redelik soortgelyke blou gedoen word as die hiperskakelverwysings bo die uitdrukkingsrubriek. Gebruikers wat sukkel om tussen sekere kleure te onderskei, kan dan dalk verward wees wanneer hulle op die teks, byvoorbeeld "SOVEEL AS 1 (SOVEEL II, uitdr.)", probeer klik en niks gebeur nie. Die kenmerkende blou waarin hiperskakels gemerk word, is dus goed, maar wanneer daar in 'n soortgelyke kleur gepresiseer word sonder 'n hiperskakel, kan dit minder effektief wees. Die soortgelyke kleure verg op die een plek 'n interaksie van die gebruiker en laat op 'n ander plek nie so iets toe nie.

As bogenoemde kleurkombinasies aangepas moet word om vir meer gebrui-

kers duidelik te kan wees, kan dit soos by die WAV se olyfgroen van vroeër ietwat aangepas word deur die kleure se helderheid (hoe lig of donker dit is) te verander, of 'n bietjie met die saturasie te speel. Andersins, en dit is hier waarskynlik die makliker roete, kan daar herbesin word oor die Aanlyn WAT se kleurgebruik en met 'n kleurblinde palet in gedagte beplan word. Dit beteken dat die WAT se bloue grotendeels behou kan word en goed saam met oranje, rooi of bruin gebruik kan word, maar nie so goed saam met pers, groen of dowwe grys nie (vgl. Soegoto et al. 2019: 4-5). Dit is boonop reeds duidelik in hoe goed die kleur van die etiket werk, asook hoe die kombinasie van die valer groene en bloue minder geslaagd is.

Wanneer die uitdrukkingsveld in die HAT Aanlyn oorweeg word, byvoorbeeld in die artikel van *vlieg* (Figuur 9), is dit dadelik opmerklik dat 'n eenvoudiger, skoner ontwerp gevolg is.

vlieg² ww. [het gevlieg]
1 deur middel van vlerke in die lug opstyg en voortbeweeg: voëls/insekte/vlermuise **vlieg**.
2 (lugvaartkunde) met 'n vliegtuig reis, daardeur vervoer word of as loods dit self stuur: *Ons het direk van Kaapstad na Amsterdam gevlieg*. • Die slagoffers van die busongeluk is per helikopter hospitaal toe gevlieg. • Noodvoorrade is na die dorpie gevlieg wat deur die oorstromings afgesonder is. • Ek wil leer **vlieg**.
3 (figuurlik) vinnig beweeg/verbygaan/verdwyn: Die skelm **vlieg** om die hoek. • Die skrywer se eerste boek het van die boekvinkelrakke gevlieg. • Die ruit het aan skerwe gespat toe die krieketbal daardeur **vlieg**. • As 'n mens hard aan die werk is, **vlieg** die ure behoorlik. ◊ Die dae van ons jare ... gou gaan dit verby en ons **vlieg** daarheen (Ps. 90:10).
| die tyd **vlieg** die tyd gaan gou verby.

Uitdrukkings
mekaar in die hare **vlieg**
rusie maak; stry.
hoër **vlieg** as wat jou vlerke lank is
meer onderneem as wat jy kan behartig.
te hoog **vlieg**
bo jou vermoë lewe.
in die lug **vlieg**
ontplof.
► **vlieg**: vliegduur, vliegeskader, vlieggebied, vlieghoogte, vliegklub, vliegles, vliegroete, vliegsnelheid, vliegsport, vliegtog, vliegtyd, vlieguitrusting, vliegvrees

Figuur 9: 'n Gedeelte van *vlieg* se artikel in die HAT Aanlyn

Hierdie eenvoudiger, skoner ontwerp is uiteraard in 'n mate omdat die HAT Aanlyn as handwoordeboek nie so omvattend hoef en behoort te wees soos die Aanlyn WAT nie. Sonder die insluiting van sitate by uitdrukkings is dit ook makliker om 'n gebruikersvriendelike mikro-argitektuur te bewerkstellig. Nietemin is daar goeie beginsels toegepas wat die artikel toeganklik maak. Die groen blok om die gesoekte woord is reeds bespreek soos dit in die Aanlyn WAT voorkom, maar dit is weer hier noemenswaardig en nuttig. Dit het egter soos by die Aanlyn WAT dieselfde beperkinge. Wat die HAT Aanlyn se eiesoortigheid betref, is dat die purper kleur waarin die uitdrukkings gegee word heel onderskeibaar is van die swart waarin die uitdrukkinghofie en betekenisparafrase van elke uitdrukking aangebied is.

Iets wat hierdie benadering nog beter maak, is dat die purper redelik verskil van die ander kleure in die artikel, met uitsondering van die rooierige kleur van die lemma en nommers wat die betekenisonderskeidinge merk. Laasgenoemde is egter nie 'n probleem nie, want die purper staan nie direk langs of naby genoeg aan die rooi vir die kleure om verwarrend te wees nie. Betreffende die ander kleure is dit opmerklik dat swart, blou, oranje en olyfgroen gebruik word, wat almal kontrasgewys redelik van purper verskil. Gevolglik sou daar steeds maklik tussen die data onderskei kon word selfs al was meer data, soos kollokasies en/of sitate, ingesluit. Daar kan aangeneem word dat kollokasies en/of sitate in blou gegee sou word ter aansluiting by hoe dit in die res van die artikel gedoen is. Indien wel, sou dit steeds duidelik onderskeibaar wees van die uitdrukking in purper en die gepaardgaande betekenisparafrase in swart. Die probleem in die Aanlyn WAT se hantering van uitdrukkings is dus nie uitsluitlik vanweë die kwantiteit data nie, maar in 'n mate weens 'n wanbestuur, of eerder swakker aanbieding, van die data.

3.3 Geëtiketteerde lemmas

In al drie woordeboeke kan die gebruik van kleur 'n waardevolle rol by die oordrag van pragmatiese inligting speel deurdat dit etikette duideliker sigbaar kan maak. Etikette speel 'n belangrike rol omdat dit enige afwyking van die standaardtaal aandui, byvoorbeeld ten opsigte van formaliteit, gebruiksfrekwensie of gebiedsverankering, of andersins nuttige en nodige gegewens oor die gebruiksfrekwensie oordra. Hierdie gegewens kan by sekere leksikale items veral belangrik wees, en 'n woordeboekgebruiker wat nie duidelik genoeg van die etiket bewus gemaak word nie, kan byvoorbeeld in die verleentheid gestel word wanneer hulle 'n item as neutraal interpreteer wanneer dit eintlik minder gepas is vir 'n bepaalde situasie of selfs kwetsend is. Kleur kan juis hier gebruik word om dié beperkinge op die gebruik van 'n leksikale item verteenwoordig deur die lemma vir woordeboekgebruikers duideliker aan te dui.

Dit is byvoorbeeld duidelik hoe etikette in die WAV gemerk word, soos in die artikel vir *moffie* (Figuur 10). In dié artikel word *moffie* duidelik as informeel geëtiketteer in die bepaalde betekenisonderskeidinge wat in die WAV opgeneem word. Die gebruik van purper laat die etiket beter uitstaan relatief tot sowel die blou van die lemma en meervoudsvorm wat reg voor dit verskyn as die swart waarin die betekenisparafrase kort daarna gegee word. Daarbenewens laat die kleur die etiket duidelik by die ander kontekstuele leiding aansluit, want so ook is dieselfde purper gebruik om "deur lede van die LGBTQIA-gemeenskap gebruik om na hulle self te verwys" en "dikwels kwetsend, neerhalend deur cisgender mense gebruik in verwysings na lede van die LGBTQIA-gemeenskap" te merk. Hierdie besonderhede is belangrik om uit te lig en dui duidelik die gebruiksbepalinge aan om te wys dat die leksikale item "onder" die standaardtaal is en nie as neutraal geag moet word nie.

moffie s.nw. | moffies | (informeel)
1 (deur lede van die LGBTQIA-gemeenskap gebruik om na hulle self te verwys) queer persoon: >
"Moffie" is oorspronklik 'n Afrikaanse Gaylewoord vir queer mense. Nie 'n vloekwoord nie (Du Toit
Albertze, Klyntji, 22 November 2018).
2 (dikwels kwetsend, neerhalend deur cisgender mense gebruik in verwysings na lede van die
LGBTQIA-gemeenskap) queer persoon: > Ewe skielik onthou ek hoe 'n trans suster by Stellenbosch
se treinstasie uitgejou is. Ek het net staan en toekyk terwyl die woord "moffie" skryn in my ore (Du
Toit Albertze, Klyntji, 22 November 2018).
○ ENGELS: **queer**

Figuur 10: Die artikel vir *moffie* soos in die WAV

Soortgelyk word etikette goed met kleur in die Aanlyn WAT uitgelig, soos genoem. Dit is ook hier goed in gevalle waar 'n lemma 'n potensiaal neerhalende leksikale item verteenwoordig, soos in die Aanlyn WAT se hantering van *raaiselkind* (Figuur 11).

raaiselkind s.nw., raaiselkinders. (ongewoon; neerhalend) Kind (KIND 1 b en c) wat outisties is: Knaende vroe tref een ná die ander die glashok waarbinne I.D. soos 'n verskrikte seedier vasgevang sit ná die dood van haar raaiselkind (Burg., 1 Apr. 2003, 8). Raaiselkinders ... lyk na gewone kinders maar is van kleins af net heeltemal anders (Hoofst., 21 Mei 1982, 9).

Figuur 11: Die Aanlyn WAT se benadering tot *raaiselkind*

In só 'n geval kan woordeboekgebruikers van 'n omvattende woordeboek egter wonder wat in die Aanlyn WAT as neerhalend beskou word, of verdere toeligting oor etikette en hulle gebruik verlang. Hulle kan moontlik ook wonder waarom sekere items, veral rassismes, aansienlik korter en minder omvattend as ander items bewerk word. Vir hierdie gebruikers kan dit nuttig wees as kleur dan gebruik word om die verhouding tussen die etiket en die bepaalde buitetekst, in die Aanlyn WAT se geval die hanteringsbeleid vir beledigende en sensitiewe leksikale items, duidelik te maak. Dit kan op 'n paar maniere gedoen word, maar die etiket kan byvoorbeeld in een kleur gemerk word wat die gebruiker via hiperskakelroete na die bypassende buitetekst neem wat met dieselfde kleur omraam word om die relevante teksdeel uit te lig. Nog 'n moontlikheid is 'n opskietspyskaart wat aan die kant van die skerm vertoon word, met oortjies na verskillende buitetekste in die WAT. Die oortjie na die hanteringsbeleid as tersaaklike buitetekst kan dan byvoorbeeld in dieselfde kleur gemerk word. Andersins kan 'n knoppie of ander simbool in een kleur binne die artikel gemerk word, met so 'n tipe oortjie in dieselfde kleur. Sodoende kan woordeboekgebruikers bewus gemaak word van waar hulle addisionele toeligting kan bekom om die woordeboek se spesifieke benadering beter te verstaan.

4. Gevolgtrekking

Uit die bespreking van die enkele punte is dit duidelik dat kleur 'n belangrike rol in aanlyn woordeboeke se artikel kan en moet speel. Dit is omdat dit toegang kan verbeter en 'n wesenlike bydrae tot 'n gebruikersvriendelike mikro-argitektuur kan lewer deur as nietipografiese struktuurmerker ingespan te word. Kleur alleenlik kan nie hierdie werk verrig nie, en moet ook nie die enigste middel wees waardeur bepaalde gegewens oorgedra word nie. Desondanks is dit iets wat 'n rol kan speel in hoe suksesvol 'n woordeboek as praktiese gebruiksinstrument kan dien. Daarom is dit nodig om doelbewus te dink oor die gebruik van kleur en die moontlikhede wat dit in die artikelstruktuur inhou, asook oor hoe die gebruik duidelik iewers toegelig moet word sodat woordeboekgebruikers dit kan verstaan. Heelwat van die moontlikhede in die teoretiese oorsig is egter tans nog toekomsmusik in die Afrikaanse leksikografie, soos die veel-vlakkige artikelstruktuur. Dit is nietemin opmerklik dat kleur in 'n woordeboek soos die WAV waaraan werk meer onlangs begin is reeds doeltreffend begin gebruik word, selfs al is daar nog by oorweging van die individuele kleure 'n bietjie ruimte vir verbetering. Die ontwikkeling is wel belowend en hou opwindende moontlikhede in.

Indien die Webinhoudtoeganklikheidsriglyne oorweeg word, is dit duidelik dat kleur 'n rol in sensitiwiteit kan speel en redaksies kan help om hulle begrip van sensitiwiteit te verbreed. Wat hiermee bedoel word, is dat sensitiwiteit nie net ten opsigte van pragmatiese inligting oor bepaalde leksikale items hoef te geld nie. Dit kan ook geïnterpreteer word ten opsigte van bepaalde gebruikersbehoefte en -vermoëns. Dit is veral moontlik en nodig in aanlyn woordeboeke waarin kleiner aanpassings geen party sal benadeel nie, maar wel vir dié gebruikers wat dit nodig het 'n verskil kan maak sodat 'n woordeboek vir daardie gebruikers as praktiese gebruiksinstrument kan dien.

Eindnote

1. Soos aangekondig in 'n nuusbrief van 9 April 2025, het die Aanlyn WAT 'n nuwe baadjie gekry, wat onder meer 'n nuwe kleurskema insluit. Die skermkote wat in hierdie artikel ingesluit is, is egter reeds in Februarie 2025 geneem. Die nuusbrief is hier beskikbaar:
<https://www.wat.co.za/wat-wys-sy-kleure/>

Bronnelys

Woordeboeke

- Aanlyn WAT:** *Woordeboek van die Afrikaanse Taal*.
<https://www.woordeboek-co-za/> [20 Februarie 2025]
- HAT Aanlyn:** *Handwoordeboek van die Afrikaanse Taal*.
<https://mml-hat-co-za/> [24 Februarie 2025]

WAV: *Woordeboek van Afrikaans Vandag*.

<https://afrikaans-vandag.co.za/> [20 Februarie 2025]

Ander bronne

- Almind, R.** 2005. Designing Internet Dictionaries. *Hermes, Journal of Linguistics* 18(34): 37-54.
- Berisso, K.** 2018. Addressing Color Blind Awareness in the Classroom. *Journal of Business and Management Sciences* 6(3): 93-99.
- Beyer, H.L. en P.A. Louw.** 2022. Aspekte van vernuwing in die Afrikaanse leksikografie: Standaard- en pedagogiese woordeboeke as barometer. *Lexikos* 32(3): 25-48.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Dobres, J., Chahine, N. en B. Reimer.** 2017. Effects of Ambient Illumination, Contrast Polarity, and Letter Size on Text Legibility under Glance-Like Reading. *Applied Ergonomics* 60: 68-73.
- Fuertes-Olivera, P.A. en S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlyn: De Gruyter.
- Gouws, R.H.** 2001. The Use of an Improved Access Structure in Dictionaries. *Lexikos* 11: 101-111.
- Gouws, R.H.** 2005. Oor die verhouding tussen woordeboekstrukture, woordeboekinhoud en leksikografiese funksies. *Lexikos* 15: 52-69.
- Gouws, R.H.** 2017. Leksikografie. Carstens, W.A.M. en N. Bosman (Reds.). 2017. *Kontemporêre Afrikaanse Taalkunde*: 429-463. 2de uitgawe. Pretoria: Van Schaik.
- Gouws, R.H.** 2018a. Accessibility, Access Structures and Access Procedures. Jesenšek, V. en M. Enčeva (Reds.). 2018. *Wörterbuchstrukturen zwischen Theorie und Praxis*: 35-56. Berlyn: De Gruyter.
- Gouws, R.H.** 2018b. Internet Lexicography in the 21st Century. Engelberg, S., Kämper, H. en P. Storjohann (Reds.). 2018. *Wortschatz: Theorie, Empirie, Dokumentation*: 215-236. Berlyn: De Gruyter.
- Gouws, R.H.** 2018c. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.
- Gouws, R.H.** 2023. Nuwe raamstruktuurtypes in aanlyn woordeboeke. *Lexikos* 33(1): 382-403.
- Gouws, R.H. en T.J.D. Bothma.** 2024. Aspekte van inligtingsonttrekkingstrukture in aanlyn woordeboeke. *Lexikos* 34(1): 355-370.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. en S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Gouws, R.H. en S. Tarp.** 2024. Despite Current Challenges: Lexicography Has a Bright Future. *Lexicographica* 40(1): 187-212.
- Index of Color Contrast Samples.** s.a.
<https://web.archive.org/web/20160814160227/http://trace.wisc.edu/contrast-ratio-examples> [28 Februarie 2025]
- Jamil, A. en G. Denes.** 2024. Investigating Color-Blind User-Interface Accessibility via Simulated Interfaces. *Computers* 13(2): 53-68.
- Juicy Studio.** 2025. Luminosity Colour Contrast Ratio Analyser.
<https://juicystudio.com/services/luminositycontrastratio.php> [28 Februarie 2025]

- Kilgarriff, A. en I. Kosem.** 2012. Corpus Tools for Lexicographers. Granger, S. en M. Paquot (Reds.). 2012. *Electronic Lexicography*: 31-55. Oxford: Oxford University Press.
- Klosa, A. en R.H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31(1): 142-172.
- Mihindou, G.-R.** 2006. *A Theoretical Model for a Yipunu–English–French Explanatory Dictionary of Medical Terms*. Ongepubliseerde D.Litt.-verhandeling. Stellenbosch: Universiteit Stellenbosch.
- Müller-Spitzer, C. en A. Koplenig.** 2014. Online Dictionaries: Expectations and Demands. Müller-Spitzer, C. (Red.). 2014. *Using Online Dictionaries*: 143-188. Berlyn: De Gruyter.
- Müller-Spitzer, C., Koplenig, A. en A. Töpel.** 2012. Online Dictionary Use: Key Findings from an Empirical Research Project. Granger, S. en M. Paquot (Reds.). 2012. *Electronic Lexicography*. 425-457. Oxford: Oxford University Press.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Shieh, K.-K. en Y.-K. Lai.** 2008. Effects of Ambient Illumination, Luminance Contrast, and Stimulus Type on Subjective Preference of VDT Target and Background Color Combinations. *Perceptual and Motor Skills* 107(2): 336-352.
- Soegoto, E.S., M.A. Fadhlurrahman en Y.A. Hermawan.** 2019. Video Streaming Application Design for Color Blind Users. *Journal of Physics: Conference Series* 1402(6): 1-8.
- Swasty, W. en A.R. Adriyanto.** 2017. Does Color Matter on Web User Interface Design? *CommIT (Communication & Information Technology) Journal* 11(1): 17-24.
- Tarp, S.** 2012. Do We Need a (New) Theory of Lexicography? *Lexikos* 22: 321-332.
- TPGi.** 2025. Colour Contrast Analyser (CCA).
<https://www.tpgi.com/color-contrast-checker/> [28 Februarie 2025]
- WCAG (Web Content Accessibility Guidelines) 2.2.** 2024. *W3C Recommendation 12 December 2024*.
<https://www.w3.org/TR/WCAG21/> [27 Februarie 2025]
- WebAIM.** 2025. Contrast Checker.
<https://webaim.org/resources/contrastchecker/> [28 Februarie 2025]
- Wiegand, H.E.** 1983. Ansätze zu einer allgemeinen Theorie der Lexikographie. Schildt, J. en D. Viehweger (Reds.). 1983. *Die Lexikographie von heute und das Wörterbuch von morgen: Analysen — Probleme — Vorschläge*: 92-127. Berlyn: Akademie der Wissenschaften der DDR, Zentralinstitut für Sprachwissenschaft.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts: An Overview. *South African Journal of Linguistics* 14(4): 134-149.
- Wiegand, H.E.** 1998. *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexicographie*. Berlyn: De Gruyter.
- Wiegand, H.E. en S. Beer.** 2013. Access Structures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard en H.E. Wiegand (Reds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 110-149. Berlyn/New York: De Gruyter.

Gender Bias in Computer-generated Thesauri: The Case of the Serbian Section of *Kontekst.io*, a Thesaurus of Synonyms and Semantically Related Terms

Dragana Čarapić, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro* (draganac@ucg.ac.me)

(<https://orcid.org/0000-0003-3375-0169>)

and

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

(<https://orcid.org/0000-0001-5497-1361>)

Abstract: This paper studies gender bias in the computer-generated thesaurus *Kontekst.io*, which is a search portal of synonyms and semantically related terms in Serbian, Croatian and Slovenian. Its Serbian section, which is the focus here, is based on a natural language processing (NLP) technique called word embeddings and a large internet corpus of Serbian. Gender bias is uncovered in four selected entries of this thesaurus: *žena* (woman), *muškarac* (man), *d(j)evojka* (young woman) and *momak* (young man). The analysis is first conducted semantically and the terms found are grouped into various semantic fields. After that, in the vein of the earlier studies of gender bias in traditional dictionaries and critical discourse analysis, an analysis of gender bias in the selected entries is provided. The results show that gender bias is ubiquitous and that it extends deeper than the earlier studies of gender bias in word embeddings have shown. We then give recommendations for improving this lexicographic product based on the results.

Keywords: GENDER BIAS, COMPUTER-GENERATED THESAURUS, WORD EMBEDDINGS, *Kontekst.io*, SERBIAN, LEXICOGRAPHY

Opsomming: Geslagsvooroordeel in rekenaargegenereerde tesourusse: Die geval van die Serwiese afdeling van *Kontekst.io*, 'n tesourus van sinonieme en semanties verwante terme. In hierdie artikel word geslagsvooroordeel in die rekenaargegenereerde tesourus *Kontekst.io*, 'n soekportaal van sinonieme en semanties verwante terme in Serwies, Kroaties en Sloweens, bestudeer. Die Serwiese afdeling, waarop daar hier gefokus word, is gebaseer op 'n natuurliketaalprosesseringstegniek (NTP-tegniek) genaamd woordinbedding en 'n groot internetkorpus van Serwies. Geslagsvooroordeel word in vier uitgesoekte inskrywings in hierdie tesourus blootgelê: *žena* (vrou), *muškarac* (man), *d(j)evojka* (jong vrou) en *momak* (jong man). Die ontleding word eers semanties uitgevoer en die terme wat gevind word, word in verskillende semantiese velde ingedeel. Daarna, in dieselfde trant as vroeëre studies van geslagsvooroordeel in

tradisionele woordeboeke en kritiese diskoersanalise, word 'n ontleding van geslagsvooroordeel in die uitgesoekte inskrywings verskaf. Die resultate toon dat geslagsvooroordeel alomteenwoordig is en dat dit verder strek as wat vroeëre studies van geslagsvooroordeel in woordinbedding aange-ton het. Aanbevelings wat op die resultate gebaseer is, word dan gemaak om hierdie leksikogra-fiese produk te verbeter.

Sleutelwoorde: GESLAGSVOOROORDEEL, REKENAARGEGENEREEERDE TESOURUS, WOORDINBEDDING, *Kontekst.io*, SERWIES, LEKSIKOGRAFIE

1. Introduction

The recent rise of machine learning and artificial intelligence (AI) has transformed many fields. Among these, Natural Language Processing (NLP) has gained promi-nence by enabling computers to process and generate human language with ever increasing sophistication. Central to NLP are word embeddings — vector representations of words which capture their semantic meaning based on the context in which they are used in large corpora (Lee 2020). Word embeddings, however, reflect, perpetuate and even amplify societal biases from the source data, i.e. corpora, including gender bias (Gonen and Goldberg 2019). For in-stance, one type of word embeddings (*word2vec*), trained on the *Google News* data-set, in answer to the following: "man is to computer programmer as woman is to x", provides that "x is a homemaker" (Bolukbasi et al. 2016). Gender bias was demonstrated as omnipresent and consistent across different types of word embeddings and the proposed methods of mitigating it have so far produced only a limited effect — as Gonen and Goldberg put it, it is the effect of putting "a lipstick on a pig" (2019).

One area where word embeddings are starting to find its application is the creation of thesauri, especially for lesser-resourced languages, with poorly devel-oped lexical databases (Arppe et al. 2023). Lexicographic products reflect but also shape our understanding of language and, by extension, the society (cf. Grenon-Nyenhuis 2000). Traditional dictionaries have been found to reflect and repro-duce gender stereotypes (Bergenholtz and Gouws 2006; Nübling 2009). The same is true, and much more so, of word embeddings and the products based on them, given that they reflect gender bias existing in the source data. This fact should be taken into account when making lexicographic products based on word embeddings, bearing in mind that lexicographers have a moral responsi-bility to avoid reinforcing harmful gender stereotypes (Müller-Spitzer 2023).

This paper investigates gender bias in a computer-generated thesaurus intended for public use, aiming to uncover and point to some of such biases in it. The thesaurus in question is *Kontekst.io*, which contains semantically related words and synonyms in Serbian, Croatian and Slovenian. Our study will focus on the examples from Serbian. It should be noted that in this study we consider thesaurus as a type of a dictionary, following Gouws (2017: 134).

Gender bias in word embeddings has been studied relatively extensively so far (Zhao et al. 2019; Basta et al. 2019; Yang and Feng 2020; Basta et al. 2021;

Caliskan et al. 2022; etc.). However, as far as we know, the present study will be the first to offer a linguistic analysis of gender bias in a thesaurus based on word embeddings intended for general public use. In addition, the study of gender bias in word embeddings has mostly focused on the phenomenon of gender bias in occupation terms, while the present study will go beyond that — we will cover all synonyms and semantically related terms provided under several selected entries in Serbian (*žena* (woman), *muškarac* (man), *d(j)evojka* (young woman), and *momak* (young man)).

The current study builds on the body of research examining gender bias in traditional dictionaries and in word embeddings. The analysis will be performed within the vein of critical discourse analysis, given that we can understand dictionary as a text (cf. Fuertes-Olivera and Tarp 2022). In critical discourse analysis, ideology and power are key terms, and lexicography and its products are always subject to them, as they "are never value-free, apolitical or asocial" (Chen 2019: 1). Critical discourse analysts depart from the premise that language is not a mere mirror of social phenomena, but also their constitutive factor (Vuković-Stamatović 2022: 429) and that the analysis must simultaneously include the analysis of text, discourse practice, i.e. the origin, distribution and/or use of the text, and, finally, its social context (Fairclough 1992, 1995). Given that in this study we are dealing with a computer-generated thesaurus, we cannot criticise the lexicographer's intervention but we can criticise the lack thereof, as well as the data which are the source of bias. The approach to lexicography through critical discourse analysis belongs to critical lexicography in general (Kachru 1995; Chen 2019).

The background of the paper consists of two parts: Section 2 will cover the phenomenon of gender bias in traditional dictionaries, while Section 3 will present word embeddings and discuss them through the lens of gender bias. After that, in Section 4, we will present the thesaurus *Kontekst.io* and our method, following which the analysis is provided.

2. Gender bias in traditional dictionaries

In patriarchal societies, men are seen as the primary sources of power and moral authority. They also tend to hold prominent positions in decision-making, social privilege and property ownership. Certain cultures are far more patriarchal than others, even though most societies are patriarchal to some degree and assign gendered responsibilities (James 2010). This has traditionally had an impact, in particular, on women's status in these societies and has also been reflected in their languages. Dictionaries do not only reflect the sociolinguistic reality of certain speech communities — they also tend to propagate it, especially bearing in mind that they are frequently seen as authoritative sources which can shape how we understand words. As such, a dictionary can be "an agent of social impact" (Gouws 2022: 40).

Various research studies have highlighted the presence of gender bias in dictionaries. Some recent papers on the issue include: Norri 2019; Iversen 2021;

Solonets 2021; Pettini 2021; Müller-Spitzer and Rüdiger 2022; Vacalopoulou 2022; Fuertes-Olivera and Tarp 2022; Müller-Spitzer 2023, etc., indicating that gender bias in dictionaries is really not a thing of the past (cf. Fuertes-Olivera and Tarp 2022) and that modern dictionaries still feature it, despite the increasing efforts dictionary makers have been investing in reducing it.

Gender bias in dictionaries manifests in different ways. Some of these include asymmetrical definitions, stereotypical usage examples, and a choice of collocations which reflects stereotypes.

To illustrate asymmetrical definitions for different genders, we will use an example from the *Dictionary of the Serbian Language* (Matica Srpska 2007), where the entries *man* and *woman* are defined as follows:

- *man* – a person, human being of a male gender, a person who is of a gender opposite to *woman* (2007: 746),
- *woman* – 1. a human being who has the ability to give birth, of a gender opposite to *man*. 2. a person who is married, a wife. 3. a. an adult person of a female gender. b. a female person working in a house, a maid. 4. fig. pej. a weak person, a coward (when talking about men) • *easy* ~ a woman of low morale in her relation to men. *take a woman* – to get married (2007: 369).

The asymmetry in the two definitions above can be seen even physically — the entry related to *man* is much shorter to that for *woman*. *Man* is, first of all, a *person*, who does not need any more specific defining, whereas *woman* is firstly described through her role of giving birth (meaning 1). In addition, *woman* is stereotypically portrayed in meaning 2, where she is defined through her marital status; in meaning 3b, where she is defined through household chores and working for someone else; then in meaning 4, where a *woman* means a *weak* man; and in the phrase *easy woman*, with negative connotations, again in relation to men. The only neutral meaning for *woman* is, in fact, provided in 3a. The described gender difference is a result of the Serbian patriarchal society, which is reflected in the given dictionary entries. However, as said before, dictionaries do not just describe the sociolinguistic reality, but they propagate it (and, perhaps, co-construct it?), and one of the dilemmas before dictionary makers is how to intervene and to what extent.

As Müller-Spitzer (2023: 80) notes, lexicographers have a responsibility, given that dictionaries may contribute to exclusion and perpetuating gender stereotypes. In other words, "the representation of gender in dictionaries is a matter of both language use and lexicographic-moral responsibility" (Müller-Spitzer 2023: 83). Due to this, many dictionary makers have opted to invest effort into reducing gender bias as much as possible in their dictionaries. As a result, a stark asymmetry in the definitions of genders is more rarely seen in modern dictionaries, however, gender bias remains at more subtle levels. Müller-Spitzer and Rüdiger (2022: 130) analyse the choice of examples in the entries on *man*, *woman*, *boy* and *girl* in the *Cambridge Dictionary*, and quote the following examples from these entries:

"He plays baseball, drinks a lot of beer and generally acts like one of the boys."

"Steve can solve anything — the man's a genius."

"Who was that beautiful girl I saw you with last night?"

"Both girls compete for their father's attention."

The examples contain gender stereotypes — men and boys are associated with playing sports, drinking alcohol, and intelligence; women and girls are described in terms of appearance and as fighting for men's attention.

A similar situation may be noted with collocations. For instance, the *Online OXFORD Collocation Dictionary of English* (2019) (<https://m.freecollocation.com>) provides the following adjectival collocates for the entry *girl*:

- baby; little, small, young; adolescent, teenage; bubbly, happy, lively; lovely, nice; attractive, beautiful, good-looking, gorgeous, handsome, pretty, stunning; single, unmarried.

In contrast, consider the following adjectival collocates given by the same dictionary for the entry *boy*:

- big; little, small; young; elder, eldest, older; baby; adolescent, teenage; good; naughty; bright, clever.

As can be seen, some of the collocates for boys and girls are similar, especially those referring to age and size. However, considerable differences may be noted in other semantic fields — for instance, on the one hand, *girls* have six collocates for describing pleasant physical looks, four for describing pleasant disposition, and two for describing marital/partner status, as qualities considered important for females in anglophone societies, while boys are not described at all in terms of these three aspects. On the other hand, two collocates for boys refer to their intelligence, as a quality more associated with men in gender stereotypes, while such collocates are not found for girls. In addition, *boys* are described as *naughty*, also a quality more associated with men, as more obedience is expected from girls. The choice of the collocates certainly comes from the corpus that informs the *Online OXFORD Collocation Dictionary of English* and the same is true of the thesauri created from word embeddings.

In the literature, we have not come across examples treating gender bias reflected in a choice of synonyms (and semantically related terms) for certain entries, but, as our present analysis will show, gender bias can be uncovered through this lens as well.

3. Word embeddings and gender bias

As suggested earlier, a word embedding is a numeric representation of a word in the form of a vector. Such representations are used in Natural Language Pro-

cessing (NLP) to allow for automatic text analysis and processing to complete tasks such as sentiment analysis (analysis of texts to determine if their emotional tone is positive, neutral, or negative), translation, web search, parsing through different kinds of texts, etc. They have also been used to inform and make some dictionaries, thesauri in particular (Morinaga and Yamaguchi 2018; Chaimae et al. 2020; Liang et al. 2023; Arppe et al. 2023), although this use of word embeddings has not been as frequent as the uses mentioned above.

When using the technique called word embeddings, each word is assigned a number format, i.e. a real-valued vector. Words used in similar contexts, i.e. similarly, are closer in vector space — for example, the words "tomato" and "potato" would be close together in the vector space because they have certain similarities, e.g. they are both edible domesticated plants, while "tomato" and "house" would be further apart because they are used in largely different contexts. Thus, the techniques involving word embeddings measure and classify semantic similarities between words by examining their distributional properties within large language data.

For easier understanding of what word embeddings actually are, below we provide a graphic visualisation for some words in a simple, three-dimensional vector space:

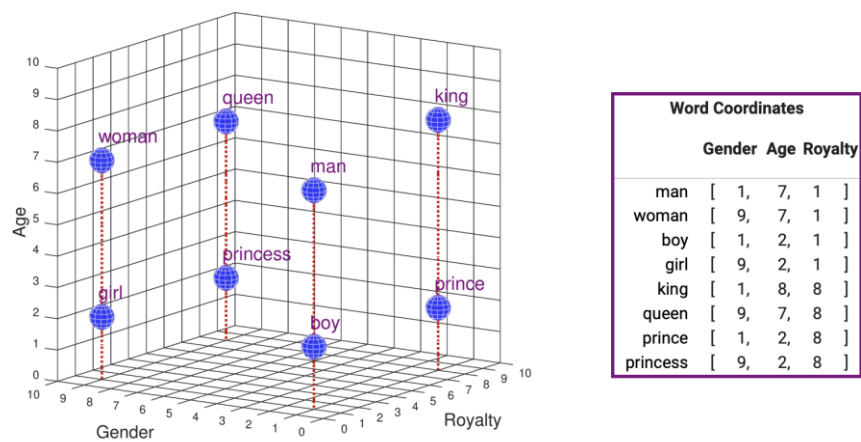


Figure 1: An illustration of word embeddings in a 3D vector space (Touretzky 2024)

In Figure 1, the words *boy*, *girl*, *man*, *woman*, *prince*, *princess*, *king* and *queen* are represented along three semantic dimensions: gender (smaller values indicate male gender, while larger indicate female), age (smaller values indicate younger age), and royalty (greater values for royal status). The values assigned are not perfectly symmetrical — for instance, for age, *king* is assigned an 8 while a queen is assigned a 7. The reason for this is that often in the corpora there is more talk

of *old kings* than *old queens*. Each of the coordinates in this 3D space can be read using three numbers, one for each of the three dimensions involved — the list of these three numbers or values for all the words depicted in the figure is given on the right. These series of numbers for each word represent their vectors. Of course, as we said, this 3D representation is very simplified and words, in practice, are represented by a much higher number of dimensions — of the order of tens and even hundreds of dimensions; in addition, the numbers are more precise and are rarely whole but rather decimal numbers. Based on the numbers assigned to each word, words can be mathematically compared in terms of their semantic similarity or semantic distance, and analogies such as "a man is to a woman as a king is to a queen" can be derived. Based on this methodology, it is possible to derive a list of synonyms and semantically related terms from a corpus, and as we have seen, these can be used to form lexical databases, inform dictionaries, and make thesauri.

As suggested earlier, gender bias is rather frequent in word embeddings, primarily as a reflection of the source data which naturally contain it. Thus, word embeddings can yield analogies such as "a father is to a doctor as a mother is to a nurse" (Bolukbasi et al. 2016). However, word embeddings do not just reflect the source data and the biases in them; the algorithms which they are based on can even inadvertently amplify them (Bolukbasi et al. 2016; Basta et al. 2019). In response to this problem, different debiasing methods have been proposed — some of these focus on correcting the source data itself, some intervene in the algorithm, while some use human-made dictionaries for creating less biased word embeddings (Zhao et al. 2019; Yang and Feng 2020, etc.). However, the results are mixed and what is certain is that gender bias always remains in the output (Gonen and Goldberg 2019; Lee 2020). Most debiasing approaches focus on the words related to occupations, but gender bias extends much deeper and can be hidden across multiple features (Caliskan et al. 2022). So far, the debiasing methods have been successful at rather "hiding gender bias but not truly removing it", Ronchieri and Biagi (2023: 730) conclude.

4. Data and method

As suggested earlier, the data used in this study derives from the computer-generated thesaurus *Kontekst.io*. In the following text, we present *Kontekst.io* in more details, following which more light is shed on the method.

4.1 *Kontekst.io*

Kontekst.io is a search portal of synonyms and semantically related terms in Serbian, Croatian and Slovenian, based on word embeddings (Plahuta 2024). On its website, it is stated that it can primarily be used as "a dictionary of synonyms".

What the homepage of *Kontekst.io* looks like can be seen in Figure 2 below:



Figure 2: The homepage of *Kontekst.io*

As can be seen, the website's homepage is simple — the user first chooses one out of the three languages offered. The second step is to type in the search term or to choose one of the popular search terms offered below.

The output offers a list of words with similar distributional properties, i.e. semantically related words including synonyms and even antonyms, along with the normalised frequency of the word per 1 million words of the corpus, its score of semantic similarity with the entered term and the examples of its use in the source corpus. Below is part of the output for the entry *lijepa* (beautiful) from the Serbian section (Figure 3); we present only part of it, as the whole output contains some 50 synonyms and semantically related terms and cannot fit here.

Kontekst.io			
slični izrazi i sinonimi u savremenom srpskom, hrvatskom i slovenskom			
SL HR SR			
Upotrebite okvir za pretraživanje			
... ili kliknite jednog od donjih primera:			
kabadahija • životna priča • švaler • hemijska industrija • teroristička grupa • štazi • civilizacijska tekovina • partibrejkers • pravoslavlje • fibrilacija			
Sličnost reči ili fraza u rezultatima zavisi od toga, koliko puta se reč ili fraza pojavlja u sličnom kontekstu kao "lijepa".			
Slični izrazi i sinonimi za		Primeri iz općenitog korpusa	
lijepa		Korpus srWac	
Kliknite za traženje		srWac je korpus srpskog jezika (Charis.ai)	
	UČESTALOST	SLIČNOST	
lepa	36.67	82%	moj mentor i ja, dijelili smo međusobno poštovanje. I iz poštovanja, izraslo je prijateljstvo, i jedna lijepa
prekrasna	1.69	80%	ničeg nejasnog i mračnog. Svoj život on je proveo u mjestima koja su uvijek bila u prirodi i vanredno lijepa
prelijepa	0.76	80%	tri zvezde imagoje ' '. Jednog jutra, Veljko tjera u polje, da pasu, veliku svoja krmaču i tri vrlo lijepa
predivna	4.46	78%	u skladu sa povodom koncert je trebalo da posle 50 minuta bude zaklju? en legendarnom " Bila je tako lijepa
divna	11.47	78%	
krasna	0.67	77%	
zgodna	5.57	77%	

Figure 3: Some synonyms and semantically related terms for *lijepa* (beautiful) in *Kontekst.io* (Serbian)

The Serbian thesaurus is based on the Serbian web corpus *srWaC* (Ljubešić and Klubička 2016), which contains 554,627,647 running words (version 1.1), and was obtained by crawling the *.rs* web domain (the Serbian national web domain). Some of the advantages of using this corpus include its size, the fact that it is quite recent and the fact that it is available for public use. Additionally, there are not many viable corpus alternatives when it comes to corpora for Serbian, which is far less resourced than many other European languages. The disadvantage of the corpus is that all the texts come from the Internet.

As we find in Ulčar et al. (2021), *Kontekst.io* was made using the *word2vec* algorithm (Mikolov et al. 2013) and a 256-dimensional vector model trained for its needs. Using their mathematical approach, Ulčar et al. (2021) have also determined that *Kontekst.io* contains a high degree of gender bias in its entries referring to occupations, but they have not looked beyond that.

According to the data from *Similarweb.com*, a website tracking 100 million websites in 190 countries, *Kontekst.io* was the ninth most popular website in the category of dictionaries and encyclopedias in Serbia in April 2024. Most of the websites preceding it were global websites, such as *Wikipedia.org*. Based on this, we can say that *Kontekst.io* is a rather popular thesaurus in Serbia. Most of its traffic comes from the referrals from the *Google* search when users are looking for synonyms for certain words (again, according to *Similarweb.com*).

4.2 Method

In this study, we analyse the synonyms and semantically related terms provided under four selected entries: *žena* (woman), *muškarac* (man), *d(j)evojka* (young woman), and *momak* (young man), in the Serbian component of *Kontekst.io*. The number of the entries studied was limited by the space provided for this paper. The entries chosen are obviously gendered words, which is why it was expected that their analysis could most clearly point to the presence and types of gender bias in *Kontekst.io*.

To uncover and analyse gender bias in the selected entries, we use a method which encompasses comparative analysis, grouping the terms in separate semantic fields and applying the connotative analysis of the specific terms. The method embodies different levels of paradigmatic relations, starting from the hyponymy — superordinate term, and continuing with the synonymous expressions allocated to the respective semantic fields. For better visualisation, the results thus processed are presented in tables, comparatively for *woman* and *man*, as well as for *young woman* and *young man*. This level of analysis is purely semantic.

Further, the semantic method is complemented with critical discourse analysis (Fairclough 1992, 1995), given that the ideology underlying the selection of the terms presented as synonyms and semantically related terms, is uncovered and discussed. The selection is computer-produced but it reflects the bias existing in the source data which feed the thesaurus. More specifically, the analysis of the entries is in the vein of the analyses of gender bias in traditional dictionaries

presented in Section 2: our own examples and those of Müller-Spitzer (2023). Thus, the inequalities in the choice and the frequency of the different synonyms and semantically related terms (provided for the said male and female entries) are noted and discussed.

5. Results and analysis

Given that two pairs of entries are analysed in this paper, the analysis will be organised in two parts, each dedicated to one of these.

5.1 The entries for *žena* (woman) and *muškarac* (man) in Serbian *Kontekst.io*

The terms provided as synonyms and semantically related terms for the entries *woman* and *man* have first been grouped into separate semantic fields, as explained in the Method subsection. The results for Serbian *Kontekst.io* are presented in Table 1 below.

Table 1: Synonyms and semantically related terms for the entries *žena* (woman) and *muškarac* (man) in Serbian in *Kontekst.io*¹

<i>Kontekst.io</i> (SERBIAN)	
ŽENA (WOMAN)	MUŠKARAC (MAN)
SUPERORDINATE TERM: Osoba (person)	SUPERORDINATE TERM: Čovek, čovjek (human, man), stvor (creature)
AGE-RELATED: Beba (baby), devojčica, djevojčica (little girl), djevojka, devojka, devojaka (young woman), cura (lass), curica (lassy DIM. ²), mlada dama (young lady), mlada devojka (young girl), starica (old woman), bakica (granny)	AGE-RELATED: Mališan (tot), dečak, dječak, dečko (boy), klinac (kid), mladić, momak (young man), tinejdžer (teenager), adolescent (adolescent), pedesetogodišnjak (fifty-year-old man), matorac (old man)
OCCUPATION-RELATED: Domaćica (housewife), dadilja (nanny), sluškinja (maid)	OCCUPATION-RELATED: Policajac (police officer), vojnik (soldier)
FAMILY-RELATED: Trudnica (pregnant woman), majka (mother), udovica (widow), porodica (family)	FAMILY-RELATED: Muž (husband), roditelj (parent)
ANIMAL-RELATED: Kuja (bitch PEJ. ³)	ANIMAL-RELATED: Mužjak (male), pas (dog), bik (bull), majmun (monkey), konj (horse)
EVIL-RELATED: Veštica (witch PEJ.)	EVIL-RELATED: Nasilnik (bully), vampire (vampire)

PATIENT/BENEFICIARY ROLES: Pacijentkinja (patient FEM. ⁴), pacijentica (patient FEM.), zatvorenica (prisoner FEM.), mušterija (customer FEM.), klijentkinja (client FEM.)	PATIENT/BENEFICIARY ROLES: Pacijent (patient)
APPEARANCE-RELATED: Plavuša (blonde), crnkinja (black woman), belkinja (white woman), ženica (woman DIM.)	APPEARANCE-RELATED: Belac, bijelac (white man), crnac (black man)
FOREIGNER: Francuskinja (French woman)	FOREIGNER: Stranac (foreigner), Indijanac (Indian), Amerikanac (American)
MONEY-RELATED: Bogatašica (rich woman)	MONEY-RELATED: Bogataš (rich man), beskućnik (homeless man)
SOCIAL ROLE-RELATED: Dama (lady), komšinica (neighbour FEM.), prijateljica, drugarica (friend FEM.)	SOCIAL ROLE-RELATED: Džentlmen (gentleman)
OTHER: Duša (darling), seljančica (village girl)	SEXUALITY-RELATED: Frajer (hot shot), ljubavnik (lover)

The first thing which we can notice in Table 1 is that the superordinate term for *muškarac* (man) is *čov(j)ek* (human), whereas for *žena* (woman) it is *osoba* (person). Further, the following can also be noted (the observations are presented according to the semantic fields):

- **age:** more diminutives and the adjective *young* preceding the nouns designating women, e.g. *djevojka*, *devojka*, *devojaka* (young woman), *mlada dama* (young lady), *mlada devojka* (young girl), *curica* (lassy DIM.);
- **occupation:** gender bias detected under both the entries, e.g. *domaćica* (housewife), *dadilja* (nanny), and *služkinja* (maid) under the entry *woman*, and *policajac* (police officer) and *vojn timer* (soldier), under the entry *man*;
- **family:** there are more terms in this field for *woman* than *man*, e.g. *trudnica* (pregnant woman), *majka* (mother), *udovica* (widow), *porodica* (family), are all used for *woman*, while only *muž* (husband) and *roditelj* (parent) are used for *man*.
- **animal:** only one term with negative connotations is used for *woman*: *kuja* (bitch PEJ.); however, there are five terms in this field for *man*: *mužjak* (male), *pas* (dog), *bik* (bull), *majmun* (monkey), and *konj* (horse), some of which have negative connotations, too;
- **evil:** gender bias detected in the derogatory terms *veštica* (witch PEJ.) for *woman*, and *nasilnik* (bully) and *vampire* (vampire) for *man*;
- more **appearance**-related terms for women: *plavuša* (blonde), *crnkinja* (black woman), *belkinja* (white woman), *ženica* (woman DIM.);

- more **socially-** and **family-**defined **roles** for *woman*: *dama* (lady), *komšinica* (neighbour FEM.), *prijateljica*, *drugarica* (friend FEM.);
- more **money-**related characterisations for men, e.g. two depicting *man* (*bogataš* (rich man), and *beskućnik* (homeless man)) vs. one used for *woman* (*bogatašica* (rich woman));
- **sexuality-**related terms — only used for men, whereby both the terms in this field have a positive meaning: *frajer* (hot shot), *ljubavnik* (lover);
- **other terms** for *woman*: there is one endearing term in the data and it is used for *woman* (*duša* (darling)); the other term in this category is *seljančica* (village girl) and it is condescending.

We proceed with the results for the second pair of entries analysed.

5.2 The entries for *d(j)evojka* (young woman) and *momak* (young man) in Serbian *Kontekst.io*

In the same vein, we conduct the analysis for the entries: *d(j)evojka* (young woman) and *momak* (young man). The results are given in Table 2.

Table 2: Synonyms and semantically related terms for the entries *d(j)evojka* (young woman) and *momak* (young man) in Serbian in *Kontekst.io*

<i>Kontekst.io</i> (SERBIAN)	
D(J)EVOJKA (YOUNG WOMAN)	MOMAK (YOUNG MAN)
SUPERORDINATE TERM: Dama (lady)	SUPERORDINATE TERM: Čova, čovek, čovjek (human, man), čovečuljak (man DIM.), muškarac (man)
AGE-RELATED: Bakica (granny DIM.), cura, curica (girl), devojčica, djevojčica, klinka, mala devojčica (little girl), mlada devojka (young girl), starica (old woman), tinejdžerka (teenage girl)	AGE-RELATED: Čikica (old man DIM.), dečak, dečkić (young boy DIM.), dečko, dječak, deran, klinac (young boy), matorac (old guy), mladić, momčić (young man DIM.)
OCCUPATION-RELATED: Dadilja (nanny), konobarica (waitress), služavka (maid), stažistica (intern FEM.)	OCCUPATION-RELATED: Pandur (cop), policajac (police officer)
FAMILY-RELATED: Majka (mother), sestra (sister)	FAMILY-RELATED: Muž (husband), roditelj (parent)
ANIMAL-RELATED: Kučka (bitch), mačka (cat)	ANIMAL-RELATED: Džukac (mutt), majmun (monkey), pas (dog)

EVIL-RELATED: Ludača (mad woman)	EVIL-RELATED: Gad (bastard), govnar (piece of shit DER. ⁵), kučkin sin, (son of a bitch DER.), kurvin sin (son of a whore DER.), ludak (mad man), seronja (asshole DER., SLANG), tip (guy)
APPEARANCE-RELATED: Lepa devojka (pretty girl), lepotica (beauty), plavuša (blonde)	APPEARANCE-RELATED: Crnac (black man), debeljko (fat man), lepotan (handsome man), plavušan (blonde man)
SOCIAL ROLE-RELATED: Cimerka (roommate FEM.), drugarica (friend FEM.), komšinica (neighbour FEM.), nevesta (bride), poznanica (acquaintance FEM.), prijateljica (friend FEM.), verenica (fiancé FEM.)	SOCIAL ROLE-RELATED: Drugar (friend), dasa (hot shot), klippan (loon DER.), ortak (buddy)
SEXUALITY-RELATED: Fufa, fufica, kurva (hooker), prostitutka (prostitute FEM.)	SEXUALITY-RELATED: Frajer (hot shot), jebač (fucker DER., SLANG)
MONEY-RELATED: Bogatašica (rich woman)	FOREIGNER: Indijanac (Indian), Amerikanac (American)

The following can be noted regarding the results in Table 2:

- **age:** more diminutive expressions are used preceding the noun for *young man*: *čovečuljak* (man DIM.), *čikica* (old man DIM.), *dečkić* (young boy DIM.), *momčić* (young man DIM.);
- **occupation:** gender bias is found under both *young woman* and *young man*, e.g. *dadilja* (nanny), *konobarica* (waitress), *služavka* (maid), *stažistica* (intern FEM.), in the former category, and *pandur* (cop) and *policajac* (police officer) in the latter;
- **family:** the same number of synonyms/semantically related terms are provided for both *young woman* and *young man*: *majka* (mother) and *sestra* (sister), vs. *muž* (husband) and *roditelj* (parent);
- **animal:** on the one hand, two terms are used for *young woman*: *kučka* (bitch), which has negative connotations, and *mačka* (cat), which has sexual connotations; on the other hand, three terms are used with negative connotations for *young man*: *džukac* (mutt), *majmun* (monkey), and *pas* (dog);
- **evil:** in this semantic field, there is one term for *young woman* (*ludača* (mad woman)), whereas there are seven derogatory terms for *young man*: *gad* (bastard DER.), *govnar* (piece of shit DER.), *kučkin sin* (son of a bitch DER.), *kurvin sin* (son of a whore, DER.), *ludak* (mad man), *seronja* (asshole DER., SLANG), and *tip* (guy);

- more **appearance**-related terms are used for *young man*, e.g. *crnac* (black man), *debeljko* (fat man), *lepotan* (handsome man), and *plavušan* (blonde man), than for *young woman*: *lepa devojka* (pretty girl), *lepotica* (beauty), and *plavuša* (blonde);
- there more **socially defined roles** under the entry *young woman*: *cimerka* (roommate FEM.), *drugarica* (friend FEM.), *komšinica* (neighbour FEM.), *nevesta* (bride), *poznanica* (acquaintance FEM.), *prijateljica* (friend FEM.), *verenica* (fiancé), than for *man*: *mladoženja* (bachelor), *razbojnik* (scourer), and *zatvorenik* (prisoner);
- there is one **money**-related characterisation for *young woman*: *bogatašica* (rich woman), and there are no such characterisations for *young man*;
- there are two expressions for **foreigner** under the entry *young man*: *Indijanac* (Indian) and *Amerikanac* (American);
- **sexuality**: four terms from this field are used for *young woman*: *fufa*, *fufica*, *kurva* (hooker) and *prostitutka* (prostitute FEM.), while two such terms are used for *young man*: *frajer* (hot shot) and *jebač* (fucker DER., SLANG).

A discussion of these findings is presented in Section 6.

6. Discussion

In the analysis of the selected terms in *Kontekst.io* in Serbian, several elements of gender bias have been detected.

On the one hand, the superordinate term for both *muškarac* (man) and *momak* (young man) is *čov(j)ek* (human), whereas for *žena* (woman) it is *osoba* (person) and for *d(j)evojka* (young woman) it is *dama* (lady). The male terms are thus *more encompassing*, given that they can stand for the entire human kind, i.e. both men and women, unlike the female terms, which are much more specific and exclusive in the sense that they cannot extend their meaning beyond the female sex. Even more specifically, *person* stands for just one individual. Additionally, *lady* is a term which reflects a socially preferable role for women.

Another finding is that there are more diminutives under the entries for *woman*, as well as that the adjective *young* often precedes the nouns in this category. This has to do with the general notion of woman as a weaker sex in a patriarchal society. As such, woman requires more affection and protection through the use of diminutives.

We have also noticed that there is gender bias under all the analysed entries related to occupations. As said earlier, this phenomenon has been investigated earlier (for instance, Ulčar 2021) and it has been shown that there is substantial gender bias in word embeddings when it comes to the names of occupations.

In *Kontekst.io*, more family-related terms are provided under the entry *woman*

than *man*. This result is reflective of the social position of woman, who is largely defined through the role she has within her family in a patriarchal society. Furthermore, it is noteworthy that there is no mention of *father* as a semantically related term for either *man* or *young man*, whereas *mother* is identified as a semantically related term for both *woman* and *young woman*. The only family-related word that is provided for *man* and *young man* is *parent*, a gender-neutral designation that does not specify paternal identity. Such a disbalance is not seen, however, in the second pair of the entries, given that the additional family roles for females are assumed only after a certain age.

There are more terms in the semantic field *animal* for *man* and *young man*, than under the female entries. Many of these terms used for men have negative connotations, as they are associated with negative character traits: being *aggressive*, *violent* and *fierce*. This is also the result of gender bias existing in society and, consequently, in the source data used for the thesaurus.

Likewise, in the semantic field relating to *evil*, there are many more terms for *man* than *woman*. This result points to a similar notion of *man* as a sex more capable of inflicting violence and other evil acts and it also reveals that it is socially more acceptable to address other men rather than women using such terms. Further, of the two terms supplied in this field for women, *witch* particularly stands out and also reflects gender bias.

More appearance-related terms for the entry *woman* show that woman is more socially defined by her physical appearance than man. The terms relating to *young woman* and *young man*, however, are similar in number, suggesting that appearance becomes much more relevant for women only after a certain age, the same was noted for socially defined roles for women.

More money-related characterisations for men imply the importance of *money* in defining *man*. Also, in our data, foreigners tend to be men more often than women.

There are some more sexual terms for *young woman* than *young man*, but it should be noted that all the terms from this field have negative connotations. All of them are *derogatory* slang expressions and they imply negative character traits.

As we can see, gender bias is ubiquitous in all the entries studied. This is not just gender bias which is related to the female gender, which is usually more obvious — on the contrary, in the terms under the said entries we find reflections of many stereotypical characterisations of men, too. Thus, men are seen as being aggressive and violent, money is important for how they are socially defined, and they are seen to be working in certain professions, such as being a policeman or a soldier.

Some of the gender bias described here is more "innocuous" than the other. Perhaps on the more "innocuous" end of the spectrum are the terms for occupations, which have been the most frequent topic of interest for those seeking to address gender bias in word embeddings through improving their algorithms (Section 2). We believe that it is much more serious when the terms such

as *witch*, *prostitute*, *bitch*, and *mad woman*, for example, are provided as synonyms or semantically-related terms for *woman* or *young woman*. Most of the users of *Kontekst.io* will not be familiar with the discipline of semantics or the meaning of "semantically-related", especially in terms of semantic distribution, while many will understand what synonyms are and will perhaps see the terms provided as such. As suggested earlier, the website of *Kontekst.io* itself states that it can primarily be used as "a dictionary of synonyms", while most of its traffic comes from the *Google* search for synonyms. That being said, it does not help that this search portal does not contain instructions or a help page providing an explanation for common users on what these terms mean and what the results provided by the thesaurus actually show. We recommend that such guidance be provided visibly for the users.

Moreover, we have noticed that for derogatory terms and other terms with negative connotations, which were quite frequent in the results, no such designations were provided in order to point to users that these terms are not in common, everyday use. Providing the necessary flags, as is customary in dictionaries, should be the next step in improving *Kontekst.io*. Perhaps these can be imported from other dictionaries.

Additionally, it goes without saying that all products, including *Kontekst.io*, which are based on word embeddings, must be further improved towards eliminating more gender bias, whereby the process should not solely or primarily focus on the occupation terms. Lexicographic moral responsibility is not absolved if the thesaurus is created by a machine, i.e. an algorithm — people are always behind those algorithms, in the same way that they are behind the corpora informing the machines.

Gender bias can probably never be fully eliminated in any dictionary in general and in computer-generated ones in particular — we have demonstrated here how deep it can go in many different directions and for both the genders here studied. However, when making lexicographic products which are publicly available, one must be aware of the harms which such products may do, as they do not only reflect the existing gender bias but also may perpetuate and co-construct it.

7. Conclusion

This paper investigated gender bias in the computer-generated thesaurus *Kontekst.io*, aiming to uncover examples of this bias in it and give recommendations for improving it as a lexicographic product. *Kontekst.io* is a search portal for synonyms and semantically related terms in Serbian, Croatian and Slovenian, and this paper focused on its Serbian section — specifically, the terms provided under the entries: *žena* (woman), *muškarac* (man), *d(j)evojka* (young woman), and *momak* (young man).

Gender bias in the studied entries was found to be ubiquitous and to run deeper than the earlier studies of this phenomenon in word embeddings have

shown. Earlier studies were mostly conducted by non-linguists who focused on the more obvious examples of gender bias and concentrated more on improving the algorithms than on detecting all types of gender bias in their products.

Based on the results, we gave some recommendations how *Kontekst.io* (and other similar lexicographic products) may be improved. These include providing clear definitions and instructions to users explaining what kind of terms are supplied in the entries, providing flags marking the use of certain words as derogatory, slang etc., and working further on the algorithms, thus eliminating more gender bias, especially beyond the occupation terms.

Ultimately, we believe that gender bias cannot be reduced to a certain satisfactory level without human intervention in lexicographic products such as the one studied here. However, we do understand the necessity of such products in lesser-resourced languages such as Serbian, but we hope that they can be additionally improved and the users more instructed into what they really are and what their limitations are, so as to avoid harmful consequences such as perpetuating and co-constructing gender bias.

Our analysis focused on the entries where gender bias might be more salient. Further investigation should focus on more entries and other language sections of *Kontekst.io* in order to uncover the more subtle forms of gender bias in them and produce further recommendations for improving this and other similar lexicographic products.

Endnotes

1. Note: Mistyped words have been excluded here.
2. Diminutive.
3. Pejorative.
4. Feminine.
5. Derogatory.

References

- Arppe, A., A. Neitsch, D. Dacanay, J. Poulin, D. Hieber and A. Harrigan. 2023. Finding Words that Aren't There: Using Word Embeddings to Improve Dictionary Search for Low-resource Languages. *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*: 144-155. Toronto, Canada: Association for Computational Linguistics.
- Basta, C., M.R. Costa-jussà and N. Casas. 2019. Evaluating the Underlying Gender Bias in Contextualized Word Embeddings. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*: 33-39. Florence, Italy: Association for Computational Linguistics.
- Basta, C., M.R. Costa-jussà and N. Casas. 2021. Extensive Study on the Underlying Gender Bias in Contextualized Word Embeddings. *Neural Computing and Applications* 33(8): 3371-3384.
- Bergenholtz, H. and R. Gouws. 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13-45.

- Bolukbasi, T., K.-W. Chang, J.Y. Zou, V. Saligrama and A.T. Kalai.** 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Lee, D., M. Sugiyama, U. Luxburg and I. Guyon and R. Garnett. 2016. *Advances in Neural Information Processing Systems* 29. (30th Annual Conference on Neural Information Processing Systems 2016 (NIPS 2016), 5-10 December 2016, Barcelona, Spain): 4349-4357. La Jolla, CA, USA: NIPS Foundation.
- Caliskan, A., P.P. Ajay, T. Charlesworth, R. Wolfe and M.R. Banaji.** 2022. Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES 2022)*: 156-170. Oxford: AIES.
- Chaimae, A., M. Rybinski, E.Y. Yacine and J.F. Aldana Montes.** 2020. Comparative Study of Arabic Word Embeddings: Evaluation and Application. *International Journal of Computer Information Systems and Industrial Management Applications* 12: 349-362.
<https://cspub-ijcisim.org/index.php/ijcisim/article/view/469>
- Chen, W.** 2019. Towards a Discourse Approach to Critical Lexicography. *International Journal of Lexicography* 32(3): 362-388.
- Fairclough, N.** 1992. *Discourse and Social Change*. Cambridge: Polity.
- Fairclough, N.** 1995. *Critical Discourse Analysis*. London/New York: Longman.
- Fuertes-Olivera, P.A. and S. Tarp.** 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132.
- Gonen, H. and Y. Goldberg.** 2019. Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings but Do not Remove Them.
arXiv preprint arXiv:1903.03862
- Gouws, R.H.** 2017. Van tematiese na alfabetiese na tematiese ordening in woordeboeke — wisselwerking tussen teorie en praktyk. *Stellenbosch Papers in Linguistics Plus* 53: 133-148.
- Gouws, R.H.** 2022. Dictionaries as Instruments of Exclusion and Inclusion: Some South African Dictionaries as Case in Point. *Lexicographica* 38(1): 39-61.
<https://doi.org/10.1515/lex-2022-0003>
- Grenon-Nyenhuis, C.** 2000. The Dictionary as a Cultural Institution. *Intercultural Communication Studies* 10(1): 159-166.
- Iversen, S.H.** 2021. The (Re)presentation of Knowledge about Gender in Children's Picture Dictionaries. Goga, N., S.H. Iversen and A.-S. Teigland (Eds.). 2021. *Verbal and Visual Strategies in Nonfiction Picturebooks: Theoretical and Analytical Approaches*: 67-79. Oslo: Scandinavian University Press.
- James, K.** 2010. Domestic Violence within Refugee Families: Intersecting Patriarchal Culture and the Refugee Experience. *Australian and New Zealand Journal of Family Therapy* 31(3): 275-284.
- Kachru, B.B.** 1995. Afterword: Directions and Challenges. Kachru, B. B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta*: 417-424. Tübingen: Max Niemeyer.
- Lee, E.** 2020. Gender Bias in Dictionary-derived Word Embeddings. Technical Report (CS230: Deep Learning, Fall 2020, Stanford University, CA).
http://cs230.stanford.edu/projects_fall_2020/reports/55476615.pdf
- Liang, H., Y.M.M. Ng and N.L. Tsang.** 2023. Word Embedding Enrichment for Dictionary Construction: An Example of Incivility in Cantonese. *Computational Communication Research* 5(1): 1-26.
- Ljubešić, N. and F. Klubička.** 2016. *Serbian Web Corpus srWaC 1.1*, Slovenian Language Resource repository CLARIN.SI, ISSN 2820-4042.
<http://hdl.handle.net/11356/1063>.

- Mikolov, T., I. Sutskever, K. Chen, G. Corrado and J. Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. *arXiv preprint arXiv:1310.4546*
- Morinaga, Y. and K. Yamaguchi. 2018. Improvement of Reverse Dictionary by Tuning Word Vectors and Category Inference. Robertas Damaševičius, R. and G. Vasiljeviene (Eds.). 2018. *Information and Software Technologies: 24th International Conference, ICIST 2018, Vilnius, Lithuania, October 4–6, 2018, Proceedings* 24: 533–545. New York: Springer.
- Müller-Spitzer, C. 2023. Gender Stereotypes in Dictionaries: The Challenge of Reconciling Usage-based Lexicography with the Role of Dictionaries as Social Agents. *Lexikos* 33(2): 79–94.
- Müller-Spitzer, C. and J.O. Rüdiger. 2022. The Influence of the Corpus on the Representation of Gender Stereotypes in the Dictionary. A Case Study of Corpus-based Dictionaries of German. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs and P. Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*: 129–141. Mannheim: IDS-Verlag.
- Norri, J. 2019. Gender in Dictionary Definitions: A Comparison of Five Learner's Dictionaries and their Different Editions. *English Studies* 100(7): 866–890.
- Nübling, D. 2009. Zur lexikografischen Inszenierung von Geschlecht. Ein Streifzug durch die Einträge von Frau und Mann in neueren Wörterbüchern. *Zeitschrift für germanistische Linguistik* 37(3): 593–633.
<https://doi.org/10.1515/ZGL.2009.037>
- Online OXFORD Collocation Dictionary of English. 2019.
<https://m.freecollocation.com>
- Pettini, S. 2021. One is a Woman, so That's Encouraging too. The Representation of Social Gender in "Powered by Oxford" Online Lexicography. *Lingue e Linguaggi* 44: 275–295.
- Plahuta, M. 2024. *Kontekst.io*.
<https://virostatiq.com/> [Accessed 20 July 2024]
- Ronchieri, E. and C. Biagi. 2023. Comparing Methods for Mitigating Gender Bias in Word Embedding. Bui, T.X. (Ed.). 2023. *Proceedings of the 56th Hawaii International Conference on System Sciences, January 3–6, 2023*: 722–731. Honolulu: HICSS.
<https://hdl.handle.net/10125/102720>
- Similarweb. 2024. <https://www.similarweb.com/website/blog.context.io/> [Accessed 10 July 2024]
- Solonets, P.V. 2021. *Gender and Dictionary: Russian Perspective*. Unpublished Master's Thesis. Braga: University of Minho.
- Touretzky, D. 2024. Word Embedding Demo (webpage), Carnegie Mellon University blogs.
<https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/tutorial.html>
- Ulčar, M., A. Supej, M. Robnik-Šikonja and S. Pollak. 2021. Slovene and Croatian Word Embeddings in Terms of Gender Occupational Analogies. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 9(1): 26–59.
- Vacalopoulou, A. 2022. Gender Stereotypes in Greek Children's Dictionaries. *Dictionaries* 43(1): 167–192.
- Vujanić, M., D. Gortan-Premk, M. Dešić, R. Dragičević, M. Nikolić, L.J. Nogo, V. Pavković, M. Radović-Tešić, N. Ramić, R. Stijović and E. Fekete. 2007. *Rečnik srpskog jezika* [Dictionary of the Serbian Language]. Novi Sad: Matica Srpska.
- Vuković-Stamatović, M. 2022. "Accessing the EU Is Like Running on a Treadmill in the Gym": How the EU Accession Process is Metaphorically Presented in the Online Media of Serbia, Montenegro, and Bosnia and Herzegovina. *Annales. Series historia et sociologia* 32(3): 427–448.

- Yang, Z. and J. Feng.** 2020. A Causal Inference Method for Reducing Gender Bias in Word Embedding Relations. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(5): 9434-9441.
<https://ojs.aaai.org/index.php/AAAI/article/view/6486>
- Zhao, J., T. Wang, M. Yatskar, R. Cotterell, V. Ordonez and K.-W. Chang.** 2019. Gender Bias in Contextualized Word Embeddings.
arXiv preprint arXiv:1904.03310

'n Omvattender aanbod van semantiese data in aanlyn woordeboeke*

Rufus H. Gouws, *Departement Afrikaans en Nederlands,
Universiteit Stellenbosch, Suid-Afrika*
(rhg@sun.ac.za)
(<https://orcid.org/0000-0002-3423-058X>)

Opsomming: Betekenis is die datatipe waaraan die meeste gebruikers van verklarende woordeboeke 'n behoefte het. Weens ruimtebeperkings kan gedrukte woordeboeke slegs in 'n beperkte mate verslag doen van betekenisaspekte wat vir woordeboekgebruikers ter sake is. In die vasstelling van betekeniskenmerke wat in woordeboekverklarings aangebied word, laat leksikograwe hulle grootliks lei deur insigte vanuit ouer benaderings in die linguistiek. Gevolglik word 'n streng skeiding tussen linguistiese en ensiklopediese data dikwels steeds gehandhaaf. Na 'n oorsig oor aspekte van woordeboektipologie, woordeboekstrukture en semantiese data in gedrukte woordeboeke word daar in hierdie artikel voorstelle gemaak vir die benutting van nuwe woordeboekstrukture om 'n omvattender weergawe van semantiese data in aanlyn verklarende woordeboeke te bewerkstellig. In die verskillende semantiese subkommentare van 'n artikel word daar voorsiening gemaak vir 'n horisontale ordening van verskillende vakke wat elk 'n datatipe huisves. Meer ruimte word geskep vir semantiese data en ook vir data wat buite die bestek van tradisionele woordeboekinskrywings val. 'n Sterker fokus word ook geplaas op betekenisbetrekkinge en maniere waarop hierdie betekenisleiding in die semantiese kommentaar aangebied kan word. Nuwe tipes geïntegreerde mikrostrukture word voorgestel om die data op gepaste maniere toeganklik te maak vir gebruikers, en voorstelle word gemaak om inligting ook aan woordeboek-eksterne bronne te onttrek. Semantiese data moet so goed en so omvattend as moontlik in verklarende woordeboeke aangebied word.

Sleutelwoorde: BETEKENISPARAFRASE, GEÏNTEGREERDE MIKROSTRUKTUUR, PARALLEL-GEÏNTEGREERDE MIKROSTRUKTUUR, SEMANTIESE DATA, SEMANTIESE KOMMENTAAR, SEMANTIESE SUBKOMMENTAAR, SKAGSTRUKTUUR, SOEKTONNEL, SOEKUNIVERSUM, UITGESTELDE GEÏNTEGREERDE MIKROSTRUKTUUR, VAKKE, VEELFASIGE GEÏNTEGREERDE MIKROSTRUKTUUR, VLAKKE

Abstract: A More Comprehensive Presentation of Semantic Data in Online Dictionaries. Meaning is the type of data most frequently needed by dictionary users. Due to space restrictions printed dictionaries can only offer a limited reflection of those aspects of meaning relevant to dictionary users. In determining features of meaning to be used in explanations of meaning, lexicographers are often guided by insights from older approaches in linguistics. Consequently, a strict division is often still maintained between linguistic and encyclopaedic data. Following an overview

* Met waardering opgedra aan Frikkie Lombard.

of some aspects of dictionary typology, dictionary structures and semantic data in printed dictionaries, this paper offers suggestions for the use of new types of dictionary structures to ensure a more comprehensive presentation of semantic data in online monolingual dictionaries. In the different subcomments on semantics in dictionary articles, provision is made for a horizontal ordering of different compartments to accommodate different types of data. More space is allocated to semantic data and also to data that fall beyond the scope of traditional dictionary entries. A stronger focus is placed on semantic relations and ways in which this type of semantic guidance could be accommodated in the comment on semantics. New types of integrated microstructures are suggested to make data accessible for target users, and suggestions are also made for the retrieval of information from dictionary-external sources. Semantic data should be presented as good and as comprehensive as possible in monolingual dictionaries.

Keywords: COMMENT ON SEMANTICS, COMPARTMENTS, DELAYED INTEGRATED MICROSTRUCTURE, INTEGRATED MICROSTRUCTURE, LEVELS, MULTI-PHASED INTEGRATED MICROSTRUCTURE, PARALLEL INTEGRATED MICROSTRUCTURE, PARAPHRASE OF MEANING, SEARCH TUNNEL, SEARCH UNIVERSE, SEMANTIC DATA, SHAFT STRUCTURE, SUBCOMMENT ON SEMANTICS

1. Inleiding

Semantiese data is dié datatipe waarvoor die meeste gebruikers van algemene eentalige en tweetalige woordeboeke hulle tot hierdie naslaanbronne wend. In tweetalige woordeboeke is die soektog veral gerig op die vind van 'n gepaste vertaalekwivalent om die betekenis van 'n brontaalvorm in die doeltaal uit te druk. In eentalige woordeboeke is dit veral 'n verklaring van betekenis waaraan gebruikers 'n behoefte het. Die aanbod van semantiese data is egter nie beperk tot aanduiders van vertaalekwivalente en betekenisparafrases nie. Na gelang van die tipologiese aard van 'n woordeboek, bied woordeboekartikels op semantiese vlak ook leiding ten opsigte van onder meer betekenisbetrekkinge, deursigtige en ondeursigtige gelede woorde, deiktiese waardes van linguistiese uitdrukkinge, semantiese gebruiksbeporinge asook pragmatiese data. Aangesien algemene eentalige en tweetalige woordeboeke ook vir ander datatipes voorsiening moet maak, lei die ruimteprobleme in gedrukte woordeboeke daartoe dat daar selde 'n uitvoerige aanbod van semantiese data kan wees; selfs in omvattende woordeboeke. Na gelang van die woordeboeksoort maak leksikograwe 'n keuse uit die verskillende tipes semantiese data en bied dan slegs aanduiders van hierdie tipes aan. Die omvang van die weergawe van elk van hierdie aanduidertipes word ook deur die woordeboek se tipologiese beperkinge bepaal. Hierdie keuse word normaalweg op grond van die behoeftes van die teikengebruiker gemaak, maar die meeste woordeboektypes kan beswaarlik aan al die behoeftes van al die voornemende gebruikers voldoen. Alhoewel aanduiders van tipes semantiese data dikwels die besetting van soeksones in 'n woordeboekartikel oorheers, bly dit 'n beperkte aanbod wat lei tot 'n beperkte leksikografiese bewerking van die leksikale items wat deur die lemmas verteenwoordig word.

Aanlyn woordeboeke het nie dieselfde ruimteprobleme as gedrukte woordeboeke nie. Alhoewel die aanbiedingsruimte beperk is weens die skermgrootte van die rekenaar, tablet of slimfoon, laat die bergingsruimte groot hoeveelhede data toe. Leksikograwe van aanlyn woordeboeke kan, waar nodig, dus voorsiening maak vir 'n uitvoeriger aanbod van elke datatipe in vergelyking met die aanbod wat in 'n ooreenstemmende gedrukte woordeboek verskyn.

Hierdie artikel is gerig op die vind van moontlike oplossings om die omvangs-probleem van die aanbod van semantiese data in aanlyn woordeboeke te oorkom. Met inagneming van ander datatipes wat ook in woordeboeke aangebied word, is die fokus in hierdie artikel uitsluitlik op semantiese data. Die aard van die woordeboeksoort, die struktuur van die woordeboekartikel en verskillende soeksones in die artikel as soekgebied word onder die loep geneem.

2. Ter agtergrond

Voordat voorstelle vir 'n nuwe aanbod van semantiese data in aanlyn woordeboeke bespreek word, is dit noodsaaklik om kortliks aandag te gee aan enkele aspekte van gedrukte woordeboeke, by name woordeboektipologie, woordeboekstrukture asook enkele aspekte van semantiese data in gedrukte woordeboeke. Hierdie laasgenoemde aspek word voorafgegaan deur enkele verwysings na die semantiek as linguistiese studierigting.

2.1 Aspekte van woordeboektipologie

In die inleidende paragraaf is daar verwys na semantiese data in algemene eentalige en tweetalige woordeboeke. Hierdie verwysing het veral te make met die soort tipologiese indeling wat op gedrukte woordeboeke gerig is. Woordeboektipologie vorm 'n wesenlike deel van leksikografiese besprekings. Scerba (1940) het reeds gewys op die belang van tipologiese verskille tussen woordeboeke. Zgusta (1971: 198, e.v.) bied 'n uitvoerige tipologiese klassifikasie wat as basis dien vir talle latere besprekings van woordeboektipologie.

Die vroeë klassifikasie van woordeboeksoorte was uiteraard op gedrukte woordeboeke gerig en dit geld vandag steeds in 'n hoë mate — ook in talle beskouinge van die aanlyn leksikografie. Die tradisionele woordeboekklassifikasie het woordeboeke nie in waterdigte kompartemente geplaas nie, maar wel voorsiening gemaak vir oorvleueling tussen verskillende tipes. Desondanks was daar duidelike kriteria waarvolgens enige woordeboek in 'n bepaalde kategorie geplaas is. Verskillende woordeboeksoorte het verskillende gebruikersgroepe gehad en die aard en aanbod van die verstrekte data is in terme van die woordeboeksoort, die teikengebruikers, en die leksikografiese behoeftes en woordeboekvaardighede van die teikengebruikers bepaal. Algemene eentalige en tweetalige woordeboeke is in die eerste plek as linguistiese instrumente gesien wat linguistiese data aan hulle gebruikers beskikbaar moes stel. Die keuse en aanbod van data is grootliks deur die woordeboeksoort bepaal en dit het beperkte geleentheid

geskep vir verandering en vernuwing.

Wat die hantering van betekenisaspekte in eentalige verklarende woordeboeke betref, het die onderskeid tussen omvattende woordeboeke, standaardwoordeboeke en pedagogiese woordeboeke aan elk van hierdie woordeboeksoorte 'n bepaalde datapakket voorgehou waaraan die leksikograwe hulle moes hou.

2.2 Aspekte van woordeboekstrukture

In die aanbod van semantiese data in woordeboeke speel verskillende strukture 'n rol. Vir die plasing en vind van semantiese data is die dataverspreidingsstruktuur, die toegangstruktuur en die data-onttrekkingstruktuur van die hoogste belang. Vir die doel van hierdie bespreking is dit egter veral die artikelstruktuur en sy onderskeie komponente en elemente wat ter sprake gebring word.

2.2.1 Artikelstruktuur

Algemene eentalige en tweetalige woordeboeke vertoon 'n vaste artikelstruktuur met data wat veral in óf die vorm- óf die semantiese kommentaar gehuisves word. Woordeboekartikels bevat twee tipes elemente, te wete aanduiders en struktuurmerkers (vgl. Wiegand 1989a en Wiegand en Smit 2013). Aanduiders is mikrostrukturele elemente waaraan leksikografiese inligting onttrek kan word. Struktuurmerkers is elemente waaraan geen leksikografiese inligting onttrek kan word nie, maar wat bepaalde datatipes en soeksones merk. Semantiese data word deur middel van aanduiders in woordeboekartikels weergegee, en hierdie datatipe word deur middel van struktuurmerkers van ander datatipes onderskei.

Die semantiese kommentaar bevat verskillende tipes semantiese data soos die aanduiders van die betekenisparafrase, antonieme, sinonieme en, in die geval van tweetalige woordeboeke, vertaalekwivalente. In die geval van lemmas wat 'n polisemiese leksikale item verteenwoordig, bevat die semantiese kommentaar 'n semantiese subkommentaar vir elkeen van die verskillende polisemiese waardes. 'n Semantiese subkommentaar bevat die tersaaklike betekenisparafrase of vertaalekwivalent. Die semantiese kommentaar, sowel as die semantiese subkommentare, bied ook ruimte vir die weergawe van koteksaanduiders. Sulke aanduiders word nie net ter bevrediging van 'n teksproduksiefunksie verstrekk nie, maar ook om die teksbegripfunksie ter wille te wees. Daarom is dit belangrik dat daar 'n duidelike verband tussen 'n betekenisparafrase of vertaalekwivalent en die verstrekte koteksinskrywings moet wees. Vir 'n ondubbelsinnige vergestaltung van hierdie verband word daar voorsiening gemaak vir verskillende tipes mikrostrukture. Dit is veral die geïntegreerde en die niegeïntegreerde mikrostrukture (Wiegand 1989a: 483, 488) wat hier ter sake is en wat die verwarring teenwerk wat spruit uit die gebruik van 'n primitiewe mikrostruktuur (vgl. Gouws 2003: 107).

In die geval van 'n geïntegreerde mikrostruktuur tree een of meer koteks-

aanduiders in dieselfde semantiese subkommentaar, 'n integraat (sien Wiegand 1989a: 484 en Wiegand en Smit 2013: 176), as die betekenisparafrase of vertaal-ekwivalent op. Figuur 1, 'n deelartikel uit die HAT, bevat vyf semantiese subkommentare en telkens word die betekenisparafrase-aanduiders aangevul deur een of meer koteksaanduiders wat in dieselfde integraat geïntegreer is. Die nabyheid van koteksaanduiders aan die betekenisparafrase-aanduiders skep 'n verhouding van onmiddellike adressering wat vir die gebruiker 'n kitsonttrekking van inligting moontlik maak.

groot-heid *s.nw.* **1** [geen mv.] eienskap van groot te wees; omvangrykheid; belangrikheid: *grootheid van liggaam* ○ *Watter maatstawwe word gebruik om die grootheid van 'n kunsversameling te bepaal?* **2** [**grootheide**] (*wisk.*) hoeveelheid: *meetbare/verdeelbare/vermenigvuldigbare grootheide*. **3** [geen mv.] verhewenheid; heerlikheid; luister; waardigheid; hoogheid: *die grootheid van God*. **4** [geen mv.] verhewenheid van gees; karakteradel: *grootheid van gees*. **5** iemand van belang/aansien: *onder die grootheide van sy/haar tyd gereken word*. ♦ **die onbekende grootheid wees** die faktor wees waarvan die invloed/uitwerking nie voorspel/bepaal kan word nie.

Figuur 1: Uit die HAT

In die geval van 'n niegeïntegreerde mikrostruktuur word al die koteksinskriftings wat vir elk van die betekenisonderskeidings verstrek word saam in een teksblok geplaas wat gewoonlik op die teksblokke volg wat die verskillende semantiese subkommentare huisves. Sien in hierdie verband figuur 2, 'n artikel uit die *Prisma Groot woordenboek Afrikaans en Nederlands*:

rond (voorz.), **rond** (voors.) **1** (om iets heen) *om, rond* **2** (in de buurt van, omheen) *rondom, om en by* **3** (met betrekking tot) *rondom* ♦ **1** rond het vat zitten stalen banden *om die vat is daar staal bande*; rond de tafel zitten *om die tafel sit*; er staat een wisselend magnetisch veld rond de stroomdraad *daar is 'n wisselende magnetiese veld om die stroomdraad*; ze renden de kamer rond *hulle het deur die kamer rondgehardloop* **2** het zijn enkele van de vergeten dorpen rond Dokkum *dit is 'n paar van die vergete dorpe rondom Dokkum*; een band van geladen deeltjes rond de aarde *'n band van gelaai deeltjes rondom die aarde*; rond de klok van vijf uur *om en by vyf uur*; de koers schommelde rond de honderd punten *die koers het om honderd punte rond geskommel*; zij is rond de dertig *sy is om en by dertig*; rond Pasen *om en by Paasfees* **3** rond het verschijnsel bestaan daar verskeie legenden *rondom die verskynsel bestaan daar verskeie legendes*; de geruchten rond zijn benoeming *die gerugte rondom sy benoeming*

Figuur 2: Uit die *Prisma Groot woordenboek Afrikaans en Nederlands*

Die artikel van die lemma *rond* vertoon drie semantiese subkommentare. Elkeen bevat, in hakies, 'n bondige betekenisverklaring, gevolg deur die tersaaklike vertaalekwivalent(e). Na afloop van die derde semantiese subkommentaar is daar 'n diamantvormige nietipografiese struktuurmerker wat ter wille van kitstoe-gang die artikeldeel merk waar die koteksaanduiders verskyn. Hierna is daar vir elk van die voorafgaande semantiese subkommentare 'n artikelgleuf wat die gepaste koteksaanduiders bevat. Elk van hierdie artikelgleuwe het 'n nommer wat ooreenstem met 'n nommer, as polisemiemerker, van een van die voorafgaande semantiese subkommentare. Daardeur word elke koteksgleuf verbind met 'n semantiese subkommentaar wat dieselfde nommer het. Alhoewel elke semantiese subkommentaar nie koteksaanduiders as deel van die betrokke inte-graats bevat nie en daar 'n verhouding van afstandsadressering tussen 'n koteks-aanduiders en die tersaaklike vertaalekwivalent bestaan, maak die eksplisiete nume-ringsafparing dit vir gebruikers moontlik om probleemloos die tersaaklike koteks-aanduiders vir 'n bepaalde vertaalekwivalent te vind.

2.3 Enkele aspekte van die semantiek

Algemene verklarende woordeboeke word veral geraadpleeg vir die onttrek-king van taalkundige inligting. Dit is veral data wat met die betekenis van 'n leksikale item te make het waarna gesoek word en dan veral daardie aandui-ders wat 'n verklaring van die betekenis van 'n woord bied — die aanduiders van betekenisparafrases. As houers van taalkundige data is woordeboeke van die afgelope dekades daarop gerig om 'n suiwer linguistiese verslag van die lek-sikon te bied. Ook wat die weergawe van semantiese data betref, is daar gepoog om nielinguistiese data uit die betekenisparafrase-aanduiders te weer. Woorde-boeke is skerp gekritiseer indien daar 'n vermenging van linguistiese en ensiklo-pediese data in die verklaring van betekenis aangebied is. Vergelyk in hierdie verband die kritiek op betekenisverklaring in vroeër dele van die WAT, byvoor-beeld Combrink (1979) en Odendal (1979).

Volgens Gouws (om te verskyn) poog natuurlike semantiese metataal, 'n teoretiese benadering in die linguistiek, om betekenis in te kort tot 'n beperkte stel semantiese primitiewe. Semantiese primitiewe is die eenvoudigste linguis-tiese konsepte. Hulle kan nie gedefinieer word deur eenvoudiger terme te gebruik nie (vgl. Wierzbicka 1972 en Levisen en Waters 2017). Om betekenis met behulp van semantiese primitiewe te identifiseer, lei tot 'n suiwer linguistiese beteke-nisverklaring wat universele waarde het. Fodor (1977: 155-156) beweer dat in generatiewe grammatika "a lexical item is substituted for a constituent of a phrase marker whose terminal elements are semantic primitives". Met behulp van sulke primitiewe kan die betekenis van die woord *doodmaak* beskryf word as "veroorzaak om nielewend te word". Ook in die strukturele semantiek lei die benadering van komponentanalise daartoe dat 'n woord se betekenis in kom-ponente verdeel word wat slegs ruimte laat vir linguistiese data en geensins vir gebruiks- of ensiklopediese data nie. Kenmerkend hier is die weglating van alle

persoonlike en nielinguistiese kenmerke.

Gouws (om te verskyn) wys daarop dat die kognitiewe semantiek 'n ander benadering volg. Hiervolgens is taal gewortel in die individu se ervaring, insluitend sintuiglike ervarings. 'n Betekenisverklaring wat op hierdie benadering gebaseer is, strek dus verder as semantiese primitiewe en komponente en bevat ook buitelinguistiese kenmerke. Volgens DiMuccio-Failla (2025: 27) se kognitiewe benadering tot betekenis, kan woorde beskou word as verteenwoordigers van ervaringskategorieë. Dit het noodwendig 'n invloed op betekenisverklaring en waar hierdie benadering in woordeboeke gevolg word, sal die betekenisparafrase veel meer insluit as die data wat deur semantiese primitiewe en komponente weergegee word. Klotz en Herbst (2016: 72) het reeds aangedui dat 'n woordeboekdefinisie 'n beskrywing is van die betekeniskenmerke van 'n leksikale item gebaseer op sy gebruik in 'n taalgemeenskap.

Uit die voorafgaande bespreking kan die afleiding gemaak word dat woordeboeke wat slegs op 'n streng linguistiese interpretasie van betekenis gerig is, nie noodwendig die volle omvang van dit wat die taalgemeenskap as betekenis ervaar, weergee nie. Nuwe benaderings in die aanlyn leksikografie maak voorsiening vir aanpassings in hierdie verband.

2.4 Semantiese data in gedrukte woordeboeke

In die hieropvolgende paragrafe word enkele aspekte van semantiese data in gedrukte verklarende woordeboeke bespreek. Dit bied 'n agtergrond vir die bespreking in 'n daaropvolgende afdeling oor semantiese data in aanlyn woordeboeke.

2.4.1 Komponentanalise

Linguistiese teorieë is ten dele ter sake in die leksikografie. Die leksikografie is egter 'n dissipline uit eie reg en nie 'n subdissipline van die linguistiek nie. Daarom speel linguistiese teorieë nie 'n bepalende rol in besluite oor die opname en aanbieding van data in woordeboeke nie, alhoewel aspekte van sulke teorieë dikwels wel deur leksikograwe oorweeg word. Bepaalde benaderings in die taalwetenskap kan aangepas word vir 'n spesifieke leksikografiese gebruik.

'n Voorbeeld van so 'n aanpassing is te vinde in die linguistiese siening van semantiese primitiewe en komponentontleding. In die linguistiek lei hierdie benaderings tot 'n suiwer taalkundige analise van woordbetekenis. Die individu se ervaring van betekenis asook Wittgenstein (1953) se siening dat die betekenis van 'n woord deur sy gebruik in die taal bepaal word, word nie deur so 'n linguistiese benadering in ag geneem nie. In die leksikografiese hantering van semantiese data kan die benutting van byvoorbeeld betekenis Komponente 'n waardevolle hulp wees, maar 'n toepassing daarvan moet nie die eindpunt van die semantiese analise wees nie. Hier is 'n aanpassing nodig om vir die eiesoortigheid van leksikografiese betekenisverklaring voorsiening te maak. Na aan-

leiding van die werk aan Louw en Nida (1988), dit is die innoverende Grieks-Engelse woordeboek van die Nuwe Testament wat op semantiese velde gebaseer is, stel Louw (1985) 'n benadering bekend wat linguistiese teorie vir die leksikografie aanpas. In die analise van betekenis word semantiese komponente gebruik — Louw werk met drie komponenttipes, te wete gemeenskaplike komponente, diagnostiese komponente en supplementêre komponente. Die gebruik van die eersgenoemde twee tipes lei tot 'n produk vergelykbaar met dié vanuit die linguistiese benadering van komponentanalise. Hierdie analise word dan, waar nodig, aangevul met die gebruik van supplementêre komponente. Hierdie komponenttipe bied data wat nie streng semanties is nie en wat iets van 'n woord se gebruik of sy ensiklopediese waarde weergee. Die insluiting van supplementêre komponente bied 'n leksikografiese werkswyse wat die tradisionele komponentontleding van die strukturele taalkunde verbind met aspekte van die kognitiewe semantiek om 'n gepaste woordeboekbewerking moontlik te maak.

Die gebruik van supplementêre komponente het 'n invloed op die aard en omvang van die betekenisparafrase aangesien bykomende, selfs nielinguistiese, data daarby ingesluit word. Betekenisparafrases in die artikels van gedrukte woordeboeke bied ongelukkig dikwels nie voldoende ruimte vir 'n data-aanbod wat alle gebruikers se kennisbehoefte bevredig nie; veral omdat die betekenisparafrase nie 'n genoegsame hantering van supplementêre komponente toelaat nie. Dié situasie kan ingrypend verbeter word in aanlyn woordeboeke.

2.4.2 Betekenisbetrekkinge

Die leksikale semantiek sluit meer in as die betekenisverklaring van leksikale items. Die leksikon vertoon 'n netwerk van semantiese betrekkinge wat deel is van die studie-objek van die leksikale semantiek. As draers van betekenisdata moet verklarende woordeboeke ook van betekenisbetrekkinge verslag doen. Aan betrekkinge soos antonimie en sinonimie gee verklarende woordeboeke gereeld aandag, maar die spesifieke aard van 'n bepaalde sinonimiese betrekking, byvoorbeeld of dit absolute al dan gedeeltelike sinonieme is wat aangedui word, word selde vermeld. Koteksaanduiders om die voorkoms van die tersaaklike sinonimiese betrekking te illustreer, ontbreek meestal ook.

Naas sinonimie vorm hiponimie deel van die verskillende tipes betekenisbetrekkinge van semantiese insluiting. Waar woordeboeke die genus en differentia-metode gebruik in die betekenisparafrase van 'n woord wat 'n hiponiem is, is daar 'n implisiete aanduiding van 'n hiponimiese verhouding deurdat die superordinaat as gemeenskaplike komponent verstrek word. Dit is duidelik waar HAT die betekenis van die woord *kool* verklaar as "enigeen van die groentesoorte en -kultivars van die geslag *Brassica* met dikkerige groen of rooi blare wat styf inengevleg is en 'n harde kop vorm". Hier tree *groentesoort* as superordinaat van *kool* op. Hoofsaaklik weens ruimtebeperkinge ontbreek breër semantiese leiding ten opsigte van ander hiponieme van dieselfde superordinaat in die meeste woordeboekartikels. In die betekenisparafrase van 'n woord wat 'n

superordinaat is, word voorbeelde van hiponieme dikwels genoem. Vergelyk HAT se verklaring van *groente* as: "plante/plantdele wat 'n mens rou of gaarge-maak (kan) eet, soos kool, blomkool, aartappels, (geel)wortels en ertjies ...". Hier word meer semantiese leiding verstrekk, maar gebruikers kry steeds slegs 'n beperkte deel van die betrokke semantiese veld.

Antonimie is een tipe betekenisteenoorgesteldheid. Woordeboeke toon dikwels antonieme aan of hulle skep 'n soeksone vir die weergawe van betekenisteenoorgesteldes. Vir die woordeboekgebruiker op soek na inligting oor tipes teenoorgesteldes is die tipiese leksikografiese bewerking onvoldoende. Daar word nie genoegsaam inligting verstrekk om die gebruiker te help om tussen byvoorbeeld antonimie, omgekeerdheid en komplementariteit (Lyons 1977: 270-280) te onderskei nie.

2.5 Onverklaarde gelede woorde

Eentalige verklarende woordeboeke het dikwels die gebruik om onverklaarde gelede woorde op te neem. Hierdie woorde kwalifiseer as lemmakandidate, maar weens ruimtebeperkings kry hulle geen volledige bewerking nie. Die implikasie van opname as onverklaarde lemma is dat die gelede woord semanties deursigtig is en dat die betekenis afgelei kan word van die betekenis van die samestellende dele. Deursigtigheid is nie altyd maklik vasstelbaar nie en leksikograwe mag hulle nie op die intuïsie van hulle gebruikers beroep en aanvaar dat dit wat vir die leksikograaf deursigtig is ook vir die gebruiker sal wees nie.

Begripsprobleme duik maklik op, onder meer waar die woordvariant van een of meer stamme van 'n gelede woord polisemies is. Die WAT bied byvoorbeeld agt betekenisonderskeidings en 'n reeks subonderskeidings in die bewerking van die selfstandige naamwoord *kop*. In die lys onverklaarde lemmas verskyn onder meer *kopprofiel* en *kopskub*. Behalwe dat die betekenis van hierdie woorde nie sonder meer deursigtig is nie, het die gebruiker ook geen duidelikheid oor watter een van die betekenisonderskeidings van die lemma *kop* in elke onverklaarde gelede woord ter sake is nie. Vergelyk ook die bespreking van hierdie probleem in Gouws (1989: 77, e.v.).

Sommige woordeboeke bied meer hulp in hierdie verband, byvoorbeeld *Nasionale woordeboek* wat in sy aanbieding van onverklaarde lemmas 'n benadering van neslemmatisering (vgl. Wiegand 1989b: 391) volg, en telkens aantoon watter polisemiese waarde van die woord wat deur die voorafgaande hooflemma verteenwoordig word in die eerste stam van die gelede woord geaktiveer word. Dit is duidelik uit die bewerking in figuur 3. Die artikel van die lemma *stroom* is sesvoudig polisemies. In die lys onverklaarde lemmas word met behulp van 'n deeltrajekpolisemiemerker "(by 2)" aangetoon dat die tweede polisemiese waarde ter sake is by die onverklaarde lemmas *stroombed* en *stroombedding*, terwyl polisemiese waarde ses geaktiveer word in die sublemmas *stroombaan* — *stroomverdeling*.

stroom I (strome) s.nw. 1. *bewegende massa van 'n vloeistof, veral water.* – op, na, in die rigting van waar 'n stroom, bv. 'n rivier, vloei. Die reën val in strome. 'n – lawa. Strome bloed. 2. *stroming, beweging in 'n bepaalde rigting.* Lug-, seestrome. Gevaarlike strome langs die kus. Met die – saam, sonder om teenstand te bied. Teen die – op, teen die algemene neigings, opvattinge. – af, op gaan, met, teen die stroom. Die – volg, met die meerderheid saamgaan. 3. *rivier, spruit.* Die – kronkel tussen koppies. 4. *voortbewegende menigte.* 'n – mense, motors. 'n – van vloeke, woorde, idees. 'n – boeke, menigte. 5. *bewegende materie.* 'n – gas, lug, suiker. 6. *hoeveelheid elektriese vermoë.* Hoeveel – gebruik ons? 'n Sterk (elektriese) –. 'stroombed, –bedding (by 2); –baan, –draad, –lewering, –meter, –spanning, –verbruik, –verdeling (by 6).

Figuur 3: Uit *Nasionale woordeboek*

'n Ondubbelsinnige vasstelling van die spesifieke betekeniswaarde van onverklaarde lemmas is nie probleemloos nie. Hier moet leksikograwe aandag gee aan Gove (1966: 183) wat aanvoer dat die *self-* in *self-explanatory* na die interpreterder van die woord moet verwys en nie na die woord as sodanig nie — vergelyk ook Gouws (1989: 78).

In talle verklarende woordeboeke skiet die implisiete betekenisleiding by die weergawe van onverklaarde gelede woorde te kort. Teikengebruikers van 'n betrokke woordeboek kan nie sonder meer die betekenisinligting aan die aangebode data onttrek nie. Dit striem suksesvolle semantiese leiding in die artikels van verklarende woordeboeke.

3. Aanlyn woordeboeke

3.1 Woordeboektipologie

Tarp (2009: 24) onderskei tussen 'n kontemplatiewe en 'n transformatiewe benadering in die leksikografie. Eersgenoemde bring mee dat leksikograwe op dieselfde manier voortwerk aan hulle woordeboeke as hulle voorgangers, terwyl die laasgenoemde benadering daartoe lei dat leksikograwe vernuwend optree in die beplanning en samestelling van woordeboeke. Ook in die oorgang van gedrukte na aanlyn woordeboeke is daar dikwels 'n kontemplatiewe benadering met dieselfde werkswyse wat gehandhaaf word. Dit geld ook die tipologiese klassifikasie, want die indeling wat vir gedrukte woordeboeke geld, kan sonder meer of met slegs geringe aanpassings ook in die aanlyn omgewing

gebruik word. Die aanlyn omgewing het egter 'n transformatiewe potensiaal en bied moontlikhede wat tot wesenlike aanpassings in die bestaande woordeboektipologie kan lei. Dit het implikasies vir baie aspekte van die leksikografiepraktik, onder meer vir die data-aanbod en die dataverspreidingsstruktuur.

In die aanlyn leksikografie speel toegang tot data 'n deurslaggewende rol. Belangriker as die beskikbaarheid van 'n bepaalde woordeboeksoort is die beskikbaarheid van die gepaste data — ongeag of dit in 'n woordeboek of in 'n databasis gevind word. Toegang tot data kan ook geïndividualiseer word sodat elke gebruiker 'n unieke toegang tot die spesifieke data kry waaraan hulle 'n behoefte het. In hierdie verband bied Gouws (2006) en Bergenholtz en Bergenholtz (2017) aanduidings van meerdere woordeboeke wat aan een databasis onttrek kan word, terwyl Bergenholtz, Bothma en Gouws (2011) wys op toegang tot vooraf geselekteerde datatipes. Die beginsels wat in hierdie publikasies ter sprake gebring is, kan in die hedendaagse aanlyn leksikografie met veel meer verfyning en presiesheid toegepas word — ook wat die beskikbaarstelling van semantiese data betref. Dit kan lei tot 'n vernuwende aanpassing in die tipologiese klassifikasie van leksikografiese bronne.

Aan 'n noukeurig beplande en gemerkte databasis kan verskillende woordeboeke onttrek word. Die data in die databasis van so 'n "moederwoordeboek" (vergelyk Gouws 2006), kan gemerk word vir spesifieke woordeboeksoorte, byvoorbeeld 'n eentalige verklarende skoolwoordeboek, aanleerderwoordeboek, standaardwoordeboek of omvattende woordeboek. Na gelang van die gebruikersprofiel, soos aangedui deur 'n gebruiker wanneer die databasis benut word, kan verskillende bewerkings van dieselfde lemma gebied word. Vir teksbegripdoeleindes kan verskillende aanduiders van die betekenisparafrase aan die databasis onttrek word en vir teksproduksiedoeleindes kan verskillende koteksaanduiders verskyn. Tradisionele woordeboektipologie is hier nie so belangrik nie, maar wel die behoeftes van die gebruiker — en dit word in die gebruikersprofiel gespesifiseer. Waar 'n gebruiker se profiel hom/haar as 'n aanleerder van 'n taal spesifiseer, sal die tersaaklike aanduiders onttrek word. Aan dieselfde databasis onttrek die gebruiker wat as moedertaalspreker met 'n behoefte aan omvattende betekenisleiding gespesifiseer is, ander tersaaklike inligting.

Die databasis kan ook so opgestel word dat woordeboeke wat slegs enkele datatipes bevat daaraan onttrek kan word. Die databasis bevat al die leksikografiese data vir 'n woordeboek wat die volle spektrum van datatipes dek, maar 'n gebruiker wat slegs 'n behoefte aan, byvoorbeeld, uitspraakhulp het, onttrek slegs aanduiders wat die uitspraak van 'n betrokke woord weergee. Op dieselfde manier kan 'n woordeboek onttrek word wat slegs semantiese data bevat. Telkens kan die leksikograaf aandui watter tipes aanduiders ter sake is. In die geval van 'n uitspraakwoordeboek kan dit byvoorbeeld 'n fonetiese transkripsie van die woord wees of dit kan primêre en sekondêre klem, asook sillabeverdelings aantoon. In die geval van 'n betekeniswoordeboek is verskeie aanduidertipes moontlik, byvoorbeeld 'n betekenisparafrase, antonieme en sinonieme.

3.3 Woordeboekstrukture

Aanlyn woordeboeke vertoon 'n verskeidenheid strukture wat ooreenstem met asook verskil van strukture in gedrukte woordeboeke. Sommige strukture, byvoorbeeld die makrostruktuur, word met bepaalde aanpassings gehandhaaf. Ander strukture, byvoorbeeld die raamstruktuur, val weg in aanlyn woordeboeke, terwyl nuwe strukture, byvoorbeeld die skermkootstruktuur, na vore tree. Dit is veral die artikelstruktuur en strukture wat in en vanuit die artikel ingespan word wat vernuwingsmoontlikhede in die data-aanbod, ook dié van semantiese data, tot voordeel kan strek.

3.3.1 Die artikelstruktuur in aanlyn woordeboeke

In die hieropvolgende bespreking is die fokus uitsluitlik op aspekte van die semantiese kommentaar van die artikels van aanlyn woordeboeke. In die semantiese kommentaar vertoon aanlyn woordeboeke dikwels dieselfde statiese struktuur wat in gedrukte woordeboeke voorkom. Vergelyk in hierdie verband figuur 4, die artikel van die lemma *taksonomie* in die aanlyn WAT waarvan die statiese artikelstruktuur baie ooreenstem met dié van artikels in die gedrukte dele van die WAT:

taksonomie (tak-so-no-mie) s.nw. (met mv. in bet. 2 b en 3 b) (←Gr. *taxis*, “rangskikking” en *nomos*, “norm, reël”)

1 Wetenskap of leer van klassifikasie, asook die beginsels en wette wat binne sodanige klassifikasie geld.

Vgl. **SISTEMATIEK 1**.

2 a Afdeling v.d. biologie wat gerig is op die benaming, beskrywing en klassifikasie van lewende organismes volgens verwantskappe en verskille: *Reeds in Linnaeus se tyd is sowel kwantitatiewe as kwalitatiewe data in die taksonomie gebruik* (J. Heyns: Taksonomie, 1973, 15). *Die komitee doen ... 'n voortdurende studie oor die wêreld se wild, hou by met die menings wat taksonomie betref, en hersien die kategorieë van trofeediere* (J. Lötter, vert.: Jag, 2002, 146).

Vgl. **SISTEMATIEK 2**.

b mv. **taksonomieë**. Voorbeeld van 'n klassifikasie binne die taksonomie (**TAKSONOMIE 2 a**): *Die grootste deel van my lewe was ek besig met die taksonomie van ons insekte* (Volksbl., 23 Mei 2017, E). *Bespreking van die taksonomie en beskrywings van die anatomie en uiterlike voorkoms van die bontebok is te vinde* (Koedoe, 1976, 67). *Goeie wetenskaplike versamelings en taksonomieë het .. voortgespruit uit studies van blaarvlooie en blaarspringers, parasietwespe, dopluise en sprinkane* (Panorama, Des. 1987, 39).

Figuur 4: Uit die aanlyn WAT

Hierdie artikel vertoon 'n mikrostrukturele ordeningstruktuur met aanduiders in die semantiese kommentaar wat vertikaal georden is. 'n Vertikale hiërargiese ordening kom algemeen in woordeboekartikels voor en in gedrukte woordeboeke bied dit gemaklike interne soekroetes. 'n Vertikale ordening lei tot 'n skagstruktuur met die gebruiker wat in die interne toegangstruktuur 'n soekroete van bo na onder volg.

Aanlyn woordeboeke vertoon soms ook 'n horisontale ordening in die arti-

kelstruktuur, soos te sien in figuur 5 uit die ANW. Die linkerkantste kolom in die skermkoot vertoon die makrostrukturele deeltrajek waarbinne die lemma *hondenfluitje* pas. Die volgende kolom is die begin van die artikel en bevat 'n spyskaart van datatipes en ander inskrywings waartoe die gebruiker in en vanuit die artikel toegang het. Die daaropvolgende kolom bied 'n gedeeltelike bewerking van die eerste polisemiese waarde van *hondenfluitje*, terwyl die regterkantste kolom bepaalde grammatiese aanduiders bevat.

The screenshot shows the ANW website for the word 'hondenfluitje'. The left sidebar contains a letter index from A to Z. The main content area is divided into sections: 'Per betekenis' (Meanings), 'Per onderwerp' (Topics), 'Semagram (extra betekenisinformatie)' (Semagram), and 'Algemene voorbeelden' (General examples). The right sidebar contains grammatical information under 'Woordsoort' (Word type), 'Spelling en flexie' (Spelling and inflection), and 'Woordvorming' (Word formation).

Woordsoort	
Type	zelfstandig naamwoord
Naamtype	soortnaam
Geslacht	onzijdig
Lidwoord	het
Betekenisklasse	zaaknaam

Spelling en flexie		
Vorm	Aftrekking	
Enkelvoud	hondenfluitje	hon.den.fluit.je
Meervoud	hondenfluitjes	hon.den.fluit.jes

Woordvorming	
Type	samenstelling
Linkerlid	hond
Linkerlidtype	zelfstandig naamwoord
Rechterlid	fluitje

Figuur 5: Uit die ANW

Dit is 'n dinamiese artikelstruktuur — 'n klik op 'n spyskaartitem vul die middel-en/of regterkantste blok met die tersaaklike nuwe data. 'n Klik op byvoorbeeld die spyskaartinskrywing *Voorbeelden*, lei tot figuur 6 waar die regterkantste blok van figuur 5 verdwyn en die oorspronklike middelblok, nou die regterkantste blok, deur die bondige betekenisfrase en meegaande voorbeeldaanduiders beset word.

Die voorbeeldaanduiders wat hier verstrekk word, is ter illustrasie van die spesifieke betekenisonderskeiding van die woord wat deur die lemma verteenwoordig word. Dit word in dieselfde integraat as die betekenisparafrase-aanduiders verstrekk en gesamentlik vertoon om as voorbeeld van 'n geïntegreerde mikrostruktuur te dien. Anders as in die geval van figuur 4, die voorbeeld uit die WAT, is die voorbeeldaanduiders nie deel van die eerste skermkoot waarin betekenisparafrase 1 vertoon word nie. Die voorbeeldaanduiders word eers na 'n verdere klik sigbaar in die integraat. Hier is dus sprake van 'n uitgestelde geïntegreerde mikrostruktuur. Die sigbaarmaking van die geïntegreerde mikrostruktuur word uitgestel tot na die klik van 'n data-identifiseerder.

<https://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2075> (Article)

59

<https://www.ivdnt.org/article/hondenfluitje#subjectvoorbeelden>

Figuur 6: Uit die ANW

Die horisontale plasing van aanduiders en teksblokke kan ook tot verdere tipes geïntegreerde mikrostrukture lei. Vergelyk in hierdie verband figuur 7–10, dele van die Duitse woordeboek *lexico* se hantering van die lemma *Pferd*:

					elektro
Pferd	Lesartenübergreifende Angaben				
 Orthografie					
Normgerechte Schreibung:	Pferd				
Worttrennung:	Dieses Wort ist nicht trennbar.				
 Herkunft und Wandel					
Etymologische Angaben:	anzeigen >				
Wandel 1700 bis 1945:	-				
Wandel seit 1945:	-				
 Wortbildungsprodukte					
(automatisch ermittelt)	weiter >				
Lesartenbezogene Angaben 					
Lesart ' 'Reittier'	weiter >				
Mit <i>Pferd</i> bezeichnet man ein großes Säugetier mit langen Beinen, das vom Menschen bevorzugt als Reittier und gelegentlich auch als Zug- und Lasttier genutzt wird.					
Lesart ' 'Turngerät'	weiter >				
Mit <i>Pferd</i> bezeichnet man ein Turngerät mit vier langen Beinen, einem ledernen Aufbau und zwei Griffen auf der Oberseite.					
Lesart ' 'Schachfigur'	weiter >				
Mit <i>Pferd</i> bezeichnet man die Schachfigur, die als einzige über die anderen Figuren hinweg auf ein freies Feld bewegt werden kann.					

Figuur 7: Uit *ellexico*

Die semantiese kommentaar van hierdie artikel bevat verskeie subkommentare as integrale waarin die verskillende polisemiese waardes van die woord *Pferd* bewerk word. Deur 'n vertikale skagstruktuur in die veelvlakkige semantiese kommentaar beweeg 'n gebruiker na die relevante semantiese subkommentaar. Figuur 8 is die produk van 'n klik op die merker *weiter* (= verder) van 'n subkommentaar, hier dié van die betekeniswaarde *Reittier* (= rydier), as beperkte artikel (vergelyk Gouws en Bothma 2024: 358):



elexiko

Pferd

Lesart: 'Reittier'

zur Übersichtsseite Lesarten im Überblick

Bedeutungs-erläuterung Kollo-kationen Konstruk-tionen Sinnverwandte Wörter Gebrauchs-besonderheiten Grammatik

Erläuterung der Bedeutung / Funktion

Mit **Pferd** bezeichnet man ein großes Säugetier mit langen Beinen, das vom Menschen bevorzugt als Reittier und gelegentlich auch als Zug- und Lasttier genutzt wird.

Belege anzeigen » Illustrationen anzeigen »

Wortklasse: Individuativum

Figuur 8: Uit *elexico*

Die betekenisparafrase is die enigste aanduider in hierdie beperkte artikel, maar daar is ook die data-identifiseerders *Belege* (= sitate) en *Illustrationen* (= illustrasies) wat elk deur die merker *anzeigen* (= wys) gevolg word. 'n Klik op die data-identifiseerder *Belege* se *anzeigen* lei tot figuur 9, 'n beperkte artikel wat naas die betekenisparafrase-aanduider ook die voorbeeldaanduiders vertoon, waarvan slegs enkeles in hierdie afdruk van 'n deel van die skermkoot inpas. Ook hier word 'n uitgestelde geïntegreerde mikrostruktuur benut:



Pferd
Lesart: 'Reittier'

zur Übersichtsseite Lesarten im Überblick

Bedeutungs-erläuterung Kollo-kationen Konstruk-tionen Sinnverwandte Wörter Gebrauchs-besonderheiten Grammatik

Erläuterung der Bedeutung / Funktion

Mit **Pferd** bezeichnet man ein großes Säugetier mit langen Beinen, das vom Menschen bevorzugt als Reittier und gelegentlich auch als Zug- und Lasttier genutzt wird.

Belege verbergen × Illustrationen anzeigen »

Sie sind die Stars der Spanischen Reitschule in Wien und untrennbar mit der Geschichte Österreichs verbunden - die edlen Lipizzaner. Jetzt steht fest, die **Pferde** bleiben "rotweißrot". Die Würfel im kuriosen Streit um die Lipizzaner sind gefallen: Die weißen **Pferde** - weltweit ein renommiertes Aushängeschild für die hohe Kunst der Dressur - bleiben Österreicher! Bekanntlich hatte Italien die Zucht der Rasse für sich beansprucht. Landwirtschaftsminister Molterer konnte jetzt seinen Kollegen in Rom überzeugen, auf jegliche Forderungen zu verzichten. (Neue Kronen-Zeitung, 16.01.1999, S. 10, Lipizzaner bleiben Österreicher!)

«Seit der Abschaffung der Kavallerie hat ein kompletter Wandel stattgefunden», stellt Reitclub-Präsident Kurt Hättenschwiler fest. Hatte der Reitclub Flavil bei der

Figuur 9: Uit *ellexico*

Gouws en Bothma (2024: 360) toon aan hoe die verskillende vlakke wat elk een semantiese subkommentaar bevat verskillende vakke het, geïdentifiseer deur horisontaal-geordende data-identifiseerders, wat per soektonnel bereik kan word (Gouws en Bothma 2024: 362). 'n Klik op enigeen van die data-identifiseerders wat die vakke identifiseer, lei tot 'n beperkte artikel wat die lemma as gidselement bevat asook die eenwoordweergawe van die betrokke betekenisonderskeiding — hier *Reittier*. Die data in elkeen van hierdie vakke is gerig op die gebruik van die woord in die betrokke betekeniswaarde van die tersaaklike semantiese subkommentaar. In hierdie semantiese subkommentaar is dit nie net die voorbeeldaanduiders wat geïntegreer word nie, maar ook byvoorbeeld die kollokasies, konstruksies en verwante woorde. Figuur 10 wys hoe 'n klik op

die data-identifiseerder *Konstruktionen* tot 'n vak lei wat verskeie konstruksies vertoon:

elexiko

Pferd
Lesart: 'Reittier'

zur Übersichtsseite Lesarten im Überblick

Bedeutungs-erläuterung Kollo-kationen **Konstruk-tionen** Sinnverwandte Wörter Gebrauchs-besonderheiten Grammatik

Konstruktionen: Typische Verwendungen

Verwendungen mit Attribut [verbergen x](#)
das Pferd als Arbeitstier
Pferde im Galopp

Verwendungen in Verbalphrasen und Sätzen [verbergen x](#)
auf ein Pferd setzen
vom Pferd fallen
[Personenname] stürzte vom Pferd
rund [Zahl] Pferde sind gemeldet
bei dem Rennen sind [Zahl] Pferde am Start

Verwendungen als Attribut [verbergen x](#)
mit Pferd und Wagen
alles rund ums Pferd
zu Fuß oder mit dem Pferd
eine von Pferden gezogene Kutsche
von Pferden gezogene Wagen
Zucht- und Halteprämien für Pferde
Beziehung zwischen Mensch und Pferd
Harmonie zwischen Pferd und Reiter

Figuur 10: Uit *elexiko*

Verskillende tipes semantiese data word in die vakke van 'n enkele vlak aangebied en die dinamiese aard van die artikel maak dit moontlik vir 'n gebruiker om slegs dit te sien waarna hulle op soek is sonder om deur al die aanduiders hoef te worstel.

4. 'n Omvattender aanbod van semantiese data in Afrikaanse aanlyn woordeboeke

Die voorafgaande voorbeelde uit *lexiko* wys hoe 'n vernuwend gebruik van strukture in die woordeboekartikel tot 'n omvattender data-aanbod kan lei sonder om gebruikers aan data-oorlading bloot te stel. Afrikaanse verklarende woordeboeke kan op 'n soortgelyke manier verskillende vlakke en vakke in woordeboekartikels benut vir die aanbieding van semantiese data. In die hieropvolgende bespreking is die fokus op die aanbod van semantiese data in verklarende aanlyn woordeboeke. Tipologiese subklassifikasie van verklarende woordeboeke word nie hier in ag geneem nie, maar daar word van die standpunt uitgegaan dat leksikograwe oor die aard en omvang van semantiese data besluit na gelang van die tipologiese aard van 'n beplande woordeboek. Die voorstelle wat vervolgens gemaak word, sou ook op 'n enkele moederwoordeboek of databasis gerig kon wees waaraan verskillende woordeboeke onttrek kan word. Dit word egter nie hier bespreek nie.

4.1 Vlakke en vakke

'n Omvattende aanbod van semantiese data vereis 'n artikelstruktuur waarin data-oorlading vermy word, maar waarin vindbare data verstrekkend word waaraan inligting op 'n ondubbelsinnige manier onttrek kan word. Die uitleg van die woordeboekartikel en die meegaande gebruik van bepaalde strukture speel hier 'n beduidende rol. Müller-Spitzer en Koplenig (2014) gee aandag aan verskillende artikeluitlegmoontlikhede. Hulle gebruikersnavorsing het getoon dat gebruikers voorkeur gee aan die tipe uitleg soos in *lexico* waar versteekte data deur 'n klik op 'n data-identifiseerder gevind kan word (Müller-Spitzer en Koplenig 2014: 195). Gouws en Bothma (2024: 359-360) verwys in hierdie uitlegtipe na die benutting van vlakke en vakke om vertikale en horisontale soekposisies te skep. So 'n uitleg kan van groot waarde wees in die benutting van 'n dinamiese artikelstruktuur waar alle data nie gelyktydig gesien word nie en waar die gebruiker kan kies wanneer watter aanduiders vertoon moet word. Hier speel sowel 'n skag- as 'n tonnelstruktuur 'n wesenlike rol om die gepaste semantiese subkommentaar en die verlangde datatipe te vind.

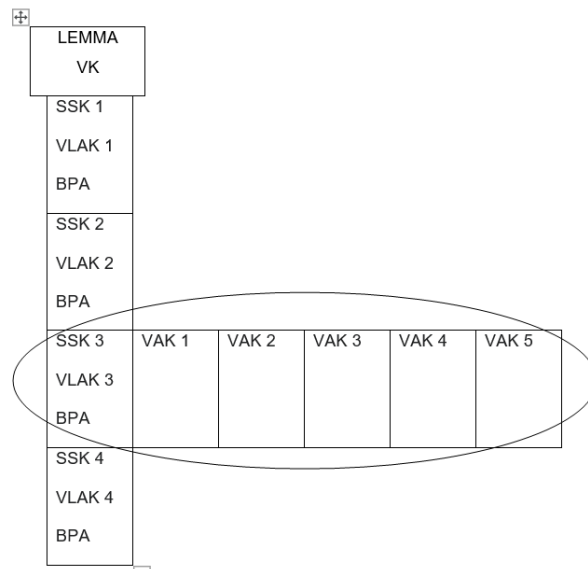
Gegee die potensiaal van 'n dinamiese artikelstruktuur en die voordele van 'n vertikale ordening van vlakke wat verskillende semantiese subkommentare huisves en die horisontale ordening van vakke wat verskillende tipes aanduiders bevat wat tot een semantiese subkommentaar behoort, kan leksikograwe aanpassings in die dataverspreidingstruktuur aanbring wat 'n veel uitvoeriger aanbod van semantiese data tot gevolg kan hê. Voorstelle in dié verband word in die hieropvolgende bespreking gemaak.

4.2 'n Veelfasige geïntegreerde mikrostruktuur

'n Basiswoordeboekartikel bestaan uit twee kommentare, naamlik 'n vormkom-

mentaar (VK) en 'n semantiese kommentaar (SK) (Wiegand en Smit 2013: 169 e.v.). Die ordening en plasing van aanduiders in die semantiese kommentaar word deur die tipe mikrostruktuur bepaal. Vir hierdie bespreking is slegs die geïntegreerde mikrostruktuur ter sake. Volgens Wiegand (1989a: 483) bevat 'n geïntegreerde mikrostruktuur alle artikel-interne aanduiders wat nie tot die vormkommentaar behoort nie, wat in die bestek van 'n bepaalde betekenis-aanduiders val en wat tot dieselfde semantiese subkommentaar as die betrokke betekenis-aanduiders behoort.

Dit is belangrik om daarop te let dat Wiegand hier nie net na koteksaanduiders verwys nie, maar wel na alle aanduiders in die bestek van die betekenis-aanduiders. Meerdere aanduidertipes kan dus tot 'n geïntegreerde mikrostruktuur behoort. 'n Horisontale ordening van vakke soos in figuur 11 se derde semantiese subkommentaar bied ruimte aan aanduiders van verskillende datatipes. Hierdie verskillende vakke is almal deel van die semantiese subkommentaar SSK 3 en val binne die bestek van hierdie semantiese subkommentaar se betekenisparafrase-aanduiders (BPA). Hulle tree nie op, soos in die geval van 'n niegeïntegreerde mikrostruktuur (figuur 2), in 'n teksblok los van die semantiese subkommentare nie. Gevolglik is hier wel sprake van 'n geïntegreerde mikrostruktuur — weliswaar 'n ander tipe as wat tot dusver bespreek is. Die aanduiders van die verskillende vakke word nie gesamentlik vertoon nie. Slegs 'n enkele vak se aanduiders word vertoon na 'n klik op die betrokke data-identifiseerder. Die inhoud van die verskillende vakke word daarom elk afsonderlik deel van die integraat waarin die betekenisparafrase-aanduiders optree. Binne hierdie semantiese subkommentaar geld 'n veelfasige geïntegreerde mikrostruktuur.



Figuur 11: 'n Veelfasige geïntegreerde mikrostruktuur

Die semantiese subkommentaar van 'n artikel met 'n veelfasige geïntegreerde mikrostruktuur kan, naas die vertikaalgeordende soekgleuf (dit is die gidsvak waarin die betrokke polisemiese waarde se betekenisparafrase aangedui word), 'n minimum van twee en 'n onbeperkte maksimum vakke bevat met data wat binne die bestek van die betekenisparafrase val. Die geïntegreerde aard van die mikrostruktuur word nie deur die gelyktydige integrasie van die inhoud van al die vakke gesamentlik bepaal nie, maar deur die integrasie, in verskillende fases, van die inhoud van 'n individuele vak en die gidsvak. 'n Klik op die data-identifiseerder van, byvoorbeeld, vak 1, skep 'n geïntegreerde mikrostruktuur, ook 'n uitgestelde geïntegreerde mikrostruktuur soos in die ANW (figuur 6), met vak 1 wat saam met die gidsvak hierdie geïntegreerde mikrostruktuur vorm. Telkens skep die klik op 'n data-identifiseerder 'n onmiddellike adressering tussen die data van daardie vak en die optrede van die leksikale item wat deur die lemma verteenwoordig word in die polisemiese waarde soos deur die betekenisparafrase aangetoon. Figuur 12 wys hoe so 'n integrasie van die gidsvak en 'n horisontaal-geordende vak 'n parallel-geïntegreerde mikrostruktuur vorm.

Lemma				
VK				
SSK 1				
Vlak 1				
BPA				
SSK 2				
Vlak 2				
BPA				
SSK 3	Vak 1	Vak 2	Vak 3	Vak 4
Vlak 3				
BPA				
SSK 4				
Vlak 4				
BPA				

Figuur 12: 'n Parallel-geïntegreerde mikrostruktuur

4.2.1 Besetting van die vakke

'n Vernuwende aspek van verklarende aanlyn woordeboeke is die dataverspreiding en -aanbod in die verskillende vakke van 'n artikelvlak. Elke vertikaalgeordende vlak wat die bewerking van 'n afsonderlike semantiese subkommentaar huisves, begin met 'n gidsvak wat gevolg word deur horisontaalgeordende vakke wat elk 'n parallel-geïntegreerde mikrostruktuur vertoon. Die gidsvak bevat tipieserwys 'n herhaling van die lemmateken met, indien nodig, een of meer leksikografiese etikette om bepaalde afwykings te merk van die optrede van die woord wat deur die lemma verteenwoordig word in die spesifieke poli-

semiese waarde. Daarnaas word die tersaaklike betekenisparafrase-aanduider ook verstrek. Daar is 'n onmiddellik lemmatiese adressering tussen die betekenisparafrase-aanduider en die lemma wat as gidselement van die gidsvak optree. So 'n lemmatiese adressering geld ook vanuit elke afsonderlike vak wat parallel aan die gidsvak aangebied word.

Die leksikograaf het die vryheid om te bepaal in watter volgorde die data in die verskillende vakke verstrek moet word. Dit moet nie 'n arbitrêre besluit wees nie, maar moet tematiese samehang asook die behoeftes van die aangewese teikengebruiker in ag neem. Deur elke vak van 'n gepaste data-identifiseerder te voorsien, vergelyk figuur 13, kry die gebruiker 'n oorsig oor die datatipes wat in die betrokke semantiese subkommentaar aangebied word. Let wel, slegs enkele datatipes word in hierdie figuur vertoon. Veel meer tipes kan probleemloos in 'n veelfasige geïntegreerde mikrostruktuur geakkommodeer word.

SSK 3	Vak 1	Vak 2	Vak 3	Vak 4	Vak 5
Lemma BPA	Betekenis +	Koteks 1 Voorbeeld sinne	Koteks 2 Kollokasies Voorbeeldsinne	Sinonieme 1. Absoluut Voorbeeldsinne 2. Gedeeltelik Voorbeeldsinne	Teenoorgesteldes Antonieme Omgekeerdes Komplementêre vorme

Figuur 13: Besetting van vakke

4.2.1.1 Betekenis +

Hier word die eerste vak beset deur data ter aanvulling van die betekenisparafrase. Soos vroeër in hierdie artikel genoem, is taal gewortel in die individu se ervaring en die individu se interpretasie van 'n woord se betekenis is gebaseer op hierdie ervarings. Van hierdie ervaringsdata val buite die eng linguistiese benadering van betekenisverklaring waarvan die hoofbetekenisparafrase blyke gee. In 'n vak wat onmiddellik op die gidsvak volg, kan 'n heenkome gebied word aan bykomende, selfs ook buitelinguistiese, data wat die leksikograaf as belangrik beskou ter ondersteuning en ter aanvulling van die data in die betekenisparafrase. Op hierdie manier word voorsiening gemaak, onder meer deur die benutting van supplementêre kenmerke, vir die bevrediging van 'n breër kognitiewe funksie van die betrokke woordeboek. 'n Parallel-geïntegreerde mikrostruktuur plaas hierdie betekenis +-data in dieselfde integraat as die data in die gidsvak.

4.2.1.2 Koteksleiding

4.2.1.2.1 Voorbeeldsinne

'n Tipiese kenmerk van 'n geïntegreerde mikrostruktuur is die plasing van koteksaanduiders in dieselfde integraat as die betekenisparafrase-aanduider. 'n Klik op die data-identifiseerder van vak 2 skep 'n parallel-geïntegreerde mikrostruktuur, as onderdeel van die veelfasige geïntegreerde mikrostruktuur, wat koteksaanduiders met die betekenisparafrase verbind. Die woordeboekgebruiker wat slegs 'n teksbegripbehoefte het se woordeboekraadpleging hoef nie verder as 'n kyk na die gidsvak te gaan nie. Waar 'n teksproduksiefunksie bevredig moet word, lei 'n klik op die koteksmerker die gebruiker na die ondersteunende voorbeeldsinne — wat poëme en/of sitate kan wees.

Woordeboeke verstrek voorbeeldsinne ter wille van teksproduksiebehoeftes, maar voorbeeldsinne ondersteun ook beter teksbegrip. Vir leksikograwe is dit nie altyd ewe maklik om te weet of 'n bepaalde optrede van 'n woord 'n eie betekenisonderskeiding verteenwoordig al dan 'n bepaalde gebruik van die woord in 'n bestaande betekenisonderskeiding nie. Behoort die optrede van die woord *kort* in 'n *kort referaat* en 'n *kort rok* tot verskillende betekenisonderskeidings of is dit verskillende gebruike van dieselfde betekenisonderskeiding, te wete "van beperkte afmeting"? Vergelyk ook Gouws (1987). In 'n vak wat voorbeeldsinne bevat, kan 'n groter verskeidenheid voorbeelde verstrek word ter illustrasie van die bepaalde betekenisonderskeiding, maar ook van verskillende gebruike van 'n woord in dieselfde betekenisonderskeiding. Dit verteenwoordig 'n uitbreiding van koteksaanduiders en bied toegevoegde waarde aan die aanlyn woordeboek.

4.2.1.2.2 Kollokasies

Ter wille van omvattender koteksleiding kan dit nuttig wees om tussen voorbeeldsinne en kollokasies te onderskei. Dit word soms ook in gedrukte woordeboeke gedoen, maar dikwels met beperkter aandag aan kollokasies. Daarom bied vak 3 verdere koteksleiding, te wete kollokasie-aanduiders. Die verstrekking van kollokasies kan belangrik wees vir teksproduksiedoeleindes. Kollokasies is nie leksikale items in eie reg nie. Desondanks kan woordeboeke hoër status aan hulle gee as net dié van koteksaanduiders. Naas die gee van kollokasies as koteksaanduiders gerig op die lemma kan verdere koteksaanduiders verskaf word waarin nie net die woord optree wat deur die lemma verteenwoordig word nie, maar die kollokasie as sodanig. Kollokasies word dan verhef tot sekondêre bewerkingsseenhede wat as adres van 'n eie koteksaanduider optree. So 'n koteksaanduider is deel van 'n tweeledige adresseringsverhouding, naamlik 'n lemmatiese adressering asook 'n nielemmatiese adressering met die kollokasie wat die adres is. Die tweede koteksvak in figuur 13 (vak 3) kan kollokasies bevat met die lemma as een van die komponente — hetsy as basis hetsy as kollokator. Naas hierdie kollokasies kan die vak ook voorbeeldmateriaal bevat as koteksaanduiders van die betrokke kollokasies.

'n Vak wat kollokasies bevat asook voorbeeldmateriaal met die kollokasie as adres verteenwoordig 'n verdere tipe geïntegreerde mikrostruktuur. Die onmiddellike adressering van koteksaanduiders en kollokasie skep 'n eie integraat en bied 'n vak-interne geïntegreerde mikrostruktuur. Die geïntegreerde mikrostruktuur wat in die semantiese subkommentaar geld, is die primêre geïntegreerde mikrostruktuur, terwyl die geïntegreerde mikrostruktuur wat slegs binne 'n vak geld as 'n sekondêre geïntegreerde mikrostruktuur beskou kan word.

4.2.1.3 Betekenisbetrekkinge

Vak 4 en 5 is gemerk vir bepaalde tipes betekenisbetrekkinge, te wete sinonimie en semantiese teenoorgesteldes. Die leksikon bevat 'n netwerk betekenisbetrekkinge en in verdere vakke sou daar ook aan ander tipes betekenisbetrekkinge aandag gegee kon word, byvoorbeeld hiponimie. Vergelyk ook die bespreking in 2.4.2.

4.2.1.3.1 Sinonieme

Aanduiders van sinonieme is woordeboekinskrywings waarby gebruikers baie kan baat. In woordeboeke word sinonieme op verskillende maniere en in verskillende artikelgleuwe verstrek, telkens as inskrywings met 'n lemmatiese adressering. Sinoniemaanduiders kry selde in woordeboeke 'n eie bewerking om die aard van die betrokke sinonimiese betrekking duidelik te maak. Woordeboeke onderskei nie konsekwent tussen absolute en gedeeltelike sinonimie nie en dit kan striemend wees vir die kommunikatiewe sukses van die woordeboekgebruiker.

'n Omvattender aanbod van semantiese data behoort ook tot 'n omvattender verslag van betekenisbetrekkinge te lei. In vak 4 sou daar dus tussen absolute en gedeeltelike sinonieme onderskei kan word en gepaardgaande voorbeeldmateriaal sou die korrekte gebruik kon illustreer. As 'n voorbeeld: in die bewerking van die woord *besope* behoort *dronk* nie sonder verdere inskrywings as sinoniem verstrek te word nie. *Besope* het 'n verhouding van gedeeltelike sinonimie met *dronk* en die aanduiders van *dronk* as sinoniem moet vergesels wees van 'n merker van die spesifieke polisemiese waarde van *dronk* wat in 'n sinonimiese betekenisbetrekking tot *besope* staan. Sonder hierdie aanvullende inskrywing sou die gebruik van *dronk* in die betekenisonderskeiding "lighoofdig" ten onregte ook as sinonimies tot *besope* gesien kon word. Naas 'n aanduiding van die tipe sinonimiese verhouding asook van spesifieke betekenisonderskeidings van woorde wat by so 'n verhouding betrokke is, kan voorbeeldmateriaal daartoe bydra om die sinonimiese verhouding te ekspliseer en die gebruiker van beter semantiese leiding te voorsien. Waar sinoniemaanduiders deur koteksaanduiders ondersteun word, vertoon die woordeboek nogmaals 'n vak-intern geïntegreerde mikrostruktuur.

4.2.1.3.2 Betekenisteenoorgesteldes

Binne die breë veld van semantiese opposisie kan verskillende tipes betrekkinge van teenoorgesteldheid onderskei word, byvoorbeeld antonimie, omgekeerdheid en komplementariteit. 'n Woordeboek wat as taalkundenaslaanbron dien, moet 'n presiese hantering van hierdie betrekkinge bied en moet dit vermy om 'n term soos *antoniem* oorkoepelend te gebruik vir die verskillende tipes teenoorgesteldheid.

Vak 5 kan data bevat om verskillende tipes betekenisteenoorgesteldheid voor te stel en gebruikers moet op 'n ondubbelsinnige manier tussen die tipes kan onderskei. 'n Benoeming van die onderskeie tipes betekenisteenoorgesteldes tesame met die gebruik van relevante koteksaanduiders sal ook 'n waardevolle rol kan speel om gebruikers vertrouwd te maak met die onderskeie gebruike van die verskillende woordpare. Ook hier word 'n vak-interne geïntegreerde mikrostruktuur weer eens benut.

4.2.1.4 Ander data in die semantiese subkommentaar

Ruimtebesparing is geen groot kwessie in aanlyn woordeboeke nie. Daar is voldoende bergings- en aanbiedingsruimte om afstand te doen van ruimtebesparingsmetodes soos teksverdigting asook makrostrukturele nesting en nisting. Die vormkommentaar van 'n woordeboekartikel kan 'n gleuf bevat waarin verskillende tipes ongelede woorde verstrek word ter illustrasie van woordboumoontlikhede met die woord wat deur die lemma verteenwoordig word as boumiddel. Sulke gelede woorde word om morfologiese redes opgeneem en nie om as onverklaarde lemmas implisiete betekenisleiding te gee nie. As deel van die semantiese opdrag van 'n aanlyn woordeboek kan nisting en nesting uitgeskakel word ten gunste van 'n opname van daardie betrokke gelede leksikale items wat as lemmas kwalifiseer, as gidselemente van artikels wat 'n volledige bewerking bied. Hier is dan 'n eksplisiete oordrag van die gelede woorde se betekeniswaarde en nie, soos in die geval van onverklaarde lemmas, slegs 'n implisiete aanduiding nie. In aanlyn woordeboeke behoort eksplisiete betekenisoordrag prioriteitswaarde te hê.

Die aanbod van semantiese data in verklarende woordeboeke kan nog veel verder uitgebrei word en die vernuwende benutting van leksikografiese strukture dra daartoe by om 'n geskikte huisvesting van data te bewerkstellig. Woordeboeke is nie meer net draers van tekstipes nie, vergelyk Wiegand (1996), maar hulle bevat ook nietekstuele data soos klank- en beeldmateriaal. Sulke data kan ook as deel van die semantiese kommentaar aangebied word. In die horisontale ordening van data kan 'n semantiese subkommentaar 'n vak bevat met illustrasieprente wat in 'n verhouding van ostensiewe adressering (Gouws 1994) tot die lemma staan.

Aanlyn woordeboeke het verskillende soekposisies, te wete:

- die *soeksone*, dit is 'n gleuf in 'n woordeboekartikel waar 'n bepaalde aanduiding of datatipe geplaas word;

-
- die *soekgebied*, dit is die individuele woordeboekartikel;
 - die *soekveld*, dit is die woordeboek se sentrale teks wat die verskillende artikeltrajekte bevat;
 - die *soekstreek*, dit is die volle woordeboek met sy sentrale teks plus die verskillende buitetekste;
 - die *soekdomein*, dit is die portaal waarbinne aanlyn woordeboeke geplaas kan word; en
 - die *soekuniversum*, dit is die omgewing buite die soekdomein, byvoorbeeld die internet of ander woordeboekportaal-eksterne bronne waartoe gebruikers via die woordeboek toegang het (Gouws 2023: 383, e.v.)

'n Optimale benutting van soekposisies kan ook die uitbreiding van die aanbod van semantiese data bevoordeel. Hier is dit veral die soekdomein en die soekuniversum wat nuwe moontlikhede skep.

4.3 Semantiese data vanuit die soekdomein en die soekuniversum

Die soeksone, soekgebied, soekveld en soekstreek is tradisionele soekposisies binne 'n woordeboek en terreine waar semantiese data geplaas kan word — soos gesien in die besetting van vlakke en vakke in woordeboekartikels, dit is in 'n woordeboek se soekgebiede. Die soekdomein en soekuniversum is soekposisies buite die woordeboek wat as vertrekpunt van 'n woordeboekraadplegingsprosedure dien, maar wat 'n verrykingswaarde aan 'n woordeboek bied wat die aard en omvang van onttrekbare inligting betref.

Daar is reeds vroeër in hierdie artikel verwys na Bergenholtz en Bergenholtz (2017) se voorstelle om meerdere woordeboeke aan een databasis te onttrek. Met behulp van slim sagteware kan leksikografiese data ook aan sowel woordeboeke as ander bronne onttrek word. Die leksikograwe van 'n verklarende aanlyn woordeboek wat in 'n woordeboekportaal optree, moet bewus wees van die tipe inhoud van die ander woordeboeke in die portaal. 'n Woordeboekportaal-interne kruisverwysingstelsel kan dit dan moontlik maak om vanuit 'n artikel in een woordeboek na 'n verlangde datatipe in 'n ander woordeboek in dieselfde soekdomein te beweeg. 'n Tipe semantiese data wat nie in 'n gegewe woordeboek aangebied word nie kan vanuit daardie woordeboek in 'n ander woordeboek van die portaal gevind word. Wisselwerking binne 'n soekdomein kan ook tot 'n omvattender beskikbaarstelling van semantiese data lei.

Die soekuniversum maak dit moontlik om toegang te kry tot data wat buite die grense van 'n woordeboek en woordeboekportaal val. Deur die gebruik van 'n leksikografiese datatrekkingstruktuur, vergelyk Gouws (2018), is die onttrekking van inligting aan data uit leksikografiese en ander bronne moontlik. Die inligting wat onttrek word aan die data in 'n koteksvak van 'n woordeboekartikel kan aangevul word met data vanuit die woordeboek se korpus. Hier moet die gebruiker daarop bedag wees dat die korpusdata nie geredigeer is nie. Bykomende korpusdata kan bydra tot hoër vlakke van teksbegrip en tot 'n beter interpretasie van die gebruik van 'n bepaalde woord. Dit bied ook meer insig in die

breër konteks van 'n sitaat. So 'n datatrekking brei die omvang van die betekenis- en gebruiksleiding uit wat 'n woordeboek bied en verhoog die status van die woordeboek as 'n gesaghebbende en verteenwoordigende bron van taalkundige kennis.

'n Datatrekkingstruktuur skep ook die moontlikheid om vanuit 'n woordeboekartikel inligting aan internetbronne te onttrek. Veral ensiklopediese inligting ter bevrediging van 'n woordeboek se kognitiewe funksie kan op hierdie manier verkry word. Hierdie soort datatrekking is reeds 'n keuse in verskeie woordeboeke. Vergelyk in hierdie verband figuur 14, 'n skermkoot van 'n deelartikel van die lemma *auto* in *dict.cc*. 'n Klik op die inligtingsmerker naas 'n bepaalde woord bied 'n opskietspyskaart van waar gebruikers deur 'n verdere klik na onder meer Google gelei kan word om toegang te kry tot aanvullende semantiese en ander data:

English » Restrict search to this language	German » Restrict search to this language
edit NOUN an auto autos SYNO auto automobile car machine ...	edit NOUN das Auto die Autos SYNO auto... selbst... [altes] Auto ...
auto-aggressive {adj} psych.	autoaggressiv
auto-centered {adj} {past-p} [Am.]	autozentriert
auto-centred {adj} {past-p} [Br.]	autozentriert
auto-discriminative {adj} comp.	selbsterkennend

Nouns	
car	automot. Auto {n} 5824
automobile [esp. Am.]	automot. Auto {n} 95
auto [Am.] [coll.] [automobile]	automot. Auto {n} 51
	automot. Auto {n} 35
	automot. Personenwagen {m} 29
	automot. Auto {n} 17
	Personenkraftwagen {m} <PKW, Pkw> 13
	Anrufbeantwortungsmodus {m}
	Auspuffgase {pl}
	Auspuffgase {pl}
	Autoaggression {f}
	Autoaufbereitung {f}
	Autobiografie {f}
	Autobranche {f}
	Autodafé {n}
	Autodiebstahl {m}
	Autoerotik {f}
	Autofähre {f}
	Autohändler {m}
	Autohändlerin {f}
	Autohersteller {m}
	Autohersteller {pl}
	Autoimmunerkrankung {f}
	Autoimmunkrankheit {f}

Figuur 14: Uit *dict.cc*

5. Ten slotte

Die aanbieding van semantiese data is een van die hoofopdragte van die opstellers van verklarende woordeboeke. Die moontlikhede wat die aanlyn omgewing bied, kan daartoe lei dat woordeboeke 'n veel omvattender en uitvoeriger aanbod van semantiese data vertoon. Leksikograwe moet op 'n kundige en dikwels vernuwende manier met woordeboekstrukture omgaan. Daar moet weg-beweeg word van 'n dataverspreiding en -plasing wat alle data gelyktydig sigbaar maak vir gebruikers. Deur data in verskillende vlakke en vakke aan te bied en volgens 'n veelfasige geïntegreerde mikrostruktuur te orden, kan gebruikers slegs daardie semantiese inligting onttrek waaraan hulle 'n behoefte het — sonder om deur al die soeksones in 'n omvangryke woordeboekartikel te worstel. Woordeboek-interne data kan aangevul word deur aan gebruikers toegang tot data in die soekuniversum te gee. Sulke prosedures verhoog die waarde van 'n woordeboek as 'n omvattende bron van taalkundige kennis.

Om woordeboeke relevant te hou vir die teikengebruikers moet leksikograwe pogings aanwend om meer en meer gerigte data in hulle woordeboeke op te neem, maar ook om die woordeboek as vertrekpunt vir soektogte in ander bronne te laat dien. Hiervoor skep die dekking van semantiese data 'n ideale geleentheid.

6. Bronnelys

6.1 Woordeboeke

ANW = *Algemeen Nederlands Woordenboek*. <http://anw.inl.nl/> [April 2025]

De Villiers, M. et al. (Reds.). 1988⁷. *Nasionale woordeboek*. Kaapstad: Nasou.

dict.cc = <https://www.dict.cc/> [Mei 2025].

elexiko = <http://www.owid.de/wb/elexiko/start.html> [April 2025]

HAT = Luther, J. et al. (Reds.). 2015⁶. *HAT: Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson.

Louw, J.P. en E. Nida. 1988. *Greek-English Lexicon of the New Testament Based on Semantic Domains*. New York: United Bible Societies.

Martin, W. et al. (Reds.). 2011. *Prisma Groot woordenboek Afrikaans en Nederlands*. Houten: Prisma.

WAT = *Woordeboek van die Afrikaanse Taal*.

6.2 Ander bronne

Bergenholtz, I. en H. Bergenholtz. 2017. One Database, Four Monofunctional Dictionaries. *Hermes* 50: 119-125.

Bergenholtz, H., T.J.D. Bothma en R.H. Gouws. 2011. A Model for Integrated Dictionaries of Fixed Expressions. Kosen, I. en K. Kosen (Reds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users: Proceedings of eLex 2011, Bled, Slovenia, 10–12 November 2011*: 34-42. Ljubljana: Trojina, Institute for Applied Slovene Studies.

Combrink, J.G.H. 1979. Die sesde deel van die WAT. *Standpunte* 140 (32–2): 49-64.

- DiMuccio-Failla, P.V.** 2025. A Theory for a Usage-Based Cognitive Lexicography. Giacomini, L. en V. Piuino (Reds.). 2025. *Patterns of Meaning in Lexicography and Lexicology*: 19-90. Berlyn/Boston: De Gruyter.
- Fodor, J.D.** 1977. *Semantics: Theories of Meaning in Generative Grammar*. New York: Crowell.
- Gouws, R.H.** 1987. Lexical Meaning versus Contextual Evidence in Dictionary Articles. *Dictionaries* 9: 87-96.
- Gouws, R.H.** 1989. *Leksikografie*. Kaapstad: Academica.
- Gouws, R.H.** 1994. Ostensiewe adressering in vertalende woordeboeke. *Lexikos* 4: 61-85.
- Gouws, R.H.** 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. *Lexikos* 13: 92-110.
- Gouws, R.H.** 2006. Die zweisprachige Lexikographie Afrikaans-Deutsch — Eine metalexikographische Herausforderung. Dimova, A., V. Jesenšek en P. Petkov (Reds.). 2006. *Zweisprachige Lexikographie und Deutsch als Fremdsprache*: 49-58. Hildesheim: Georg Olms.
- Gouws, R.H.** 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.
- Gouws, R.H.** 2023. Nuwe raamstruktuurtypes in aanlyn woordeboeke. *Lexikos* 33(1): 382-403.
- Gouws, R.H.** (om te verskyn). Definitional Methods in Lexical Semantics.
- Gouws, R.H. en T.J.D. Bothma.** 2024. Aspekte van inligtingsonttrekkingstrukture in aanlyn woordeboeke. *Lexikos* 34(1): 355-370.
- Gove, P.B.** 1966. Self-Explanatory Words. *American Speech* 41: 182-198.
- Hausmann, F.J. et al. (Reds.).** 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*. Berlyn: De Gruyter.
- Klotz, M. en T. Herbst.** 2016. *English Dictionaries. A Linguistic Introduction*. Berlyn: Erich Schmidt.
- Levisen, C. en S. Waters (Reds.).** 2017. *Cultural Keywords in Discourse*. Amsterdam: John Benjamins.
- Louw, J.P.** 1985. A Semantic Domain Approach to Lexicography. Louw, J.P. (Red.). 1985. *Lexicography and Translation*: 157-197. Kaapstad: Bybelgenootskap van Suid-Afrika.
- Lyons, J.** 1977. *Semantics*. Cambridge: Cambridge University Press.
- Müller-Spitzer, C. en A. Koplenig.** 2014. Questions of Design. Müller-Spitzer, C. (Red.). 2016. *Using Online Dictionaries*: 189-204. Berlyn/Boston: De Gruyter.
- Odendal, F.F.** 1979. Plus positief en plus negatief. *Tydskrif vir Geesteswetenskappe* 19(1): 24-41.
- Scerba, L.** 1940. Opyt obshchei teorii leksikografii. *Izvestiia Akademii Nauk SSSR, Otdelenie literatury i iazyka* 3: 89-117. (Engelse vertaling deur D.M.T.Cr. Farina 1995: Towards a General Theory of Lexicography. *International Journal of Lexicography* 8(4): 314-350).
- Tarp, S.** 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, H., S. Nielsen en S. Tarp. (Reds.). 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 17-32. Bern: Peter Lang.
- Wiegand, H.E.** 1989a. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Reds.). 1989-1991: 462-501.
- Wiegand, H.E.** 1989b. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. (Reds.) 1989-1991: 371-409.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts. An Overview. *SA Journal of Linguistics* 14(4): 134-149.
- Wiegand, H.E. en M. Smit.** 2013. Microstructures in Printed Dictionaries. Gouws, R.H. et al. (Reds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 149-214. Berlyn/New York: De Gruyter.

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2108> (Volle uitgawe / Full volume)
<https://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2075> (Article)

74 Rufus H. Gouws

Wierzbicka, A. 1972. *Semantic Primitives*. Frankfurt: Athenäum.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Zgusta, L. 1971. *Manual of Lexicography*. Den Haag: Mouton.

Lexicographic Description of a Polysemous Word in a Learner's Dictionary Based on Its Lexical Prototype

Anton Karasev, *Département d'Anglais, Laboratoire Ligérien
de Linguistique (VETD-DO), Université de Tours, Tours, France*
(anton.karasev@univ-tours.fr)
(<https://orcid.org/0009-0005-7311-6922>)

Abstract: This paper presents a practical application of the lexical prototype theory to lexicographic practice, proposing a draft model of a learner's dictionary entry for a polysemous word based on its lexical prototype. The lexical prototype (LP) is regarded as the semantic core of a polysemous word that functions at the level of the language system and is represented in actual speech by contextual meanings: the primary meaning (PM), metonymic, and metaphorical senses. Drafting a dictionary entry based on the LP requires reconstructing the LP through analysing the contextual meanings in relation to the PM in order to establish the minimal set of semantic features that are shared by all the senses of the word. The proposed method of lexicographic description of polysemous words offers a new perspective on polysemy as a linguistic phenomenon as well as moves away from the list-based method.

Keywords: LEXICOGRAPHY, LEARNER'S DICTIONARY, DICTIONARY ENTRY, POLYSEMY, LEXICAL PROTOTYPE, PRIMARY MEANING, METONYMY, METAPHOR, COGNITIVE LINGUISTICS

Opsomming: Die leksikografiese beskrywing van 'n polisemiese woord in 'n aanleerderswoordeboek gebaseer op die leksikale prototipe daarvan. In hierdie artikel word 'n praktiese toepassing van die leksikale prototipe-teorie in die leksikografiese praktyk aangebied en 'n konsepmodel van 'n aanleerderswoordeboekinskrywing vir 'n polisemiese woord gebaseer op die leksikale prototipe daarvan, word voorgestel. Die leksikale prototipe (LP) word beskou as die semantiese kern van 'n polisemiese woord wat op die vlak van die taalsisteem funksioneer en word in werklike spraak deur kontekstuele betekenisse verteenwoordig: die primêre betekenis (PB), metonimiese, en metaforiese betekenisse. Die opstel van 'n woordeboekinskrywing wat op die LP gebaseer is, verg die rekonstruksie van die LP deur die analisering van die kontekstuele betekenisse in verhouding tot die PM om sodoende die minimale stel semantiese kenmerke wat deur al die betekenisse van die woord gedeel word, te bepaal. Die voorgestelde metode vir die leksikografiese beskrywing van polisemiese woorde bied 'n nuwe perspektief op polisemie as 'n linguïstiese verskynsel en beweeg terselfdertyd weg van die lysgebaseerde metode.

Sleutelwoorde: LEKSIKOGRAFIE, AANLEERDERSWOORDEBOEK, WOORDEBOEKINSKRY-
Lexikos 35(2) (AFRILEX-reeks / series 35A: 2025): 75-93

WING, POLISEMIE, LEKSIKALE PROTOTYPE, PRIMÊRE BETEKENIS, METONIMIE, METAFOOR, KOGNITIEWE LINGUISTIEK

Introduction

When making a dictionary, a lexicographer faces a difficult task: on the one hand, the work must be theoretically grounded, on the other hand, it is supposed to satisfy the needs of the target audience, which determines the popularity of the dictionary and keeps it in demand (Slonimskaya 1990: 49). As a rule, dictionary users are reluctant to choose what they are not accustomed to. As a result, the structure and content of a lexicographic work are shaped as much by the expectations of the target audience as by linguistic factors. Therefore, dictionary-making is most often built on the achievements of previous generations of lexicographers both in compiling the lexicon and in writing dictionary entries (Jackson 2002: 27). Thus, the current consensus is that polysemy must be presented in the dictionary in a linguistically justified manner (Lodwig and Barret 1967: 53-54). Despite significant progress in the field of cognitive semantics, the methods of lexicographic description of lexical meanings have not changed much over time (Karasev 2014a; Halas 2016: 125). Nearly a century after the first learner's dictionaries came into being — which once stood at the forefront of both practical and theoretical lexicography (Cowie 2007: 14; Jackson 2002: 129) — no significant progress has been made in this field. For example, presenting polysemy as a list of numbered definitions introduced by Palmer has remained largely unchanged since the 1930s (Cowie 2007: 44). This is how practising lexicographers contribute to the development of linguistics: the methods they use to represent the semantic structure of words give rise to discussions whether dictionary definitions correspond to the actual lexical meaning (Arnold 2010: 108). There have been attempts to create lexicographic descriptions based on prototype theory (Halas 2016), and to describe meanings in a way that reflects their mental representations (Rivelis 2009). Opponents of such attempts argue that since we cannot know exactly how meanings are stored in the mind, it is impossible to describe them in the same way a dictionary. If this were to be discovered, the verbal description of meanings would differ significantly from how those meanings are represented in the mind of a native speaker (Zaliznyak 2004). Nevertheless, such an attempt still seems justified and, especially in the field of learner's lexicography, as it may be useful to reconstruct the systemic meaning before examining actual meanings (Guillaume 1973: 143).

This paper proposes a theoretically sound model for a learner's dictionary entry of a polysemous word based on lexical prototype theory (Arkhipov 2003: 79; Levchina 2003: 269; Pesina 1998, 2005a, 2005b; Shiryayeva 2008), developed in the works of Arkhipov and his followers. According to this theory, a word's meaning within the language system reflects a generalisation of its various uses in everyday speech. Grasping this generalised meaning can be especially helpful for

foreign language learners, as it allows them to recognise connections between specific contextual meanings, or senses, that might otherwise be unclear.

Polysemy in learner's dictionaries

The value of a dictionary lies in its ability to present the linguistic information in a visible, complete diverse manner. It serves as an essential tool of understanding language phenomena as it influences both their perception and, to some extent, linguistic consciousness. One might as well speak of a special "dictionary view" of reality (Devkin 2000: 10, 14-15; Atkins and Rundell 2008: 2). At the same time, a dictionary — especially a learner's dictionary — serves as an authority for the majority of its users. Therefore, it is essential to develop lexicographic principles that ensure descriptions of linguistic phenomena are as accurate and faithful to linguistic reality as possible. In this regard, an important issue is the correspondence between a word's meaning and its dictionary definition. Specifically, it is necessary to understand whether the definition reflects the actual lexical meaning, the reference underlying that meaning, or whether it merely provides a verbal description of a real-world object (Ufimtseva 1984: 135). Another key question concerns the extent to which the lexicographic treatment of polysemous words in dictionaries aligns with linguistic consciousness.

In theoretical literature, as well as in lexicography textbooks, the issue of polysemy and its lexicographic representation is usually described as resolved (Apresyan 1984: 409; Jackson 2002: 88; Dubchinskiy 1994: 10, 24). In these works, the authors outline the existing ways of presenting word meanings in dictionaries. Other researchers describe it as problematic (Hanks 2000: 211–212), though they don't propose specific solutions. Interestingly, the idea of polysemy as existence of several independent, though related meanings of a word has been shaped by dictionary making practice (Apresyan 1984: 186; Katsnelson 2011: 40; Mednikova 2010: 17). This leads to a vicious circle: a practical lexicographer finds theoretical justification for the list-based method of describing polysemy in the linguistic literature, while a theoretical linguist uses dictionary data as a basis for describing the phenomenon of lexical polysemy. It is also interesting to note that English-language lexicography does not seek to establish or highlight the closeness of separate senses of a word within dictionary entries (Litvin 1984: 5-6); instead, more attention is paid to reflecting the relation between lexical units (Soldatov 1997: 120). This leads to the fact that the meanings of polysemous words in monolingual learner's dictionaries of English are often presented as lengthy lists of definitions.

Corpus studies confirm that the meanings of a polysemous word do not exist as 'checklists': "The numbered lists of definitions found in dictionaries have helped to create a false picture of what really happens when language is used." (Hanks 2000: 205). Since human memory doesn't have the capacity to store every variation of the meanings of all known words (Brudny 1971: 19), the information about language units is retained in an individual's memory not as extended

dictionary definitions, but in a more compact form, together with the mechanisms that enable the actualisation of meanings in context (Pertsov 2001: 43; 22). Thus, a dictionary presents a purely theoretical scheme which reflects the lexicographer's interpretation of the totality of actualisations of language-system units in speech, created through the "cataloguing" of meanings attested in texts (Arkhipov 2004: 79). This scheme does not represent the actual meaning of the word itself, but rather its "meaning potentials" that is, "potential contributions to the meanings of texts and conversations in which the words are used, and activated by the speaker who uses them" (Hanks 2000: 211). In this sense, the dictionary constitutes a minimised model of the speech usage of language units and does not provide a reliable representation of a word's semantics (Chudinov 2003: 34).

This means that the word meaning at the level of the language system must differ significantly from its realisation in speech. Given that human memory is not "a warehouse of individual traces duplicating the events of the external world, but a creative process of constructing these events" (Zalevskaya 2000: 53), the meaning at the language-system level, regardless of the number of specific meanings recorded in the dictionary, must be sufficiently compact and abstract. Each contextual meaning correlates with the systemic meaning in the same way that an allophone or allomorph relates to a phoneme or morpheme, respectively (Pesina 2005c: 76; Jakobson 1985: 177). The idea of "semantic compactness," as well as the distinction between invariants at the level of language and variants at the level of speech — originally distinguished in phonology and later in morphology — has led to the emergence of hypotheses about the existence of a core meaning, or an invariant meaning of a word at the level of the language system (Arkhipov 2003; Budagov 1947; Shmelev 1973; Smirnitsky 1954: 23; Vinogradov 1947: 34).

This systemic meaning of a word is contrasted with its contextual meanings, or senses. The systemic meaning can be established out of context, while contextual, or derived, meanings are determined only through actual use (Kuryłowicz 1955: 73-81). The systemic meaning contains a certain feature that links various senses of a word and serves as the basis for new derived meanings (Zvegintsev 1957: 205-207). Despite the existence of such studies, their results, with rare exceptions (Rivelis 2009), are not reflected in lexicography.

Polysemy and lexical prototype

Cognitive linguistics, and in particular prototype theory, offers a perspective on polysemy that is more consistent with linguistic consciousness: it views polysemy as the presence of several interrelated meanings within a word, characterized by gradedness, fuzziness, flexibility, and clustering of senses (Geeraerts 2006: 141; Nikitin 2008: 71), all organized around a prototype — the best representative of the semantic structure (Rosch 1978: 54; Lakoff 1984). Among those who support the existence of a general, systemic meaning for polysemous words, prototype

theory has led to the hypothesis that such a best representative may be so effective that it can serve as the invariant meaning (Pesina 2005c: 65). I. Arkhipov proposed calling this semantic core the lexical prototype (LP) (Arkhipov 1998: 15). The LP represents a minimal set of semantic features shared by all the senses of the word (Leshcheva 1985: 5). It includes both the essential features associated with the form of the linguistic sign and the results of semantic derivation (Pesina 2005a: 19). Therefore, the LP serves, first, as a sufficient condition for interpreting a word's actual meaning (Pesina 1998: 6; Levchina 2003: 8), and second, as the foundation for the formation of all the meanings of a polysemous word (Pesina 2005c: 79).

Research on polysemy is closely associated with the study of lexical-semantic variation, according to which a same lexeme is represented in different contexts as its lexico-semantic variants (LSVs) (Ufimtseva 1962: 91; Ufimtseva 2010: 31; Yakovlyuk 2009: 156-157). In this framework, the LP operates as the invariant unit at the language-system level (systemic meaning), while each LSV represents its realisation at the level of speech (actual meaning) reflecting the natural mechanisms of semiosis (Karasev 2014c).

The concept of LP has practical applications in lexicography. For example, Arkhipov, Novikov and Pesina (2006) suggest placing the LP at the beginning of a dictionary entry as a generalised systemic meaning of a polysemous word that serves as a motivational basis for all the senses listed below. This dictionary entry is, however, list-based and the LP serves only a supplementary function. Independently of research on the LP, a similar model of a dictionary entry was proposed in (Halas 2016), where a lexicographic description of a polysemous word is based on the radial set model with reference to the prototypical meaning. This entry constitutes a complex hierarchical structure that demonstrates the semantic relationships between the word's senses. Nevertheless, it still does not depart from the traditional list-based method.

As noted above, the LP represents the minimal set of the most essential features sufficient for the adequate interpretation of the contextual meaning (Arnold 2010: 8), and is likely to be the systemic meaning of the word that enables its functioning in speech (Arutyunova 1999: 269).

Within the framework of LP theory, a polysemous word has three types of meanings: the primary meaning (PM), metonymic meanings, and metaphorical meanings. The PM also corresponds to the prototypical meaning in the theory of Brugman and Lakoff (Brugman and Lakoff 1988). The semes underlying metonymic and metaphorical transfers are part of the LP. Since, in real communicative situations, actual meanings are derived based on context, a dictionary entry structured around the LP includes the LP itself as well as examples of the headword in various contexts, accompanied by prompts — similar to signposts in LDOCE — that enable the dictionary user on their own to derive the meaning of each LSV with reference to the LP.

The systemic meaning exists as a part of the cognitive structure of a language speaker, and therefore is hidden from direct access. However, it can be estab-

lished based on the actual meanings of a word that are catalogued in existing dictionaries (Alefirenko 2005: 110). The LP can be described as the minimal semantic core of the word meaning, which functions at the level of the language system and which can serve as a basis for the construction of a consistent semantic structure of a word (Pesina 1998: 8).

Dictionary definitions provide detailed descriptions of contextual meanings, or LSVs, which offer valuable insight into how language speakers understand meanings, they are especially useful for establishing the LP.

The procedure for establishing the LP involves a componential analysis of dictionary definitions and an interpretation of contextual meanings to identify the semantic components that constitute the features common to all senses. This type of analysis, applied to nouns, verbs, and adjectives, is set out in detail in the dissertation (Karasev 2014a) as well as in related papers (Karasev 2013, 2014b), with no apparent variation across parts of speech. This paper proposes a draft model of a learner's dictionary entry for the verb *see* based on its LP.

Procedure for establishing the LP

To compile a dictionary entry for a polysemous word based on its LP, the first step is to formulate the LP of this verb. There are two stages of the procedure:

1. Identification of the PM of the word through the analysis of definitions from several dictionaries;
2. Interpretation of the word's LSVs based on the PM, in order to examine how the word's senses relate to PM and determine the minimal set of semantic features that constitute the LP.

Below are the definitions of the primary meaning of the verb *see* from seven general-purpose and two learner's dictionaries.

General-purpose dictionaries

<i>Cambridge Dictionary</i>	to be conscious of what is around you by using your eyes
<i>Oxford English Dictionary</i>	to perceive with the eyes
<i>The Chambers Dictionary</i>	to perceive by the sense operated in the eyes
<i>American Heritage Dictionary</i>	to perceive with the eye
<i>Merriam-Webster Dictionary</i>	to perceive by the eye
<i>Webster's Online Dictionary</i>	perceive by sight or have the power to perceive by sight

Learner's dictionaries

<i>Collins COBUILD</i>	when you see something, you notice it using your eyes.
<i>Longman Dictionary of Contemporary English</i>	to notice or examine someone or something, using your eyes
<i>Oxford Advanced Learner's Dictionary</i>	to become aware of somebody / something by using your eyes

As shown in the tables above, these definitions consistently feature, in various forms, the same semantic component [EYE PERCEPTION]: *perceive with the eye(s)/by sight/by the sense operated in the eyes*. Thus, the PM of the verb *see* can be formulated as: *perceive with the eyes*. The determinant may be adapted to the context if necessary. All the sentences containing the LSVs of the verb *see* below were taken from the British National Corpus. The selection of contexts roughly reflects the meanings of the verb *see* traditionally included in dictionaries.

To further identify the LP, the verb's LSVs are analysed against its PM. For this purpose, the contexts in which the LSVs of *see* are used were examined in detail.

LSV in context	PM	Interpretation
(1) If the operation is successful, he will be able to <i>see</i> again	<i>perceive with the eyes</i>	If the operation is successful, he will be able to <i>perceive with the eyes</i> again
(2) Can you <i>see the</i> bird in that tree?	<i>perceive with the eyes</i>	Can you <i>perceive</i> the bird in that tree <i>with your eyes</i> ?

In contexts (1) and (2), the verb *see* is used in its PM. In (1), the verb is intransitive and means *have the power to perceive by sight* (MWD), whereas in (2) it is transitive and means *have the power to perceive by sight*. Since seeing objects is possible only if one has *the power of sight* and to *have the power to perceive by sight* implies the ability to *see* objects, this distinction is irrelevant for the purposes of this study. Therefore, the same meaning is realised in both cases and it corresponds to the PM: *see*₁ — *perceive with the eyes*.

(3) Did you <i>see</i> that documentary on Channel 4 last night?	<i>perceive with the eyes</i>	Did you <i>perceive</i> that documentary on Channel 4 <i>with your eyes</i> last night?
--	-------------------------------	---

Context (3) suggests a purposeful action directed at *the documentary* as its object: *obtain the content of the documentary by perceiving it with the eye*. This action is based on the physical ability to see, but it is not limited to it and involves the

interpretation of what has been seen. The sense of the verb *see* in the analysed context is metonymic, because the interplay of concepts occurs within the same domain: the EYE PERCEPTION. This is the case of a metonymic transfer: sensory perception — obtaining sensory information as the goal of perception (Ageeva 1990: 10). This contextual meaning of *see* can be phrased as *see*₂ — *watch something (film, television, etc.) by perceiving with it the eyes*.

The same mechanism works in contexts (4) and (5):

- | | | |
|---------------------------------------|-------------------------------|---|
| (4) Please <i>see</i> who's knocking. | <i>perceive with the eyes</i> | Please <i>perceive</i> who's knocking <i>with your eyes</i> . |
| (5) <i>See</i> chapter 12. | <i>perceive with the eyes</i> | <i>Perceive</i> chapter 12 <i>with your eyes</i> . |

Unlike *see*₂, the focus in (4) and (5) is on the processing of information: looking at something → perceiving it with the eyes → receiving the information. The same meaning is found in context (6).

- | | | |
|---|-------------------------------|--|
| (6) I must <i>see</i> your passport before you can enter the country. | <i>perceive with the eyes</i> | I must <i>perceive</i> your passport <i>with my eyes</i> before you can enter the country. |
|---|-------------------------------|--|

This context doesn't imply a requirement to physically observe the passport, but rather to examine its content, that is to look at it in order to obtain information. Since the verb *see* refers to a purposeful action based on visual perception, this is a metonymic sense that can be presented as follows: *see*₃ — *look up, refer to, or find something out by perceiving it with the eyes*.

- | | | |
|---|-------------------------------|--|
| (7) We still <i>see</i> each other a couple of times a month. | <i>perceive with the eyes</i> | We still <i>perceive with</i> each other <i>with our eyes</i> a couple of times a month. |
|---|-------------------------------|--|

In (7), the verb *see* is also used in a metonymic sense, as it does not describe eye perception itself, but rather a certain kind of interaction that requires visual perception. This transfer is based on the idea of meeting someone one knows socially or by chance while *perceiving them with the eyes*, therefore: *see*₄ — *visit or meet someone one knows socially or by chance while perceiving them with the eyes*.

In a similar sense, the verb *see* is used in contexts (8) and (9).

- | | | |
|--|-------------------------------|--|
| (8) Can I <i>see</i> you home? | <i>perceive with the eyes</i> | Can I <i>accompany</i> you and <i>perceive with my eyes</i> that you are at home? |
| (9) My mother used to <i>see</i> me across the road. | <i>perceive with the eyes</i> | My mother used to <i>go with</i> me across the road <i>while perceiving me with her eyes</i> |

What makes the verb *see* special is that visual perception is the primary means of obtaining information about the world (Arutyunova 1999: 11; Dretske 1969: 77). That is why it can be used to denote a range of actions that somehow involve visual perception. In contexts (8) and (9), the verb *see* refers to the idea of accompanying someone somewhere and making sure that they are there *by perceiving it with your eyes*. In real communication, speakers tend to choose the most economical way of expressing this complex idea using a single word *see*. The combinability of the verb — *see someone somewhere* — hints at this concept. Thus, the verb *see* is used metonymically in a sense directly related to visual perception: *see₅* — *accompany somebody to make sure they get there by perceiving it with the eyes*.

- | | | |
|--|-------------------------------|--|
| (10) Please <i>see</i> that the lights are switched off before you leave. | <i>perceive with the eyes</i> | Please <i>make sure that</i> the lights are switched off before you leave [AS IF] <i>by perceiving it with your eyes</i> . |
| (11) Could you <i>see</i> that the kids are in bed by 8 o'clock? | <i>perceive with the eyes</i> | Could you make sure that the kids are in bed by 8 o'clock [AS IF] <i>by perceiving it with your eyes?</i> |
| (12) I have had them checked to <i>see</i> that they are sound, and they are no danger to any structure. | <i>perceive with the eyes</i> | I have had them checked to make sure that they are sound, and they are no danger to any structure [AS IF] <i>by perceiving it with my eyes</i> . |
| (13) It's up to you to <i>see</i> that the job's done properly. | <i>perceive with the eyes</i> | It's up to you to make sure that the job's done properly [AS IF] <i>by perceiving it with your eyes</i> . |

In contexts (10), (11), (12) and (13), the verb *see* is used in a metaphorical sense, as it is related to the PM through the comparative semantic feature [AS IF]. In forming this meaning, the source domain of EYE PERCEPTION is mapped onto the target domain MENTAL PERCEPTION, which yields the meaning *visually check that something is done*. It may seem that in contexts (10) and (11) the verb *see* is used literally, since the object of visual perception is, at first sight, physically visible objects: lights and kids. However, the position of the direct object of *see* is not taken by these nouns, but by object clauses, which are always associated with verbs of mental activity. This metaphorical sense of *see* can be formulated as follows: *see₆* — *make sure that something is done as if by perceiving the result with the eyes*.

It is interesting to note in relation to the semantics of the verb *see* that, due to speakers' capacity of abstract thinking and imagination that stems from it, the boundary between verbally described, actually perceived, and mentally modulated objects is, to some extent, blurred. Thus, a native English speaker uses the verb *see* not only to refer to actually visible objects but also to imagined objects and situations. As the interpretation of meanings *see*₂, *see*₃, *see*₅ and *see*₆ demonstrates, visual perception is closely connected with the processing of information, which is why English speakers use this verb to refer to mental processes. The same is observed in context (14).

- (14) I can't *see* why he's so up- *perceive with the eyes* I can't *perceive*, [AS IF] *with*
 set. *my eyes*, why he's so upset.

As in examples (10)–(13), in this case there is a contextual cue — an object clause as the direct object. Consequently, the contextual meaning of the verb *see* is associated with mental activity and belongs to the domain of MENTAL ACTIVITY or more precisely to its subdomain of INFORMATION PERCEPTION. Since the PM of the verb *see* belongs to the domain of EYE PERCEPTION, the sense found in context (14) is metaphorical. Visual perception is the basis of interaction with the world and the main source of information. Visual images become the foundation for logical reasoning, which is why the processes of visual perception and mental processing are closely related. In such situations, first of all, the [PERCEPTION] seme is actualised, promoting the metaphorical shift from EYE PERCEPTION to MENTAL PERCEPTION. As a result, the verb *see* is reinterpreted as a verb of intellectual activity. The resulting metaphorical sense can be worded as follows: *see*₇ — *understand something, as if by perceiving it with the eyes*.

- (15) I don't *see* the situation *perceive with the eyes* I don't *perceive* the situation
 quite as negatively as [AS IF] *with my eyes*, quite
 you do. as negatively as you do.

The sense of *see* in (15) is metaphorical, as it is derived from the PM on the basis of the comparison feature [AS IF]. This transfer results from the mapping of the EYE PERCEPTION domain onto the MENTAL PERCEPTION domain. The context provides the following cues. The object of perception is the abstract concept of *situation*, which is a mental construct. The adverb *negatively* expresses an attitude, which is not typical of sensory perception, but rather of cognitive information processing. As in the case of *see*₇ first of all the [PERCEPTION] feature is actualised. The mode of perception is determined by the interpreter based on these contextual cues. The difference between this meaning and *see*₇ is that in combination with the circumstance of the mode of action, the meaning of the verb becomes inseparably linked to the attitude expressed by the adverb *negatively*. The same is observed in example (16).

- | | | |
|--|-------------------------------|--|
| (16) Having a child makes you <i>see</i> things differently. | <i>perceive with the eyes</i> | Having a child makes you <i>perceive</i> things differently [AS IF] <i>with the eyes</i> . |
|--|-------------------------------|--|

The adverb *differently* in (16) specifies the meaning of the verb *see*, thus serving as a contextual cue that facilitates the interpretation of its sense which can be presented as *see*₈ — *consider something in a particular way as if by perceiving it with the eyes*.

- | | | |
|--|-------------------------------|--|
| (17) Can you really <i>see</i> her as the president? | <i>perceive with the eyes</i> | Can you really <i>perceive</i> her as the president [AS IF] <i>with the eyes</i> ? |
|--|-------------------------------|--|

Contextual cues such as *really* and *as the president* indicate that the person referred to by the pronoun *she* does not hold a leading position in a certain institution. In context (17), as in (15) and (16), the seme [PERCEPTION] is foregrounded, and this LSV of *see* is associated with intellectual activity. The same sense of *see* is found in examples (18) and (19).

- | | | |
|---|-------------------------------|--|
| (18) He could <i>see</i> a great future for her in music. | <i>perceive with the eyes</i> | He could <i>perceive</i> a great future for her in music [AS IF] <i>with his eyes</i> . |
| (19) She's got a new book coming out but I can't <i>see</i> it doing very well. | <i>perceive with the eyes</i> | She's got a new book coming out but I can't <i>perceive</i> it doing very well [AS IF] <i>with my eyes</i> . |

Such contextual elements as *great future* and *can't see it doing very well* trigger an image of entities that do not exist in the physical world but are conceived of as real. Considering this, the sense of *see* in contexts (17), (18) and (19) can be described as *see*₉ — *imagine something which is not true as if by perceiving it with the eyes*.

- | | | |
|---|-------------------------------|---|
| (20) As we <i>saw</i> in Chapter 2, the reasons for the war were complex [173]. | <i>perceive with the eyes</i> | As we <i>found out</i> from Chapter 2 [AS IF] we <i>perceived</i> this information <i>with our eyes</i> , the reasons for the war were complex. |
| (21) I'll go and <i>see</i> what he wants [181]. | <i>perceive with the eyes</i> | I'll go and <i>find out</i> what he wants [AS IF] I'll <i>perceive</i> this information <i>with the eyes</i> . |

In contexts (20) and (21), the verb *see* is used metaphorically: the domain of EYE PERCEPTION is mapped onto the domain of OBTAINING INFORMATION. The verb *see*

in this context does not refer to eye perception, but rather to the result of obtaining information. This sense is similar to *see*₃ and *see*₇. However, in (20) and (21), the verb is used to describe the result of thought processing of information obtained from a known source. The metaphorical transfer is based on likening visual perception to mental perception. This meaning can be presented as *see*₁₀ — *find something out as if through eye perception*.

- (22) He *saw* action in Vietnam. *perceive with the eyes* He *took part in* action in Vietnam [AS IF] he *perceived it with his eyes*.

The direct object of *see* in (22) refers to an abstract concept — *action*, which cannot be perceived visually. The metaphorical transfer is based on the mapping of the EYE PERCEPTION domain onto the EXPERIENCE domain: the idea of "seeing" is projected onto the whole perception of combat — *experience acquired as if through eye perception*. The same sense of *see* is found in (23):

- (23) The region has *seen* some of the fiercest fighting in the war. *perceive with the eyes* Some of the fiercest fighting in the war *took place* in the region [AS IF] the region *perceived it with the eyes*.

The semantic transfer in (23) is driven by the same mechanism: *experiencing something as if through eye perception*. In this case, a twofold understanding is possible. Either the noun *region* is used metonymically and refers to the people who live in that area, or it is used in its PM with the subject of perception being the region itself. Regardless of the interpretation, the meaning of the verb *see* remains metaphorical, as it is conditioned by the interaction of independent cognitive domains (cf. the interpretation in (22)). Thus, the meaning of the verb *see* in (22) and (23) takes the form of *see*₁₁ — *experience something as if by perceiving it with the eyes*.

The extensive polysemy of the verb *see* can likely be attributed to vision being the most dominant of the five senses and, consequently, essential for humans in navigating and interpreting their environment. Everything that falls within an observer's field of vision undergoes cognitive processing (Maturana 1980: 8). This is probably why all the LSVs of the verb *see* analysed above include the integral feature [PERCEIVE].

If the verb refers to the information (cognitive change triggered by external causes) that is perceived visually the [WITH THE EYES] seme is present. The object of perception may be an abstract entity, as in contexts (16) and (23). In this case, the observer can only mentally "see" it and perceive the image if it were real (i.e., there is a change in its state caused by internal causes). For a human being whose nervous system is a closed neuronal network, there is no significant distinction between a visual image triggered by internal or external changes in relative neuronal activity (Ritchie 2013: 46). Therefore, actually seeing

with the eyes and "as if seeing with the eyes" is essentially the same process. That is why the seme [WITH THE EYES] is necessarily present in the LP formula.

Besides that, since the LP of a polysemous word underlies all of its senses, its formula must encompass both metonymic and metaphorical extensions. Since metonymic transfer is based on contiguity, a metaphorical sense of a verb may be described as action, process or state associated with the PN in some way. Unlike PM, all the metonymic senses of the verb *see* discussed above are not associated with a state, but rather with a purposeful action — obtaining information. Evidence for this is the use of these LSVs in the imperative mood (e.g., *Please see₃ who's knocking. See₃ chapter 12*) or in the Present Continuous (e.g., *Did you know that she is seeing₃ an older man.*)

As the analysis shows, all the LSVs of the verb *see* are motivated by its nominative and non-derivative meaning. The meaning of *see₁* is motivated on the basis of direct denotation, so in the dictionary entry it is formulated as:

LITERAL: *perceive with the eyes*

The meanings of LSVs *see₂–see₅* are metonymic extensions of the PM:

see₂ — *to watch (film, television, etc.) by perceiving with the eyes*

see₃ — *to look up, refer to, or find something out by perceiving it with the eyes*

see₄ — *to visit or meet someone one knows socially or by chance while perceiving them with the eyes*

see₅ — *to accompany somebody to make sure they get there by perceiving it with the eyes.*

Since the meanings of these LSVs are somehow associated with the PM through the idea of contiguity, the generalised formulation for the given senses is as follows:

METONYMIC: *to do something associated, in some way, with eye perception.*

The meanings of LSVs *see₆–see₁₁* are formed on the basis of metaphorical transfer:

see₆ — *make sure that something is done as if by perceiving the result with the eyes*

see₇ — *understand something as if by perceiving it with the eyes*

see₈ — *consider something in a particular way as if by perceiving it with the eyes*

see₉ — *imagine something which is not true as if by perceiving it with the eyes*

see₁₀ — *find something out as if through eye perception*

see₁₁ — *experience something as if by perceiving it with the eye*

The meanings of these LSVs include the comparison feature [AS IF], which associates it with the PM. In accordance with the conventions of LP theory, when used in the LP formula, this seme takes the form of *as it were*. The generalised metaphorical meaning of the above LSVs is formulated as follows:

METAPHORICAL: *perceive with the eyes, as it were*

Thus, the semantic structure of the verb *see* takes the form of:

LITERAL: *perceive with the eyes*

METONYMIC: *to do something associated, in some way, with eye perception*

METAPHORICAL: *perceive with the eyes, as it were.*

To formulate the LP all the three elements must be combined:

see — *perceive with the eyes, or do something associated, in some way, with eye perception, or as it were.*

This formula describes in the most general form all three types of meanings of the verb: the PM, metonymic and metaphorical senses. The following signposts are included in the dictionary entry to help infer contextual meanings:

*see*₂ — *watch (film, television, etc.) by perceiving with the eyes* → **WATCH (TELEVISION, ETC.)**

*see*₃ — *look up, refer to, or find something out by perceiving it with the eyes* → **CHECK/LOOK/REFER TO/FIND OUT**

*see*₄ — *visit or meet someone one knows socially or by chance while perceiving them with the eyes* → **MEET/VISIT**

*see*₅ — *accompany somebody to make sure they get there by perceiving it with the eyes* → **GO WITH SOMEBODY SOMEWHERE**

*see*₆ — *make sure that something is done as if by perceiving the result with the eyes* → **MAKE SURE**

*see*₇ — *understand something as if by perceiving it with the eyes* → **UNDERSTAND**

*see*₈ — *to consider something in a particular way as if by perceiving it with the eyes* → **CONSIDER**

*see*₉ — *imagine something which is not true as if by perceiving it with the eyes* → **IMAGINE**

*see*₁₀ — *find something out as if by eye perception* → **FIND OUT INFORMATION**

*see*₁₁ — *experience something as if by perceiving it with the eye* → **EXPERIENCE SOMETHING**

Model of a dictionary entry based on LP

Based on the analysis, a dictionary entry for the verb *see* is compiled:

LEXICAL PROTOTYPE: *perceive with the eyes, do something associated with eye perception, or as it were*

MEANINGS:

LITERAL: *perceive with the eyes.*

If the operation is successful, he will be able to see again. • Can you see the bird in that tree?

• The moment we saw the house, we knew we wanted to buy it.

METONYMIC: *do something associated, in a certain way, with eye perception*

<i>Did you see that documentary on Channel 4 last night?</i>	WATCH (TELEVISION, ETC.)
<i>Please see who's knocking • Elements are usually classified as metals or non-metals (see chapter 12) (=refer to chapter 12 for further information). • I must see your passport before you can enter the country.</i>	CHECK/LOOK/REFER TO/FIND OUT
<i>We still see each other a couple of times a month • See you at the station at 6 o'clock • See you (informal, used to say goodbye).</i>	MEET/VISIT
<i>Can I see you home? • My mother used to see me across the road. • Let me see you to the door.</i>	GO WITH SOMEONE SOMEWHERE
<i>Please see that the lights are switched off before you leave.</i>	MAKE SURE

METAPHORICAL: perceive with the eyes, as it were

<i>I can't see why he's so upset. • I see what you mean. • You see, the thing is, I'm really busy right now.</i>	UNDERSTAND
<i>Having a child makes you see things differently. • I don't see the situation quite as negatively as you do.</i>	CONSIDER
<i>Can you really see her as the president? • He could see a great future for her in music. • She's got a new book coming out but I can't see it doing very well.</i>	IMAGINE
<i>As we saw in Chapter 2, the reasons for the war were complex.</i>	FIND OUT INFORMATION
<i>He saw action in Vietnam. • The region has seen some of the fiercest fighting in the war.</i>	EXPERIENCE SOMETHING

Apart from sharing some features with traditional entries in learner's dictionaries (such as the structure of microentries and the use of signposts), the method of lexicographic description of a polysemous word based on its LP presented in this paper also has a number of differences. The main difference is the absence of a list of definitions. The entry begins with a formula for LP of the headword, followed by the three types of meanings described above. The PM which describes the prototypical sense — *perceive with the eyes* — serves as the starting point for the formation of all LSVs. Unlike the PM, which refers to a state, in its metonymic senses the verb *see* denotes actions — *watch, look up, visit*, etc. — which are nonetheless associated with visual perception.

Metaphorical senses are related to the PM through the feature of comparison [AS IF/AS IT WERE]. The metaphorical nature of the actions falling under this definition is generally determined by the nature of the objects of perception — abstract concepts that are inaccessible to visual perception. Therefore, in all meta-

phorical uses, the verb *see* refers to mental processes that the speaker conceptualises as *perceive with the eyes, as it were*.

The features that underlie metonymic and metaphorical transfers are indicated in the entry by means of signposts. These signposts do not serve as definitions, they are intended to help the dictionary user navigate the entry. Supplemented by illustrative examples, they serve as contextual cues that enable the user to infer the contextual meanings of the polysemous headword. They reveal, clarify and actualise a certain aspect of the LP, which may help the learner to form a comprehensive idea of the word's meaning.

A dictionary entry of this type also offers the user a fundamentally different perspective on the nature of polysemy: the existence of a systemic meaning and its realisations in speech. The generalised and universal character of the LP formula, which encompasses the primary, metaphorical, and metonymic meanings, demonstrates the motivation of the LSVs by the LP and shows the semantic relations between them. It also helps conceptualise the distinction between metonymic and metaphorical meanings.

From a theoretical point of view, a lexicographic description of this type clearly demonstrates the real non-mirroring of the correlation between the levels of language and speech. At the system level, the meaning of a polysemous word exists as an invariant structure — the LP — while at the speech level it is rendered through a context-based meaning of a specific LSV. Thus, it can be assumed that the proposed dictionary entry reflects the cognitive mechanisms of semiosis.

In conclusion, it should be added that the proposed dictionary entry is a theoretical model that can serve as a basis for further lexicographic research. The suggested approach to the lexicographic description of polysemous words does not imply a rejection of traditional methods. A dictionary compiled according to the principles discussed here can be used as an additional tool for describing polysemous words, alongside traditional dictionaries and teaching materials.

References

- Ageeva, N.G. 1990. *Tipologiya i mekhanizmy glagol'noy metonimii v sovremennom angliyskom yazyke*. Unpublished Candidate of Sciences Dissertation. Kiev: Kiev State Pedagogical Institute of Foreign Languages.
- Alefrenko, N.F. 2005. *Spornye problemy semantiki*. Moscow: Gnozis.
- American Heritage Dictionary*.
<https://ahdictionary.com/> [10 April 2025]
- Apresyan, Yu. D. 1984. *Lingvisticheskoye obespechenye sistemy frantsuzsko russkogo perevoda ETAP-1*. Moscow: Institute of Foreign Languages of USSR Academy.
- Arkhipov, I.K. 1998. Problemy yazyka i rechi v svete prototipicheskoy semantiki. *Problemy lingvistiki i metodiki prepodavaniya inostrannykh yazykov* 6: 5-22.
- Arkhipov, I.K. 2003. *Chelovetchesky faktor v yazyke*. St. Petersburg: NIYaK.
- Arkhipov, I.K. 2004. Kommunikativny tseitnot i prototipicheskaya semantika. *Izvestiya Rossiyskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena* 4(7): 75-84.

- Arkhipov, I.K., D.N. Novikov and S.A. Pesina. 2006. Organizatsiya i metod postroyeniya slovarnoy statyi na osnove printsypa proto:picheskoy semantyki. Gumanova, M.K. (Ed.). 2006. *Proceedings of the 5th International Conference, 'Yazyki v sovremennom mire'*: 67-73. Moscow: KDU.
- Arnold, I.V. 2010. Polisemiya sushchestvitelnogo i leksiko-grammaticheskie razryady. Bukharkin, P.E. (Ed.). 2010. *Semantika. Stilistika. Intertekstualnost*: 5-24. Moscow: Knizhny dom "LIBROKOM".
- Arutyunova, N.D. 1999. *Yazyk i mir cheloveka*. Moscow: Yazuki russkoy kultury.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Brudny, A.A. 1971. Znachenie slova i psikhologiya protivopostavleniy. Leontyev, A.A. (Ed.). 1971. *Semanticheskaya struktura slova: psikhologicheskie issledovaniya*: 19-27. Moscow: Nauka.
- Brugman, C. and G. Lakoff. 1988. Cognitive Topology and Lexical Networks. Small, S., G. Cottrell and M. Tannenhaus (Eds.). 1988. *Lexical Ambiguity Resolution*: 477-507. San Mateo: Morgan Kaufman.
- Budagov, R.A. 1947. *Slovo i ego znachenie*. Moscow: Nauka.
- Cambridge Dictionary*.
<https://dictionary.cambridge.org/> [10 April 2025]
- Chudinov, A.P. 2003. *Rossiya v metaforicheskom zerkale: Kognitivnoe issledovanie politicheskoy metafory (1991–2000)*. Yekaterinburg: UGPU.
- Collins COBUILD*.
<http://www.collinslanguage.com/collins-elt-learners-of-english/cobuild#.Ul8XqFCST8k>
[8 April 2014].
- Cowie, A.P. 2007. *English Dictionaries for Foreign Learners: A History*. (A Reprint of the 1999 edition.) Oxford: Oxford University Press.
- Devkin, V.D. 2000. *Ocherki po leksikografii*. Moscow: Prometey.
- Dretske, F.I. 1969. *Seeing and Knowing*. London: Routledge & Kegan Paul.
- Dubchinskiy, V.V. 1994. *Iskusstvo sozdaniya slovaray: Konspekt po leksikografii*. Kharkov: KhPU.
- Guillaume, G. 1973. *Principes de linguistique théorique de Gustave Guillaume*. Québec/Paris: Librairie C. Klincksieck.
- Halas, A. 2016. The Application of the Prototype Theory in Lexicographic Practice: A Proposal of a Model for Lexicographic Treatment of Polysemy. *Lexikos* 26: 124-144.
<https://doi.org/10.5788/26-1-1338>
- Hanks, P. 2000. Do Word Meanings Exist? *Computers and the Humanities* 34(1-2): 205-215.
- Jackson, H. 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Jakobson, R.O. 1985. *Izbrannye raboty*. Moscow: Progress.
- Karasev, A.A. 2011. Definitii v tolkovykh slovaryakh i leksicheskoe znachenie slova. Fetisov, Anatoly (Ed.). 2011. *Proceedings of Herzen Readings: Modern Languages Interuniversity Academic Conference, St. Petersburg, Russia, May 19–20, 2011*: 188-189. St. Petersburg: Herzen State Pedagogical University of Russia.
- Karasev, A.A. 2013. Polysemous Words Monolingual Learner's Dictionary Entry (Lexical-Prototypical Approach). *St. Petersburg State Polytechnical University Journal: Humanities and Social Sciences* 3(179): 206-210.
- Karasev, A.A. 2014a. *Biosemioticheskie printsipy uchebnogo odnoyazychnogo slovary (na materiale sovremennogo angliyskogo yazyka)*. Unpublished Candidate of Sciences Dissertation. St. Petersburg: Herzen State Pedagogical University of Russia.

- Karasev, A.A. 2014b. Uchebnyy odnoiazychnyy slovar' angliyskogo yazyka v svete biosemiotiki. *Nauchnoe mnenie* 4: 202-209.
- Karasev, A.A. 2014c. Yavlenie polisemii s tochki zreniya biosemiotiki. *Izvestiya Volgogradskogo gosudarstvennogo pedagogicheskogo universiteta. Filologicheskie nauki* 2(87): 16-20.
- Katsnelson, S.D. 2011. *Soderzhanie slova, znachenie i oboznacheniye*. Moscow: Editorial URSS.
- Kurylowicz, J. 1955. Zametki o znachenii slova. *Voprosy yazykoznaniiya* 3: 73-81.
- Leshcheva, L.M. 1985. *Svyaz' leksicheskikh znacheniy v semanticheskoy strukture (na materiale prilagatel'nykh sovremennogo angliyskogo yazyka)*. Unpublished Candidate of Sciences Dissertation. Minsk: MGLU.
- Levchina, I.B. 2003. *Razvitie semanticheskoy struktury sinestezicheskikh prilagatel'nykh*. Unpublished Candidate of Sciences Dissertation. St. Petersburg: Herzen State Pedagogical University of Russia.
- Litvin, F.A. 1984. *Mnogoznachnost' slova v yazyke i rechi*. Moscow: Vysshaya shkola.
- Lodwig, R.R. and E.F. Barret. 1967. *The Dictionary and the Language*. New York: Hayden.
- Longman Dictionary of Contemporary English*.
<https://www.ldoceonline.com/> [10 April 2025]
- Maturana, H.R. 1980. Biology of Cognition. Maturana, H. and F.J. Varela (Eds.). 1980. *Autopoiesis and Cognition: The Realization of the Living*: 2-62. Dordrecht: D. Reidel.
- Mednikova, E.M. 2010. *Znachenie slova i metody ego issledovaniya*. Moscow: Vysshaya shkola.
- Merriam-Webster Dictionary*.
<https://www.merriam-webster.com/> [10 April 2025]
- Oxford English Dictionary*.
<https://www.oed.com/?tl=true> [10 April 2025]
- Pertsov, N.V. 2001. *Invarianty v russkom slovoizmenenii*. Moscow: Yazyki russkoy kultury.
- Pesina, S.A. 1998. *Leksicheskii prototip v semanticheskoy strukture slova*. Unpublished Candidate of Sciences Dissertation. St. Petersburg: Herzen State Pedagogical University of Russia.
- Pesina, S.A. 2005a. *Issledovanie semanticheskoy struktury slova na osnove prototipicheskoy semantiki (na materiale angliyskikh sushchestvitel'nykh)*. Unpublished Doctor of Sciences Dissertation. St. Petersburg: Herzen State Pedagogical University of Russia.
- Pesina, S.A. 2005b. Metodika opredeleniya sodержatel'nogo yadra mnogoznachnogo sushchestvitel'nogo angliyskogo yazyka. *Izvestiya Rossiyskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena* 5(11): 51-59.
- Pesina, S.A. 2005c. *Polisemiya v kognitivnom aspekte*. St. Petersburg: Herzen State Pedagogical University of Russia.
- Ritchie, L.D. 2013. *Metaphor*. Cambridge/New York: Cambridge University Press.
- Rivelis, E.I. 2009. Slovar' v svete kognitivnoy lingvistiki. *Studia linguistica cognitiva* 2: Nauka o yazyke v izmenyayushcheyseya paradigme znaniya: 304-350.
- Shiryaeva, A.V. 2008. Prototipicheskaya semantika kak osnovanie sinonimii glagolov. *Izvestiya Rossiyskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena* 23(54): 269-274.
- Shmelev, D.N. 1973. *Problemy semanticheskogo analiza leksiki (na materiale russkogo yazyka)*. Moscow: Nauka.
- Slonimskaya, I.M. 1990. Leksikograficheskaya teoriya i leksikograficheskaya pragmatika. Perebynos, V.I. (Ed.). 1990. *Statisticheskaya leksikografiya i uchebnyy protsess*: 48-79. Kiev: KGPIYa.
- Smirnitky, A.I. 1954. K voprosu o slove: Problema tozhdestva slova. *Trudy Instituta yazykoznaniiya AN SSSR* 4: 3-49.

- Soldatov, B.G.** 1997. Semiotika sinonimicheskogo opredeleniya v angliyskoy odnoiazychnoy leksikografii. *Proceedings of Aktual'nye problemy teoreticheskoy i prikladnoy leksikografii the 2nd International Workshop on Lexicography, Ivanovo, Russia, 25–27 September 1997*: 60-69. Ivanovo: Ivanovsky State University.
- The Chambers Dictionary.**
<https://chambers.co.uk/> [10 April 2025]
- Ufimtseva, A.A.** 1962. *Opyt izucheniya leksiki kak sistemy*. Moscow: Nauka.
- Ufimtseva, A.A.** 1984. K voprosu o tak nazyvaemom definitsionnom metode opisaniya leksicheskogo znacheniya slova. Yartseva, V.N. (Ed.). 1984. *Slovo v grammatike i slovare*: 134-142. Moscow: Nauka.
- Ufimtseva, A.A.** 2010. *Leksicheskoe znachenie: printsip semiologicheskogo opisaniya leksiki*. Moscow: LIBROKOM.
- Vinogradov, V.V.** 1947. *Russkiy yazyk. Grammaticheskoe uchenie o slove*. Moscow / Leningrad: Nauka.
- Webster's Online Dictionary.**
<http://www.websters-online-dictionary.org/> [8 April 2014]
- Yakovlyuk, A.N.** 2009. Leksiko-semanticheskoy variant kak svyazuyushchee zveno mezhdu mnogoznachnym slovom v yazyke i ego realizatsiy v rechi. *Bulletin of Chelyabinsk State University* 34(172): 156-162.
- Zalevskaya, A.A.** 2000. *Vvedenie v psikholingvistik*. Moscow: Rossiyskiy gosudarstvennyy gumanitarnyy universitet.
- Zaliznyak, A.A.** 2004. Fenomen mnogoznachnosti i sposoby ego opisaniya. *Voprosy yazykoznaniya* 2: 20-45.
- Zvegintsev, V.A.** 1957. *Semasiologiya*. Moscow: Moscovsky Universitet.

Exemplification of Sensitive Words for People with Disabilities in Monolingual English Learner's Dictionaries

Meimei Li, *Center for Linguistics and Applied Linguistics, Guangdong
University of Foreign Studies, Guangzhou,
China (lmm19854460393@163.com)*
(<https://orcid.org/0009-0009-5930-191X>)

Abstract: This paper investigates how sensitive words related to people with disabilities are exemplified in the "Big Five" monolingual English learner's dictionaries. The findings show that learner's dictionaries tend to be cautious in exemplifying such terms, with notable differences in the number of words exemplified and examples provided, as well as in the inclusion of additional examples. Despite these variations, the "Big Five" consistently offer phrase and sentence examples for sensitive words that serve both decoding and encoding purposes. The analysis reveals that the exemplification of sensitive words is influenced by factors such as part-of-speech, attitude labelling, and word currency. In general, adjectival forms of sensitive words are more likely to be exemplified than their nominal counterparts. Sensitive words that remain in current use tend to have higher exemplification rates than those considered old-fashioned. Neutral or euphemistic expressions are more frequently illustrated with examples than terms that carry negative connotations. This paper argues for broader exemplification of sensitive words, particularly neutral and euphemistic ones, and recommends optimising the presentation and the quality of additional examples.

Keywords: BIG FIVE, EXEMPLIFICATION, SENSITIVE WORDS, PEOPLE WITH DISABILITIES, INCLUSIVE LANGUAGE

Opsomming: Toeligting van sensitiewe woorde vir persone met gestremdhede in eentalige Engelse aanleerderswoordeboeke. Hierdie artikel ondersoek hoe sensitiewe woorde wat verband hou met persone met gestremdhede in die "Groot Vyf" eentalige Engelse aanleerderswoordeboeke toegelig word. Die bevindings toon dat aanleerderswoordeboeke daartoe neig om omsigtig te wees in die toeligting van sodanige terme, met opmerklike verskille in die aantal woorde wat toegelig word en die voorbeelde wat verskaf word, sowel as in die insluiting van bykomende voorbeelde. Ten spyte van hierdie variasie, bied die "Groot Vyf" konsekwent voorbeelde van frases en sinne vir sensitiewe woorde aan wat 'n dekodeerende sowel as enkoderende doel dien. Die analise toon dat die toeligting van sensitiewe woorde beïnvloed word deur faktore soos die woordsoort, etikettering wat ingesteldheid oordra, en hoe aktueel die woord is. Oor die algemeen word byvoeglike vorms van sensitiewe woorde meer toegelig as hulle naamwoordelike teenhangers. Sensitiewe woorde wat steeds in gebruik is, word meer dikwels toegelig as dié wat as outyds beskou word. Neutrale of eufemistiese leksikale items word meer dikwels met voorbeelde

toegelig as items waaraan negatiewe konnotasies geheg word. Hierdie artikel bepleit breër toeligting van sensitiewe woorde, veral neutrale en eufemistiese woorde, en beveel aan dat die aanbieding en kwaliteit van bykomende voorbeelde geoptimaliseer word.

Sleutelwoorde: GROOT VYF, TOELIGTING, SENSITIEWE WOORDE, PERSONE MET GESTREMDHEDE, INKLUSIEWE TAAL

1. Introduction

While the defining and usage labelling of sensitive words in monolingual English learner's dictionaries (MELDs) has been extensively discussed (e.g. Norri 2000, 2020; Van der Meer 2005; Coffey 2010), relatively little scholarly attention has been paid to their exemplification in MELDs. Two studies on the exemplification policy in learner's dictionaries (Stein 2002; Xu 2008) conclude that attitude-labelled and currency-labelled words are more likely to be exemplified than region or subject-marked words. However, both studies selected their target words at random, which naturally led to the inclusion of different types of words. Given the specific nature of sensitive language, findings based on the exemplification of general vocabulary may not be directly applicable to sensitive terms. It is thus unclear: (1) how learner's dictionaries exemplify (or fail to exemplify) sensitive words; (2) whether learner's dictionaries have a consistent exemplification policy for sensitive words; (3) what factors influence the exemplification of sensitive words in learner's dictionaries; (4) how we can improve the exemplification of sensitive words in learner's dictionaries. This paper, therefore, examines the exemplification of sensitive words with varying degrees of social acceptability in the "Big Five" through an analysis of twenty-one terms used to refer to people with disabilities (*cripple*, *deaf mute*, *developmentally disabled*, *differently abled*, *disabled*, *dumb*, *epileptic*, *handicapped*, *hard of hearing*, *insane*, *invalid*, *mentally handicapped*, *mentally ill*, *midget*, *neurodivergent*, *psychopath*, *retard*, *retarded*, *spastic*, *spaz*, and *visually impaired*). These terms were selected because they are frequently cited in previous scholarly research (e.g. Anderson and Fox 1988; Norri 2000, 2020; Allan and Burridge 2006) and institutional language guidelines, such as the WHO's *International Classification of Functioning, Disability and Health* (WHO 2013), demonstrating both their prominence in public discourse and their relevance to socially sensitive issues. They were looked up in the online editions of the *Oxford Advanced Learner's Dictionary of Current English* (OALD), the *Longman Dictionary of Contemporary English* (LDOCE), the *Cambridge Advanced Learner's Dictionary* (CALD), the *Collins COBUILD Advanced Learner's Dictionary* (COBUILD), and the second print edition of the *Macmillan English Dictionary for Advanced Learners* (MEDAL2 2007).¹

This article begins with a comparison of the exemplification of sensitive words for people with disabilities in the "Big Five", which is followed by an examination of factors influencing the exemplification in these dictionaries. The study concludes by presenting several tentative suggestions on how to improve the exemplification of sensitive words in learner's dictionaries.

2. Exemplification of sensitive words for people with disabilities in MELDs: Similarities

In order to investigate the exemplification policies adopted in MELDs, I examined the inclusion of examples for the sensitive words selected for the present study. The findings are presented in tabular form (see Table 1) to better summarise the similarities and differences between the dictionaries studied. A plus sign in the table means that the word is exemplified in the dictionary. A minus sign indicates that no example is provided for the word. A zero means that the word or the relevant disability sense is not included in the dictionary. The numbers placed within brackets indicate the number of examples given. For adjectives that can be used nominally as collectives (e.g. *disabled*, *handicapped*, and *insane*), I counted their nominal and adjectival uses related to disability as two distinct entries.

Since the websites of the four learner's dictionaries do not indicate the source of the additional examples, it remains unclear whether these examples are automatically extracted from corpora or selected by dictionary compilers. This lack of transparency raises concerns about their reliability. As Rundell (2015) observes, such examples are often subject to little or no filtering in terms of quality or appropriacy. In light of these concerns, this part of the comparison focuses solely on the examples included in the core entry. The additional examples will be discussed separately in Section 3.

Table 1: Exemplification of sensitive words for people with disabilities in MELDs

Lexical item	Part-of-speech (POS)	OALD	LDOCE	CALD	COBUILD	MEDAL2
cripple	n.	-	-	-	-	-
deaf mute	n.	-	-	-	-	-
developmentally disabled	adj.	0	0	+ (2)	0	0
differently abled	adj.	0	0	+ (2)	0	0
disabled	adj.	+ (7)	+ (4)	+ (1)	+ (1)	+ (1)
	n.	+ (1)	+ (1)	-	+ (1)	-
dumb	adj.	+ (1)	-	+ (1)	+ (1)	-
epileptic	adj.	+ (1)	+ (1)	+ (2)	+ (1)	+ (1)
	n.	-	-	-	0	-
handicapped	adj.	+ (2)	+ (1)	-	+ (2)	-
	n.	-	-	-	+ (1)	-
hard of hearing	adj.	+ (2)	-	+ (1)	+ (1)	-
	n.	+ (1)	-	0	0	0
insane	adj.	+ (2)	+ (1)	+ (2)	+ (1)	+ (1)
	n.	+ (1)	+ (1)	-	+ (1)	0
invalid	n.	+ (3)	+ (1)	+ (1)	+ (1)	+ (1)
mentally handicapped	adj.	-	-	0	0	-
mentally ill	adj.	0	-	0	0	0
midget	n.	-	-	-	-	-

neurodivergent	adj.	+	(2)	0	+	(2)	-	0
psychopath	n.	-	-	-	-	-	+	(1)
retard	n.	0	0	-	-	-	+	(1)
retarded	adj.	-	-	-	-	-	-	-
spastic	adj.	-	-	-	-	-	-	-
	n.	-	-	-	-	-	-	0
spaz	n.	0	0	-	-	-	0	0
visually impaired	adj.	0	-	-	+	(2)	0	0
	n.	0	0	-	+	(2)	0	0
	Words	11	7	11	12	4		
	exemplified							
	Examples	23	10	18	13	4		
	provided							

A comparison of the exemplification of the selected words in the "Big Five" reveals some similarities in how sensitive words for people with disabilities are exemplified. These similarities pertain to the number of examples provided, the types of examples used, and the functions those examples serve.

2.1 Number of examples

While the importance of examples in dictionaries has been emphasised by many lexicographers (e.g. Zgusta 1971: 263-268; Landau 2001: 305-308; Atkins and Rundell 2008: 452-462), sensitive terms for people with disabilities, particularly those considered offensive (e.g. *cripple*, *deaf mute*, *midget*, *retarded*, and *spastic*), appear to be exemplified less frequently in learner's dictionaries. The exemplification of the selected words in the "Big Five" reveals an overall lack of examples and unevenness in the coverage of examples.

Table 2 presents the exemplification frequency of the selected words in five MELDs. The data in the table shows a significant lack of examples for disability-related words in learner's dictionaries. While these dictionaries include between 18 and 25 words, the percentage of exemplified words varies widely, with some dictionaries providing very few examples. The highest rate of exemplification is found in COBUILD (63%), followed by OALD (52%), CALD (44%), LDOCE (32%), and MEDAL2 (22%). This means that even in the best case, more than a third of the words surveyed lack illustrative examples.

Table 2: Frequency of exemplification of words for people with disabilities in MELDs

	OALD	LDOCE	CALD	COBUILD	MEDAL2
Words included	21	22	25	19	18
Words exemplified	11	7	11	12	4
Percentage of exemplification	52%	32%	44%	63%	22%

Furthermore, there are significant variations in how specific words are exemplified in different dictionaries. *Disabled* (adj.), *epileptic* (adj.), *insane* (adj.), and

invalid (n.) are the only words consistently illustrated with examples in all five dictionaries. In contrast, *cripple* (n.), *deaf mute* (n.), *midget* (n.), *retarded* (adj.), and *spastic* (adj.) are not exemplified in any of them. The remaining words are treated differently across the five dictionaries.

One possible explanation for the absence of examples for sensitive terms in learner's dictionaries might be the limited corpus evidence available for such words. Many of these terms, particularly those targeting marginalised groups, occur infrequently in general corpora, making it difficult for lexicographers to extract appropriate corpus examples to illustrate their usage. Moreover, slang terms such as *spaz* are typically found in spoken language, which tends to be underrepresented in most corpora compared to written language. This further constrains the availability of suitable examples. However, limited corpus evidence alone may not fully account for the dearth of examples. For instance, while *cripple* (n.) appears 125 times in the *British National Corpus* (BNC), it is not exemplified in any of the "Big Five", whereas *spaz* (n.), which appears only once in the BNC, is exemplified in one of them (CALD). This suggests that the lack of examples for sensitive words may not be solely attributable to corpus frequency, and that other factors may also be at play.

A more compelling reason for the scarcity of examples for sensitive words in learner's dictionaries may be some sensitivity concerns. Lexicographers tend to take a cautious approach to sensitive words in case they might be criticised for normalising offensive language. Hughes (2006: 128) discusses the potential risk of including offensive terms in dictionaries:

Although modern dictionaries are predicated on the assumption that they are *descriptive* — that is, that they should reflect actual usage — there are both lingering and new social pressures about what is appropriate to appear in print. There are also tenacious assumptions that simply by printing an offensive term a dictionary is in some way dignifying the term, relaxing standards, encouraging laxity, or endorsing prejudices.

Dictionaries have to accordingly use various strategies to respond to these pressures, one of the most common being the exclusion of usage examples.

A final possible reason for the limited exemplification of sensitive words might be concern about their potential misuse. Rather than risk reinforcing negative usage or failing to convey the nuanced contexts in which such terms might appear, lexicographers may choose to avoid providing examples altogether, even when corpus evidence is available.

2.2 Types of examples

Learner's dictionaries also show similarities in the types of examples they provide. They all include both phrase examples and sentence examples for words related to people with disabilities. The number of each type of example is shown in Table 3.

Table 3: Number of different types of examples in MELDs

	OALD	LDOCE	CALD	COBUILD	MEDAL2
Phrase examples	7	5	1	3	1
Sentence examples	16	5	17	10	3
Total	23	10	18	13	4

As shown in Table 3, the "Big Five" generally give more sentence examples than phrase examples for the words examined. Phrase examples show typical collocations of the word examined or fixed expressions, such as the examples under *dumb*, *disabled*, and *epileptic* shown in (1), (2), and (3). Sentence examples, on the other hand, place the word in a specific context, as the examples for *differently-abled* and *insane* in (4) and (5) show.

- (1) a young deaf and dumb man (COBUILD)
- (2) a new home for severely disabled people (OALD)
- (3) an epileptic fit (CALD)
- (4) The same opportunities must be provided to differently-abled employees as to the rest of the workforce. (CALD)
- (5) The killer was declared criminally insane. (LDOCE)

Given the particular complexity and sensitivity of sensitive words, it is worth considering that future online learner's dictionaries could benefit from including more sentence examples for such terms. Sentence examples offer valuable contextual cues, such as syntactic patterns, collocational behaviour, and pragmatic implications. For sensitive or potentially offensive terms, understanding the context in which a word is typically used is crucial. By illustrating how a term is used in real-life situations, examples can help learners not only grasp its meaning but also develop the pragmatic competence necessary to avoid inappropriate or harmful usage. Moreover, several dictionary user studies have shown that EFL learners tend to favour full-sentence examples over short segments (e.g. Farina et al. 2019). This preference further supports the recommendation that online learner's dictionaries should consider providing more sentence examples for sensitive words.

2.3 Functions of examples

The functions of dictionary examples have been stressed by many lexicographers (e.g. Fox 1987; Atkins and Rundell 2008). Fox (1987: 137) argues that dictionary examples should "help to reinforce the meaning — not by acting as a reformulation of the definition, but by showing how the word is actually used, in an appropriate context, a typical grammatical structure, and with words that are normally associated with it". Atkins and Rundell (2008: 453-454) summarise the functions of a good dictionary example as including attestation, elucidating the meaning of words, and illustrating contextual features. However, these accounts have been criticised for failing to differentiate examples used for decoding pur-

poses and those for encoding purposes (Humb   2001; Stein 2002; Frankenberg-Garcia 2012, 2014). Thus, Humb   (2001) argues for a clear distinction between decoding and encoding examples.

The "Big Five" provide both decoding and encoding examples for the words examined, especially for neutral terms. Decoding examples are typically designed to offer contextual clues that help learners comprehend word meanings (Humb   2001). The following are examples listed under *invalid* in the "Big Five":

- (6) She had been a delicate child and her parents had treated her as an invalid. (OALD)
- (7) He was found guilty of murdering his invalid father. (OALD)
- (8) Her husband was an invalid and couldn't come to the door to open it. (OALD)
- (9) I resented being treated as an invalid. (LDOCE)
- (10) Is the invalid in bed? (CALD)
- (11) I hate being treated as an invalid. (COBUILD)
- (12) He had been an invalid for many years. (MEDAL2)

Examples (6) and (8) provide relatively clear contexts that help users infer the meaning of *invalid* through contextual cues. In contrast, examples like (7), (9), and (11) either focus on emotionally charged or abstract experiences, making the word's meaning harder to extract. Others, such as (10), lack sufficient context, reducing their effectiveness for modern learners. Ideally, examples provided in dictionaries should be natural, typical, informative, and intelligible (Atkins and Rundell 2008). However, the examples for sensitive words in the "Big Five" vary in clarity and pedagogical usefulness. While some provide meaningful and context-rich usage, others are vague or emotionally complex, potentially hindering comprehension. This variation highlights the importance of careful example selection, especially for potentially offensive words.

In contrast, encoding examples, as pointed out by Humb   (2001), should demonstrate grammatical patterns and collocations of the word. Such examples are mainly given to help facilitate production. Some examples listed under *disabled* are shown below.

- (13) physically/developmentally/intellectually disabled (OALD)
- (14) a new home for severely disabled people (OALD)
- (15) temporarily/permanently disabled (OALD)
- (16) physically/mentally disabled (LDOCE)
- (17) severely disabled (=unable to move most of your body) (MEDAL2)

Examples (13) to (17) are all encoding examples because they demonstrate common collocations of *disabled* with different modifiers (*physically, developmentally, intellectually, severely, temporarily, permanently, and mentally*). These examples do not seem particularly useful in terms of helping users understand what *disabled*

means. The provision of collocational information is aimed at facilitating users' language production.

It is worth noting that some examples for the words examined can deal with both the decoding and encoding needs of users, as the example for *disabled* in (18).

- (18) If you are elderly or physically disabled, massage can be beneficial.
(LDOCE)

The example primarily serves a decoding purpose because it helps users understand the meaning of *disabled* by linking it to a clear health-related context. However, it also has an encoding function in that it presents a common collocate of *disabled*, i.e. *physically*, which could help users produce similar sentences.

To conclude, while users of learner's dictionaries may rarely use sensitive words in their productive language, they still need to understand the meanings and usage of these terms when they encounter them. For offensive and taboo words, a clear usage label may often be sufficient to indicate their sensitive nature, especially given the difficulty of crafting example sentences that effectively convey their negative connotations without reinforcing harmful stereotypes. By contrast, exemplification is particularly important for neutral and euphemistic terms related to disability, as these are more likely to be used productively by learners and their appropriate contexts may not be self-evident. It is therefore essential that learner's dictionaries offer well-considered examples for such terms to support inclusive language use.

3. Exemplification of sensitive words for people with disabilities in MELDs: Main variations

Despite general similarities in the inclusion of illustrative examples in learner's dictionaries, the exemplification of disability-related sensitive terms in the "Big Five" shows notable intra-dictionary and inter-dictionary variations. These differences, to a great extent, reflect underlying editorial attitudes toward sensitive language.

3.1 Variations in the number of exemplified words and included examples

Learner's dictionaries differ from each other in the number of exemplified words for people with disabilities and the number of included examples. As Table 1 shows, COBUILD contains the highest number of exemplified words (12), followed by CALD and OALD, each with 11. LDOCE includes fewer exemplified words (7), while MEDAL2, the only print edition among the five dictionaries, has the fewest, with only four.

The limited exemplification of sensitive terms in MEDAL2 can be partly attributed to its print format, which imposes space constraints that may restrict

the number of examples provided. However, this factor alone does not fully explain the lack of exemplification. As previous research has shown, even the on-line version of MEDAL contains fewer examples than other MELDs, not only for sensitive terms but also for other terms such as academic words (Rees 2024).² This suggests that the sparse exemplification in MEDAL2 is more likely the result of an overall editorial policy rather than a mere consequence of dictionary format.

The relatively high number of illustrative examples in COBUILD reflects the exemplification policy of the dictionary, which states that "examples would be given for most of the words or senses of words explained in it" (Fox 1987: 137). The policy was proposed when the first edition of COBUILD was being planned and has been followed in later editions.

CALD adopts a more inclusive approach toward emerging vocabulary and euphemisms. It includes and exemplifies terms such as *developmentally disabled* and *differently abled*, whereas none of the other dictionaries include these terms. Additionally, among the three dictionaries that include *neurodivergent*, only CALD and OALD provide examples for the word.

While COBUILD exemplifies the most words, the number of examples provided for each sense is generally limited to one. In contrast, although OALD and CALD exemplify fewer words than COBUILD, they provide a greater number of examples. OALD even gives seven examples for the nominal sense of *disabled*. CALD generally provides two examples for most senses. MEDAL2, on the other hand, includes only one example per sense. Although LDOCE exemplifies seven words, it gives one example per sense except for the adjectival sense of *disabled* (with four examples).

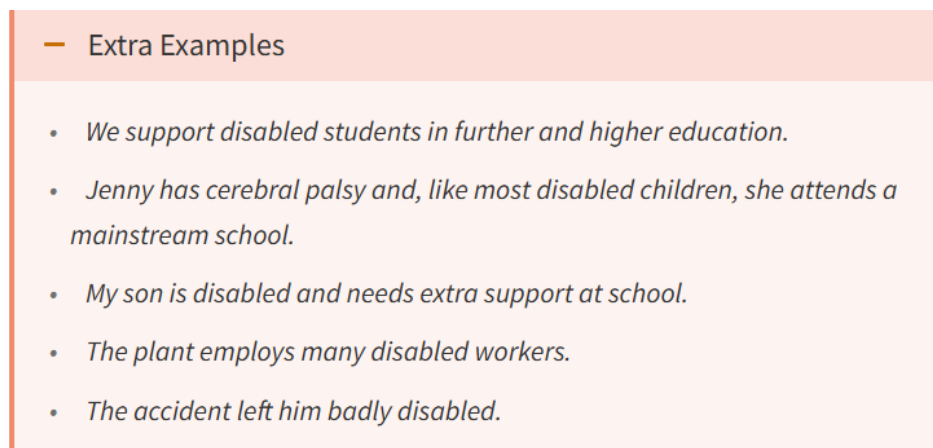
3.2 Variations in the inclusion of additional examples

A notable phenomenon of exemplification in online learner's dictionaries is the inclusion of additional examples within the extra sections of the entry. Among the four online learner's dictionaries, three (OALD, CALD, and LDOCE) provide a section of additional examples. Generally, these dictionaries do not provide information on their websites about how examples are selected. It is therefore unclear to users whether these additional examples are automatically generated from corpora or chosen by lexicographers.

OALD provides a separate **Extra Examples** section in addition to the examples included in the core entry (for an example, see the **Extra Examples** for *disabled* in Figure 1). The **Extra Examples** column is presented following the sense it illustrates.

CALD allows users to choose a display of **More examples** or **Fewer examples** following the examples provided within the core entry. An example is the entry for *disabled* shown in Figure 2. There is only one example nested within the core entry. However, if users want to see more examples, they can click on

the **More examples** link, and more examples for the word will be displayed (see Figure 3). Although CALD does not provide any information on the site about the selection of illustrative examples, Wendalyn Nichols, Publishing Manager of the Cambridge Dictionary, confirms that the examples within the core entry in CALD are all chosen by lexicographers in the process of compiling the entries according to the usual principles for learner's dictionaries. The "more examples" are also chosen by lexicographers from the *Cambridge International Corpus* using Sketch Engine (personal communication, March 28, 2025).



— Extra Examples

- *We support disabled students in further and higher education.*
- *Jenny has cerebral palsy and, like most disabled children, she attends a mainstream school.*
- *My son is disabled and needs extra support at school.*
- *The plant employs many disabled workers.*
- *The accident left him badly disabled.*

Figure 1: Extra examples for the entry *disabled* in OALD (Accessed 3 June 2025)



disabled
adjective
UK  /dɪˈseɪbld/ US 

B1

having an illness, injury, or condition that makes it difficult to do the things that other people do:

- *They are demanding equal rights for **the disabled**.*

+ More examples

Figure 2: The entry for *disabled* in CALD (Accessed 3 June 2025)

– Fewer examples

- *better access for the disabled*
- *This seat is reserved for elderly or disabled passengers.*
- *They have not been very responsive to the needs of disabled customers.*
- *The hotel has good facilities for the disabled.*
- *The company offers holidays for severely disabled people.*

Figure 3: Additional examples for the entry *disabled* in CALD (Accessed 3 June 2025)

Examples from the Corpus

disabled

- We will encourage the young to become involved and will facilitate access for the disabled.
- The governor has guaranteed health care for pregnant women, preschool children, and the disabled.
- About 70 percent of those elderly persons living with younger people are severely disabled.
- It's a challenge for the cast, some of whom are severely disabled.
- Nowadays he teaches them the cowardly art of aggressive selling to the elderly and disabled.
- There are many problems still to be overcome in providing for the disabled.
- Her son is disabled and she has to take care of him all the time.
- David goes to a special school for disabled children.
- The reason is that the opportunities for disabled drivers to test specially-adapted vehicles are limited.
- There was insufficient recognition that some of the voluntary organisations who helped with the plans do not adequately consult disabled people.
- But no-one was in doubt that the real winner was disabled sport.

Figure 4: Additional examples for the entry *disabled* in LDOCE6 (Accessed 3 June 2025)

LDOCE usually supplements its examples in the core entry with corpus examples, which are explicitly signalled by the boldface subheading **Examples from the Corpus** (for an example see Figure 4). LDOCE provides such additional examples for eleven terms surveyed in this study (*cripple*, *deaf mute*, *disabled*, *dumb*, *epileptic*, *handicapped*, *insane*, *invalid*, *mentally handicapped*, *psychopath*, and *visually impaired*). However, it is worth mentioning that LDOCE's corpus examples contravene principles of good lexicographic exemplification in several ways, leading to the suspicion that they are automatically generated, with little review from lexicographers. These deviations from good lexicographic exemplification will be examined in detail in the following sections.

Admittedly, the inclusion of extra examples in online dictionaries may benefit dictionary users, given the fact that "the average dictionary user appreciates examples in dictionaries" (Ptasznik 2023: 33). However, the presentation of additional examples in learner's dictionaries also raises all sorts of questions. While OALD and CALD provide additional examples under the senses they illustrate,

LDOCE, by contrast, presents the extra corpus examples at the end of the entry in a way that mixes examples of different senses, making it difficult for users to identify which examples illustrate a particular sense. In other words, users may be confused by the mix of examples for different senses. A more user-friendly approach to presenting corpus-derived examples might be to arrange them under the senses they illustrate or to group these examples according to the order of senses. This would enhance clarity and usability, helping learners better understand word meanings and usage in context.

Another problem is that the additional examples extracted from corpora may contain uncommon words, complex structures, idiomatic phrases, or cultural information which are difficult for learners to understand. For example, the following additional examples for *epileptic* and *invalid* are presented on the web page of LDOCE.

- (19) Thus **epileptic** seizures often began with a sustained tachycardia in spite of apnoeic pauses and severe hypoxaemia.
- (20) Ten patients had hypoxaemic events induced by **epileptic** seizures.
- (21) The survey attempted to answer critics who have dismissed international comparisons as **invalid** because of differences in cultural expectations about health care.
- (22) As it turns out, the fears that govern such organizations derive in large part from **invalid** or negative core beliefs.

The highly technical words, such as *tachycardia*, *apnoeic*, and *hypoxaemia*, in the examples for *epileptic* are difficult for foreign language learners — and maybe native speakers too — to understand. Moreover, the complex sentence structures adopted in these corpus examples, especially the example (21), may confuse learners rather than help them understand the meaning of the word. Such examples are largely useless, as Atkins and Rundell (2008: 461) argue that "if the example is incomprehensible, it is of no value".

However, a more serious problem is that some examples do not correspond to the senses they are intended to illustrate. For example, in examples (21) and (22), *invalid* is used in the general sense of "wrong" rather than in a context related to disability, and therefore does not represent a sensitive use. This is a common problem in e-lexicography (Frankenberg-Garcia et al. 2021).

A final deficiency of additional examples is that, without human intervention, they may contain grammatical errors, spelling errors, and other problems, as shown by the following example under *visually impaired* in LDOCE.

- (23) Instructions can also be obtained in Braille for the visually impaired.

The word "obtained" in this example is undoubtedly a misspelling: it should be "obtained". Such an example is unhelpful to dictionary users. It should be noted that all examples provided in dictionaries, whether selected by lexicographers or generated from a corpus, should serve the needs of dictionary users.

If learner's dictionaries choose to include additional examples, human proof-reading is desirable to ensure that poor-quality examples are eliminated. Admittedly, human intervention comes with financial costs. Given the current challenges facing dictionary publishers, such as increased competition from AI chatbots, declining advertising revenue, and the widespread expectation that dictionaries should be freely accessible, comprehensive manual oversight is often economically unfeasible. Nevertheless, intervention from lexicographers can undoubtedly improve the quality of corpus-derived examples, thereby benefiting foreign language learners.³

4. Exemplification of sensitive words for people with disabilities in MELDs: Main factors

Based on Stein (2002) and Xu (2008), I examined factors that may influence the exemplification of sensitive words for people with disabilities. The analysis indicates that the exemplification of sensitive words in the "Big Five" correlates with the parameters of part-of-speech, attitude labelling, and word currency. This section discusses how the exemplification of selected words varies with respect to these variables.

4.1 Effects of part-of-speech on exemplification

A preliminary analysis of the data in Table 1 reveals that the exemplification rates of sensitive words in the "Big Five" generally vary with the part-of-speech of the word. Adjectival senses of sensitive words are more likely to be exemplified than their nominal counterparts. For example, the adjectival senses of the words such as *epileptic*, *disabled*, and *insane* are consistently exemplified across all five dictionaries, whereas their nominal senses are either not included or not exemplified.

To quantify this exemplification trend, a comparative analysis was conducted (see Table 4). Across the dataset presented in Table 1, there are 14 adjectival terms and 14 nominal terms. The adjectival terms receive a total of 29 instances of exemplification (out of a possible 70), resulting in an average exemplification rate of 41%. In contrast, the nominal terms receive only 16 instances of exemplification (23%). This substantial difference suggests a consistent pattern favouring adjectives in the provision of usage examples.

Table 4: The exemplification rates of adj. and n.

POS	Total lexical items	Total exemplified instances	% Exemplified
adj.	14	29	41%
n.	14	16	23%

To determine whether this difference is statistically significant, a chi-square test

was performed. The result ($\chi^2 = 4.72, p < .05$) confirmed that the association between part-of-speech and exemplification is significant. In other words, the part-of-speech does have a significant effect on the exemplification of sensitive words in learner's dictionaries.

Despite this exemplification pattern, there are still some discrepancies in the exemplification of several words that can be used as both adjectives and nouns across the "Big Five" (e.g. *disabled*, *handicapped*, *hard of hearing*, and *insane*). The adjectival sense of *disabled* is exemplified in all five dictionaries, showing consistency in its treatment as an adjective. The nominal sense, however, is exemplified only in LDOCE and COBUILD. For *handicapped*, the adjectival sense is exemplified in OALD, LDOCE, and COBUILD but left unexemplified in CALD and MEDAL2. The nominal sense of the word is even less frequently exemplified. Only COBUILD provides a phrase example ("... measures to prevent discrimination against the handicapped."). As for *hard of hearing*, the adjectival use is exemplified in OALD, CALD, and COBUILD but unexemplified in LDOCE and MEDAL2. The nominal sense, however, is included in OALD and LDOCE, with illustrative examples provided only in OALD. The other dictionaries do not include the nominal usage at all. When it comes to the exemplification of *insane*, the adjectival sense is consistently exemplified in all five dictionaries. The nominal form, however, is inconsistently treated: it is illustrated by one example in OALD, LDOCE, and COBUILD, respectively, but not in CALD. MEDAL2 does not include this sense. In brief, COBUILD tends to exemplify the nominal uses of sensitive words more frequently than other dictionaries, while MEDAL2 exemplifies them the least.

Nevertheless, it is still safe to conclude that adjectival forms of sensitive words are more frequently exemplified than nominal ones. Such differences in the exemplification are largely influenced by the evolution of disability models and disability language. Nominal forms such as *a cripple*, *an epileptic*, or *a retard* are rooted in the moral model of disability, which views "disability [as] a defect caused by moral lapse or sins" (Olkin 1999: 25). The stigma-based language such as *a cripple* often reflects pity, stigma, or moral judgment (Olkin 2017). Moreover, the expressions like *the disabled* are favoured by the medical model of disability, which views disability as an inherent medical problem of the individual that must be "cured or eliminated" (Siebers 2008: 3). Using nominal forms with the definite article to group people based on certain characteristics like disability has been criticised for reducing individuals to a single attribute and failing to acknowledge their personhood (Halmari 2011; Dunn and Andrews 2015). The affirmative model represents a more recent and influential approach to understanding disability. It emphasises the positive identity and experiences of disabled individuals. Proponents of the affirmative model advocated for the identity-first language (e.g., *blind person* and *disabled person*), acknowledging disability as an integral part of identity, similar to race or gender. To recap, developments in both disability models and disability language offer insight into why adjectival forms of sensitive terms are more frequently exemplified than their nominal counterparts in learner's dictionaries.

4.2 Effects of attitude labelling on exemplification

In order to investigate the effects of attitude labelling on the exemplification of sensitive words for people with disabilities in the "Big Five", I examined the exemplification of the words labelled as having negative connotations by means of usage labels (e.g. "disapproval", "disapproving", "impolite", "not polite", "offensive", "rude", and "taboo") and other devices like definitions and usage notes. The results are tabulated in Table 5. In the Table, "L." means labelling, whereas "Ex" indicates exemplification. In the "L." column, a plus sign means that the word's offensiveness is marked by an attitude label in the dictionary. A minus sign shows that no label is provided for the word. A minus sign placed within brackets indicates that the word's offensiveness is marked by other devices such as definitions or usage notes. A zero means that the word or the relevant disability sense is not included in the dictionary.

Table 5: Effects of attitude labelling on exemplification

Lexical item	POS	OALD		LDOCE		CALD		COBUILD		MEDAL2	
		L.	Ex.	L.	Ex.	L.	Ex.	L.	Ex.	L.	Ex.
cripple	n.	+	-	(-)	-	+	-	+	-	+	-
deaf mute	n.	+	-	+	-	-	-	+	-	(-)	-
developmentally disabled	adj.	0	0	0	0	-	+	0	0	0	0
differently abled	adj.	0	0	0	0	-	+	0	0	0	0
disabled	adj.	-	+	-	+	-	+	-	+	-	+
	n.	-	+	-	+	+	-	(-)	+	(-)	-
dumb	adj.	+	+	(-)	-	-	+	+	+	(-)	-
epileptic	adj.	-	+	-	+	-	+	-	+	-	+
	n.	-	-	-	-	-	-	0	0	-	-
handicapped	adj.	+	+	(-)	+	+	-	+	+	(-)	-
	n.	+	-	(-)	-	+	-	+	+	(-)	-
hard of hearing	adj.	-	+	-	-	-	+	-	+	-	-
	n.	-	+	-	-	0	0	0	0	0	0
insane	adj.	+	+	-	+	+	+	-	+	-	+
	n.	+	+	-	+	-	-	(-)	+	0	0
invalid	n.	-	+	-	+	-	+	-	+	-	+
mentally handicapped	adj.	+	-	-	-	0	0	0	0	-	-
mentally ill	adj.	0	0	-	-	0	0	0	0	0	0
midget	n.	+	-	+	-	+	-	+	-	+	-
neurodivergent	adj.	-	+	0	0	-	+	-	-	0	0
psychopath	n.	-	-	-	-	-	-	-	+	-	-
retard	n.	0	0	0	0	+	-	+	+	+	-
retarded	adj.	+	-	(-)	-	+	-	+	-	(-)	-
spastic	adj.	+	-	-	-	+	-	-	-	+	-
	n.	+	-	-	-	+	-	+	-	0	0
spaz	n.	0	0	0	0	+	-	0	0	0	0
visually impaired	adj.	0	0	-	-	-	+	0	0	0	0
	n.	0	0	0	0	(-)	+	0	0	0	0

As Table 5 indicates, the "Big Five" show discrepancies in their exemplification of attitude-labelled sensitive words. Among them, CALD, LDOCE, and MEDAL take a more consistent approach by generally refraining from providing examples for words labelled as having negative connotations. COBUILD exemplifies more words in general, with only six of the words lacking examples: *cripple*, *deaf mute*, *midget*, *neurodivergent*, *retarded*, and *spastic*. In contrast, OALD is the most inconsistent, which makes it difficult to identify a clear principle behind its exemplification choices. Some words labelled as having negative connotations, like *dumb* (in its "unable to speak" sense), are accompanied by illustrative examples, while others (e.g. *cripple* and *deaf mute*) are not.

Notwithstanding these differences, the "Big Five" tend to adopt a cautious and selective exemplification of attitude-labelled words. To illustrate this, I tabulated the exemplification rates of attitude-labelled words and unlabelled words in Table 6.

Table 6: Exemplification rates of attitude-labelled words and unlabelled words

	Exemplified	Not exemplified	Total	% Exemplified
Labelled words	13	39	52	25%
Unlabelled words	32	21	53	60%
Total	45	60	105	43%

Based on 105 valid instances where the relevant disability-related sense of a word is included, attitude-labelled words have a significantly lower exemplification rate (25%) than unlabelled words (60%). A chi-square test of independence confirmed that this difference is statistically significant ($\chi^2(1, N = 105) = 13.41, p < .001$). This suggests that the attitude-labelled words are less likely to be exemplified in the "Big Five" than unlabelled words.

4.3 Effects of word currency on exemplification

Following Stein (2002) and Xu (2008), I also examined the effects of word currency on the exemplification of sensitive words. I surveyed the example allocation of sensitive words labelled as 'old-fashioned' (Lo.). Table 7 presents data on whether each sensitive word is marked as 'old-fashioned' (Lo=+) and whether it is accompanied by an example sentence (Ex=+). The results of the exemplification rates of currency-labelled words and unlabelled words are listed in Table 8.

It can be seen in Table 7 that the "Big Five" tend to avoid exemplifying highly offensive terms that are now outdated (e.g. *cripple*, *deaf mute*, *dumb*, *midget*, *mentally handicapped*, *retarded*, and *spastic*). By contrast, sensitive words that remain in medical, legal, or formal use (e.g. *invalid* and *insane*) are generally exemplified in MELDs. Among the 105 dictionary entries included in the analysis, 38 are

labelled as old-fashioned, while 67 are unlabelled. Of the labelled entries, only 10 (26%) are accompanied by examples, compared to 35 (52%) of the unlabelled entries. These descriptive results suggest that sensitive words labelled as old-fashioned are less likely to be exemplified than unlabelled words.

To determine whether this observed difference was statistically significant, a chi-square test of independence was conducted. The results revealed a statistically significant association between currency labelling and exemplification rates, $\chi^2(1, N = 105) = 5.64, p < .05$. This indicates that unlabelled words, i.e. more current in usage, are significantly more likely to be provided with examples in learner's dictionaries than words explicitly labelled as old-fashioned.

Table 7: Effects of currency labelling on exemplification

Lexical item	POS	OALD		LDOCE		CALD		COBUILD		MEDAL2	
		Lo.	Ex.	Lo.	Ex.	Lo.	Ex.	L0.	Ex.	L.	Ex.
cripple	n.	+	-	+	-	+	-	-	-	-	-
deaf mute	n.	+	-	+	-	+	-	+	-	+	-
developmentally disabled	adj.	0	0	0	0	-	+	0	0	0	0
differently abled disabled	adj.	0	0	0	0	-	+	0	0	0	0
disabled	adj.	-	+	-	+	-	+	-	+	-	+
	n.	+	+	-	+	-	-	-	+	-	-
dumb	adj.	+	+	+	-	+	+	-	+	+	-
epileptic	adj.	-	+	-	+	-	+	-	+	-	+
	n.	-	-	-	-	-	-	0	0	-	-
handicapped	adj.	+	+	+	+	+	-	-	+	+	-
	n.	+	-	+	-	+	-	-	+	-	-
hard of hearing	adj.	-	+	-	-	-	+	-	+	-	-
	n.	-	+	-	-	0	0	0	0	0	0
insane	adj.	+	+	-	+	+	+	-	+	+	+
	n.	+	+	-	+	+	-	-	+	0	0
invalid	n.	-	+	-	+	+	+	-	+	-	+
mentally handicapped	adj.	+	-	+	-	0	0	0	0	-	-
mentally ill	adj.	0	0	-	-	0	0	0	0	0	0
midget	n.	-	-	-	-	-	-	-	-	-	-
neurodivergent	adj.	-	+	0	0	-	+	-	-	0	0
psychopath	n.	-	-	-	-	-	-	-	+	-	-
retard	n.	0	0	0	0	-	-	-	+	-	-
retarded	adj.	+	-	+	-	+	-	+	-	+	-
spastic	adj.	+	-	+	-	+	-	-	-	-	-
	n.	+	-	-	-	+	-	-	-	0	0
spaz	n.	0	0	0	0	-	-	0	0	0	0
visually	adj.	0	0	-	-	-	+	0	0	0	0
impaired	n.	0	0	0	0	-	+	0	0	0	0

Table 8: Exemplification rates of currency-labelled words and unlabelled words

	Exemplified	Not Exemplified	Total	% Exemplified
Labelled words	10	28	38	26%
Unlabelled words	35	32	67	52%
Total	45	60	105	43%

5. Exemplification of sensitive words for people with disabilities in online MELDs: Some suggestions

Given the importance of examples in learner's dictionaries and the particular nature of sensitive words, improving their exemplification in MELDs is essential. This section thus provides some suggestions on the exemplification of sensitive words in online learner's dictionaries.

5.1 Increasing the exemplification of sensitive words

It may be difficult for learners to grasp the pragmatics of sensitive words without the help of examples. They may run the risk of using these words inappropriately, thereby causing unintended offence to others. It is thus reasonable to argue that learner's dictionaries could exemplify a greater number of sensitive words and increase the number of examples per sense. Traditional print dictionaries, due to space constraints, often grapple with the question of what types of vocabulary should be prioritised in exemplification (Xu 2008). Sensitive words, especially offensive words or taboo words, are seldom exemplified in print dictionaries (for example, in MEDAL2). However, in an era of e-lexicography, storage space is no longer a major problem for lexicographers. Lexicographers should now focus on what types of examples should be provided for different words and how many examples are needed to benefit dictionary users.

Current learner's dictionaries usually provide one or two examples per sense for sensitive words. However, numerous dictionary user studies (e.g. Summers 1988; Laufer 1993; Nesi 1996; Al-Ajmi 2008; Frankenberg-Garcia 2015) have demonstrated the effects of dictionary examples in language learning. Some studies (Frankenberg-Garcia 2014; Ptasznik 2023) show that more examples may benefit dictionary users. Zgusta (1971: 264) states outright that "exemplification is always useful". I would thus argue for an increase in the number of examples per sense for sensitive words (especially neutral and euphemistic terms) in learner's dictionaries.

While storage space is no longer a major limitation in online dictionaries, presentation space remains restricted (Lew in press).⁴ Providing too many examples within a single entry may lead to information overload, potentially hindering rather than helping users (Lew in press; Gouws and Tarp 2017). Therefore, examples for sensitive words in online learner's dictionaries should be selectively

and purposefully chosen to enhance understanding and avoid the dangers of information overload.

5.2 Improving the quality of additional examples

Considering the deficiencies in the additional examples in online learner's dictionaries, there is a clear need to improve the quality of such examples. One efficient solution could be the use of example extraction tools to identify and select good corpus examples. One of the most widely used tools is GDEX (Good Dictionary Examples; Kilgarriff et al. 2008), which was originally developed and used for the *Macmillan English Dictionary* in 2007 (Kilgarriff et al. 2008; Kosem et al. 2019). Significantly, GDEX can be used to filter out examples that do not meet specific standards. For example, it can control sentence length and word frequencies (Kilgarriff et al. 2008). Therefore, learner's dictionaries could use such example extraction tools to select high-quality corpus examples for sensitive words. For instance, examples could be limited to sentences using the defining vocabulary, which would help avoid the use of uncommon words in examples. Furthermore, human proofreading should be involved to check the quality of corpus-derived examples. This would mostly eliminate typos like *obtained* mentioned above. To sum up, a combination of human intervention and the use of example extraction tools would largely prevent the problems associated with additional examples in current learner's dictionaries and provide real benefit to users.

5.3 Optimising the presentation of additional examples

As discussed earlier, online learner's dictionaries like LDOCE place additional examples at the end of the entry, where examples of different senses are mixed together. This undifferentiated presentation makes it difficult for users, particularly EFL learners, to identify the particular senses different examples exemplify. Yamada (2013: 223) argues that the improvements of information provision in online dictionaries should focus more on "the order, arrangement, hierarchy, and presentation with differentiation and options". It is therefore argued that the presentation of additional examples in learner's dictionaries should be optimised. One possible approach, if the dictionary prefers to group extra examples together as in LDOCE, is to arrange them in the order of the senses they exemplify. Alternatively, as seen in OALD and CALD, the additional examples could be placed immediately after the sense they are intended to illustrate. Furthermore, in the digital environment, dictionaries can implement interactive features that allow users to filter or highlight examples by sense. For example, users could click on a specific sense in the definition area to view only the examples linked to that meaning.

6. Conclusion

This paper has examined the exemplification of sensitive words for people with disabilities in the "Big Five". While online learner's dictionaries tend to provide more examples overall, they remain conservative in exemplifying sensitive medical words. The dictionaries vary in the number of terms exemplified and examples provided, and in the inclusion of additional examples. Despite these differences, the "Big Five" display notable similarities in the types and functions of examples for the words surveyed: all offer phrase and sentence examples that support both the encoding and decoding needs of users. Furthermore, the findings indicate that the exemplification of sensitive words related to disability is largely influenced by three key factors: part-of-speech, attitude labelling, and word currency. First, with regard to part-of-speech, adjectival senses of sensitive words are more likely to be exemplified than their nominal counterparts. Second, sensitive words that remain in current use tend to have higher exemplification rates than those considered old-fashioned. Finally, words that carry negative connotations, typically marked by attitude labels, are less likely to be exemplified than neutral or euphemistic expressions. After a critical examination of the exemplification policies adopted by the "Big Five", this study argues for improving the quality and optimising the presentation of additional examples in online learner's dictionaries. More importantly, it recommends more frequent exemplification of sensitive words, especially neutral and euphemistic terms. As for offensive words, whether they should be exemplified may require further user research.

This study offers a preliminary exploration of the exemplification of sensitive medical words in dictionaries. There are numerous avenues for further research. A survey of users' interpretations of the examples provided for sensitive words in dictionaries would be particularly valuable. Such studies would enable lexicographers to better understand how different users, particularly users with disabilities, interpret and evaluate the lexicographic treatment of sensitive medical words. Besides user research, an investigation into the representation of people with disabilities through sentence examples could also be pursued in future studies. There is also a need for research that could have a greater concern for the rights and experiences of disabled people in lexicography. Such research would contribute to our understanding of disability in lexicographic discourse and promote inclusive language.

Acknowledgements

The author would like to thank Prof. Elsabé Taljard, and three anonymous adjudicators for their insightful comments and suggestions. Special thanks are due to Prof. M. Lynne Murphy of the University of Sussex for her constructive feedback on earlier drafts of this article. This study was inspired by academic discussions with her.

Endnotes

1. The website of the *Macmillan English Dictionary for Advanced Learners* was shut down on 30 June 2023.
2. I am grateful to one anonymous adjudicator for reminding me that MEDAL has fewer examples for words in general, not just sensitive words.
3. Thanks are due to one anonymous adjudicator who called my attention to this point.
4. Lew (in press) explains that "*storage space* refers to the capacity to hold the total content of the dictionary, whereas *presentation space* refers to how much can be presented (displayed, visualized) at a given time to the dictionary user."

References

A. Dictionaries

- CALD.** *Cambridge Advanced Learner's Dictionary*.
<https://dictionary.cambridge.org> [3 June 2025]
- COBUILD.** *Collins COBUILD Advanced Learner's Dictionary*.
<http://www.collinsdictionary.com> [3 June 2025]
- LDOCE.** *Longman Dictionary of Contemporary English*.
<https://www.ldoceonline.com> [3 June 2025]
- MEDAL2.** 2007. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan Education.
- OALD.** *Oxford Advanced Learner's Dictionary of Current English*.
<https://www.oxfordlearnersdictionaries.com> [3 June 2025]

B. Other literature

- Al-Ajmi, H.** 2008. The Effectiveness of Dictionary Examples in Decoding: The Case of Kuwaiti Learners of English. *Lexikos* 18: 15-26.
- Allan, K. and K. Burridge.** 2006. *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.
- Anderson, R.C. and R.A. Fox.** 1988. Health Education and the Developmentally Disabled. *Journal of the Multihandicapped Person* 1(4): 327-333.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- BNC.** *British National Corpus*. <http://www.natcorp.ox.ac.uk/> [19 June 2025].
- Coffey, S.** 2010. 'Offensive Items', and Less Offensive Alternatives, in English Monolingual Learners' Dictionaries. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress, Leeuwarden, 6–10 July 2010*: 1270-1281. Leeuwarden/Ljouwert: Fryske Akademy — Afûk.
- Dunn, D.S. and E.E. Andrews.** 2015. Person-first and Identity-first Language: Developing Psychologists' Cultural Competence Using Disability Language. *American Psychologist* 70(3): 255-264.

-
- Farina, D.M.T.Cr., M. Vrbinc and A. Vrbinc.** 2019. Problems in Online Dictionary Use for Advanced Slovenian Learners of English. *International Journal of Lexicography* 32(4): 458-479.
- Fox, G.** 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*: 137-149. London: HarperCollins.
- Frankenberg-Garcia, A.** 2012. Learners' Use of Corpus Examples. *International Journal of Lexicography* 25(3): 273-296.
- Frankenberg-Garcia, A.** 2014. The Use of Corpus Examples for Language Comprehension and Production. *ReCALL* 26(2): 128-146.
- Frankenberg-Garcia, A.** 2015. Dictionaries and Encoding Examples to Support Language Production. *International Journal of Lexicography* 28(4): 490-512.
- Frankenberg-Garcia, A., G.P. Rees and R. Lew.** 2021. Slipping Through the Cracks in e-Lexicography. *International Journal of Lexicography* 34(2): 206-234.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Halmari, H.** 2011. Political Correctness, Euphemism, and Language Change: The Case of 'People First'. *Journal of Pragmatics* 43(3): 828-840.
- Hughes, G.** 2006. *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-speaking World*. New York: M.E. Sharpe.
- Humblé, P.** 2001. *Dictionaries and Language Learners*. Frankfurt am Main: Haag und Herchen.
- Kilgariff, A., M. Husák, K. McAdam, M. Rundell and P. Rychlý.** 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the 13th EURALEX International Congress, Barcelona, 15-19 July 2008*: 425-432. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Kosem, I., K. Koppel, T.Z. Kuhn, J. Michelfeit and C. Tiberius.** 2019. Identification and Automatic Extraction of Good Dictionary Examples: The Case(s) of GDEX. *International Journal of Lexicography* 32(2): 119-137.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.
- Laufer, B.** 1993. The Effect of Dictionary Definitions and Examples on the Use and Comprehension of New L2 words. *Cahiers de Lexicologie* 63(2): 131-142.
- Lew, R.** in press. Space Restrictions in Paper and Electronic Dictionaries and Their Implications for the Design of Production Dictionaries.
https://wa.amu.edu.pl/~rlew/pub/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf [19 June 2025]
- Nesi, H.** 1996. The Role of Illustrative Examples in Productive Dictionary Use. *Dictionaries: Journal of the Dictionary Society of North America* 17: 198-206.
- Norri, J.** 2000. Labelling of Derogatory Words in Some British and American Dictionaries. *International Journal of Lexicography* 13(2): 71-106.
- Norri, J.** 2020. Treatment of Words for Illness and Disability in Monolingual English Dictionaries. *International Journal of Lexicography* 33(3): 227-250.
- Olkin, R.** 1999. *What Psychotherapists Should Know about Disability*. New York: Guilford Press.
- Olkin, R.** 2017. *Disability-affirmative Therapy: A Case Formulation Template for Clients with Disabilities*. New York: Oxford University Press.
- Ptasznik, B.** 2023. More Examples May Benefit Dictionary Users. *International Journal of Lexicography* 36(1): 29-55.

- Rees, G.P.** 2024. Academic Word Families in Online English Dictionaries. *Lexikos* 34(1): 437-468.
- Rundell, M.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25(1): 301-322.
- Siebers, T.** 2008. *Disability Theory*. Ann Arbor, MI: University of Michigan Press.
- Stein, G.** 2002. *Better Words: Evaluating EFL Dictionaries*. Exeter: Exeter University Press.
- Summers, D.** 1988. The Role of Dictionaries in Language Learning. Carter, R. and M. McCarthy (Eds.). 1988. *Vocabulary and Language Teaching*: 111-125. London: Longman.
- Van der Meer, G.** 2005. On Taboo and Slang in English Pedagogical Lexicography. So You've Been Looking for Them, Have You, Madam? (Dr Johnson). Gottlieb, H., J.E. Mogensen and A. Zettersten (Eds.). 2005. *Symposium on Lexicography XI: Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*: 381-390. Berlin/Boston: Max Niemeyer.
- World Health Organization (WHO).** 2013. *How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO.
- Xu, H.** 2008. Exemplification Policy in English Learners' Dictionaries. *International Journal of Lexicography* 21(4): 395-417.
- Yamada, S.** 2013. A Test of the Proposed Framework for Reviewing Online Dictionaries: M-W.com, Dictionary.com, Macmillandictionary.com, Dictionary.cambridge.org, and Oxforddictionaries.com (sic). *Dictionaries: Journal of the Dictionary Society of North America* 34: 211-224.
- Zgusta, L.** 1971. *Manual of Lexicography*. Berlin/New York: De Gruyter Mouton.

Lexicography in Action: The Traversal from Coinage and Iconicity to Iconisation

Dorra Moalla Masmoudi, *LAD Laboratory, Faculty of Arts and
Humanities, University of Sfax, Sfax, Tunisia*
(dorra.moala@fshs.usf.tn)
(<https://orcid.org/0000-0001-8055-2400>)

Abstract: This paper seeks to connect Lexicography and Critical Discourse Analysis (CDA) by highlighting that the critical potential of Lexicography lies in the coinage of dictionary entries, which can function as *semiotic acts* that aim to rally the readers around the sign-maker's world view. To illustrate this, GOT ISRAELED, a new entry in the crowdsourced *Urban Dictionary* added on 21 October 2023, is taken as an example. This entry is defined as the act of allowing someone to share something with you but this person claims the thing as his/her own and expels you. The entry has gone viral on social media and induced divergent attitudes. Its wide reach is due to its *iconic* nature and coinage during the war on Gaza. The entry has also been heavily resemiotised inside *Urban Dictionary* and its critical potential has gradually increased clearly recontextualizing the colonization of Palestine and the issue of Jewish settlements. 36 entries added between 21 October 2023 and 10 July 2024 are analyzed at three levels: *coinage*, *iconicity* and *iconisation*. The analysis draws upon Semiotic and Social Semiotic frameworks, which include DeSaussure's (1916) *sign system*, the Social Semiotic principle of *motivated sign* (Kress 1993), and Peirce's (1931) principles of *hypo-iconicity*. An SFL-based analysis exploiting subsystems within the *ideational* and *interpersonal* metafunctions is employed to explain the sign's 'iconisation' (Halliday and Matthiessen 2014, Kress and Van Leeuwen 1996/2021, Martin and White 2005). The investigation simultaneously highlights the critical potential of the sign and describes its transformation into a *bonding icon* rallying people around the writers' world view (Stenglin 2008, 2012).

Keywords: LEXICOGRAPHY IN ACTION, COINAGE, ICONICITY, ICONISATION, URBAN DICTIONARY, MOTIVATED SIGN

Opsomming: Die leksikografie in aksie: Die oorgang van nuutskepping en ikonisiteit na ikonisering. In hierdie artikel word gepoog om die Leksikografie en Kritiese Diskoersanalise (KDA) met mekaar te verbind. Dit word gedoen deur te beklemtoon dat die kritiese potensiaal van die Leksikografie geleë is in die vaslegging van woordeboekinskrywings wat as *semiotiese handelinge* kan funksioneer. Laasgenoemde is daarop gemik om die lesers te laat saamstem met die sender van die teken se wêreldbeskouing. Ter illustrasie hiervan word "GOT ISRAELED", wat op 21 Oktober 2023 as nuwe inskrywing in die skarebenutte *Urban Dictionary* bygevoeg is, as voorbeeld gebruik. Hierdie inskrywing word gedefinieer as die handeling om iemand toe te laat om iets met jou te deel, maar hierdie persoon eis dan die item as sy/haar eie op en verdryf jou. Die inskrywing het vinnig op sosiale media versprei en uiteenlopende standpunte ontlok. Die breë reikwydte daarvan

is toe te skryf aan die *ikoniese* aard en vaslegging daarvan tydens die oorlog teen Gaza. Die inskrywing is ook in die *Urban Dictionary* grootliks gehersemiotiseer en die kritiese potensiaal daarvan het geleidelik toegeneem om sodoende die kolonisasie van Palestina en die kwessie van Joodse nedersettings duidelik te herkontekstualiseer. 36 inskrywings wat tussen 21 Oktober 2023 en 10 Julie 2024 bygevoeg is, word op drie vlakke geanaliseer: *nuutskepping*, *ikonisiteit* en *ikonisering*. Die analise steun op Semiotiese en Sosiaal-Semiotiese raamwerke, wat DeSaussure (1916) se *tekensisteem*, die Sosiaal-Semiotiese beginsel van *gemotiveerde teken* (Kress 1993), en Peirce (1931) se beginsels van *hipo-ikonisiteit* insluit. 'n SFL-gebaseerde analise wat substelsels binne die *ideasionele* en *interpersoonlike* metafunksies benut, word gebruik om die 'ikonisering' van die teken te verduidelik (Halliday en Matthiessen 2014, Kress en Van Leeuwen 1996/2021, Martin en White 2005). Die ondersoek beklemtoon tegelykertyd die kritiese potensiaal van die teken en beskryf die transformasie daarvan in 'n *saambindende ikoon* wat mense rondom die skrywers se beskouings van die wêreld verenig (Stenglin 2008, 2012).

Sleutelwoorde: DIE LEKSIKOGRAFIE IN AKSIE, NUUTSKEPPING, IKONISITEIT, IKONISERING, *URBAN DICTIONARY*, GEMOTIVEERDE TEKEN

1. Introduction

Lexicography is commonly known as the scholarly field involving compiling dictionaries and editing them (Bergenholtz and Gouws 2012). Mainstream Lexicography prioritizes the objective linguistic descriptions of dictionaries as "objective value-free representations of languages" (Chen 2015: 311) and attributes little attention to the contextualization of dictionary-making. Critical Lexicography, a recent research direction, addresses this gap and showcases that dictionaries are influential ideological texts. Critical Lexicography meshes with CDA, and views language use as a form of social practice conditioned by the context of use and the dominant ideologies (Fairclough 1995). Dictionaries are, therefore, ideological sites that can be understood in terms of the context of use and power dynamics (Bertulessi 2023, Chen 2015, 2019, Fuertes-Olivera and Tarp 2022, Hornscheidt 2008, Pettini 2023). Chen (2015: 362) concentrates on the link between dictionary structures and entries and the external factors of language use stating that "lexicography and its products, dictionaries, are never value-free, apolitical or asocial. Instead, they are subject to ideology, power and politics".

Dictionary writing as "a literate writing activity" is situated within a process of social recontextualization and meaning relocation and "is inevitably bound up with subjectivity, re-representation, objectivity and ideology" (Chen 2015: 312). This entails that dictionary-making is a floating domain that shifts from context to context, practice to another and from sign-maker to sign-maker. According to Linell (1998), recontextualization in Lexicography is concretely realized across three dimensions: intra-textual recontextualization which happens within the same dictionary, intertextual recontextualization which concerns the shifts in meaning-making of the same types of texts situated within different contexts (for example two dictionaries defining the same entry) and interdiscursive recontextualization which occurs across different media and genres (for instance, a dic-

tionary entry and an argumentative text dealing with a dictionary entry). Intra-textual recontextualization is the focus of this paper, a recontextualization afforded by the participatory nature of online crowdsourced dictionaries.

The current literature on Critical Lexicography is centered upon deciphering the ideology related to gender, ethnicity and politics. By means of examining ideology-loaded entries, research studies point out that dictionaries recontextualize ethnic and gender bias in societies (Chen 2015, Fuertes-Olivera and Tarp 2022). These studies have emphasized that ideology is articulated through the selective techniques of the dictionary-maker, which include deleting, replacing, re-signifying ... etc. (Chen 2015). With regards to political ideology, studies have demonstrated that projects of dictionary compilation expose the political views of the dominant political groups (Bertulesi 2023) and perpetuate ideologies of racism and colonialism (Hornscheidt 2008).

The study at hand aligns with Critical Lexicography through showing that ideology stands as a mediating force for the coinage of new signs or the *mutability* of existing ones. The existing literature on Critical Lexicography has adopted a critical stance towards the work of ideology in dictionary-making and has argued that the role of scholars engaging with (Critical) Lexicography research is to uncover the opaque techniques of power dynamics (Chen 2019, Kachru 1995). Though this study acknowledges that Lexicography is a site of ideology encoding, it maintains that dictionaries can also be sites for *semiotic acts*, giving voice to the marginalized and oppressed groups. Chen (2019) articulates the dual work of dictionaries as domains of power and resistance. Within the latter, regional dictionaries, and dictionaries for minority groups can disrupt the hegemony of dominant dictionaries. Resistance Lexicography can be a promising direction of research within Critical Lexicography. To the best of my knowledge, there is a scarcity of substantive research on uncovering the dynamics of meaning-making in *lexicography in action*. This paper aims to address this gap. To do so, it offers to analyze 36 new entries of the lexical item GOT ISRAELED, a new coinage in the online *Urban Dictionary* that first appeared on 21 October 2023 and was widely resemiotised within the dictionary and in different generic and mediated forms outside it. This paper argues that these entries are semiotic acts introducing, in an escalating way, the agency and the subjective views of the compilers, who aim to turn the new entry into a semiotic force that recontextualizes what they consider as an unrightful colonization of Palestine. Within a period of two weeks, the entry had proliferated and shifted, in an incremental way, from an almost apolitical sign to a politicized one overtly recontextualizing the Palestinian cause through pointing to hot political issues, like the Jewish settlements and the religious grounding of the colonization of Palestine. The critical potential of these entries derives from the *iconic* nature of the sign. Studying the critical force of iconic signs based on the resemblance between form and meaning remains underexamined in the literature.

Methodologically speaking, the current studies on Critical Lexicography have pointed out that power manifestations in lexicography are mediated through the selective techniques of censoring, altering, replacing and redesigning (Chen 2019).

This study, however, contends that lexicography, as a semiotic act, exploits techniques of attitude construction and expansion. The compiler's role is to use the discourse resources to rally the reader around his/her world view. In the sign under examination, I propose to delve into how the writers of the different dictionary entries exploit the semiotic principles of *motivated sign*, *iconicity* and *iconisation* to produce an effect upon the reader.

To decipher the dynamics of attitude amassing in the dictionary entries under examination, three levels of analysis are adopted:

- The coinage of the sign
- Its iconicity
- Its iconisation

The first level draws upon De Saussure's (1916) principles of the sign system to highlight the innovative relation between *signifier* and *signified* and it explains, through the principle of the im/mutability of the sign system, the social forces triggering its mutability. Along with this, the analysis adopts a Social Semiotic perspective to underscore the motivated relation between signifier and signified in the coinage of the present sign (Kress 1993).

The second level exploits Peirce's (1931) typology of signs to explore the semiotic relation foregrounded in this iconic sign. By using Peirce's principles of *hypo-iconicity*, this analysis underscores how the principles of *direct resemblance* and *metaphoric distortion* in hypo-iconicity build up direct connection with reality and foreground the critical potential of the sign.

The third level shows how the entry has evolved from iconicity to a *bonding icon* (Stenglin 2022: 314). To highlight this, the analysis draws upon the principles of iconisation, an evolving research direction in Social Semiotics (Stenglin 2008; 2012; Martin 2010). The main tenets of iconisation are that signs turn into bonding icons when they rally people around a set of values (Stenglin 2022). This paper posits that the entry under examination has evolved into a bonding icon through amassing attitudes uniting internet users, though in divergent ways, around the sign. This is translated in its wide resemiotisation inside the dictionary and outside it in different mediated forms, its viral status, as well as the substantial positive and negative reactions the sign has triggered. In the process of its shift into a bonding icon, the entry realizes simultaneously an *ideational* role and an *interpersonal* one: it represents the world from the standpoint of the entry writer and rallies people around this world view. To highlight the interface between ideational and interpersonal meanings in the production of this sign, I adopt an analysis drawing upon the ideational and interpersonal metafunctions in SFL (Systemic Functional Linguistics) (Halliday and Matthiessen 2014, Martin and White 2005). A Systemic Linguistic analysis is undertaken building upon sub-systems within the ideational and interpersonal metafunctions, which include the systems of *represented participants*, *process types*, *participant roles* and *metaphoric distortion*, on the one hand and the systems of *interactive participants* and *graduation*, on the other. The different systems will be explained in details in section 6.

This introduction has argued that a critical reading of lexicography should be combined with the linguistic one to grasp the meaning potential of dictionary entries. Such a critical reading is especially relevant in lexical entries with an action force. Based on these premises, the paper is structured as follows: section 2 grounds the study within the research direction of language as a semiotic act, section 3 introduces the discoursal and sociopolitical context of the study. It presents the sociopolitical context of the sign coinage, introduces the dictionary under examination and narrows down the focus to the entries and their generic structure. Sections 4, 5 and 6 analyze the entries at the three levels mentioned above: coinage, iconicity and iconisation using tools from semiotics and SFL. Finally, in section 7, a framework for analyzing signs with an action force is introduced.

2. Semiosis in action

This study draws upon the principle of *semiosis as action* developed by Peirce (1907). By semiosis, Pierce (1907: 411), means "an action, or influence, which is, or involves a cooperation of three subjects, such as a sign, its object, and its interpretant". In this triadic relation, the interpretant's role is central and is understood as the thing that the sign triggers for the interpreter, which can be a feeling, an effect or a thought (Strand 2013). Peirce's semiosis constitutes a paradigm shift from DeSaussure's structure of sign, according to which a sign is inherently communicative as it is a pre-defined and fixed connection between a signifier and a signified (DeSaussure 1916). Peirce's semiotics foregrounds the signs as mediators to action. The meaning of signs is not fixed as the interpreter assigns meanings to them through processes of construction, reconstruction and validation (Strand 2013). The principle of interpreter's role in assigning meaning to signs is congruent with the principle of *reader agency* in Social Semiotics, which advocates that reading is a meaning-making action rather than a reception phenomenon (Kress 2010). The principle of reader agency is more evident in this media age, which affords a more participatory role to the reader (Moalla 2023).

The semiotic effect advocated by Peirce cannot be understood without reference to authorial agency. Social Semiotics describes the role of the designer as semiosis in action or semiosis at work, highlighting the agentive power of the designer in meaning-making and his/her interest in the process of design. Kress and Van Leeuwen (2001) equate semiosis in action with changes in representation, which are done through the action of individuals. This action cannot be dissociated from the macro-level changes in the social context. Kress and Van Leeuwen (2001: 36) describe the role of social forces in semiotic action as "all semiotic action is social: that social action changes the actor, the acted on or acted with".

To sum up, this paper suggests semiosis in action as a framework for understanding the dynamics of meaning-making in the coinage and interpretation of dictionary entries. This framework is aligned with Peirce's semiosis as action

along with the principles of CDA and Social Semiotics. It is understood at two levels. First, in line with CDA, signs are mediated forms that simultaneously recontextualize the world and reflect the authorial voices. Therefore, signs "shape how we think and act in the world" (Ledin and Machin 2018: 15). Second, signs are not fixed entities. They induce action that can be translated into a thought, a feeling or an effect. Of special interest here, is how signs trigger feelings and attitudes. This kind of action is aligned with the principle of iconisation in Social Semiotics, according to which signs align and unite people around a set of values, which are induced by the authorial discursal strategies.

3. Context of the study

Chen (2019) maintains that modern lexicography should be more critical through going beyond the text to the context of text production and its power dynamics. He proposes three phases of lexicographic analysis that include social analysis, discourse analysis and text analysis. This paper aligns with this methodology. This section proposes simultaneously to anchor the sign under examination within the sociopolitical context of its coinage and to ground it within the generic tradition of online dictionaries before undertaking a text analysis in the following sections.

The entry under examination was added to *Urban Dictionary*, an online crowdsourced dictionary about new lexis and slang words (Ro 2019). Internet users can add their lexical entries. After approval, the entry appears with the name of the writer and the date of entry. Internet users can upvote or downvote it and share it on social media platforms. From its creation by the computer science student Aaron Peckham as a humorous parody to scholarly dictionaries, *Urban Dictionary* has become very notorious with 65 million visitors every month. It is also widely used by linguists to keep record of slang words entering into the language (Ro 2019).

Urban Dictionary aligns with online dictionaries, which are increasingly appearing as alternative forms of dictionaries along with the traditional printed ones. Meta-lexicography research on online dictionaries has yielded divided views relative to their textual features and content. Advocates of this emerging genre argue that it allows selected searches of entries and rapid access to them (Kirkness 2004). At the content level, e-dictionaries, especially those treating slang words, can give voice to minority groups to counterbalance the dominance of standard language-based dictionaries (Coleman 2004). Pettini (2023) expands on this, suggesting that printed dictionaries, being produced by multinational publishing industries, reflect the interest of the dominant groups. Online dictionaries give voice to the oppressed and marginalized groups. Critical voices about slang online dictionaries emphasize the poor quality, and the inaccuracies of many of them (Liu 2015) and the lack of a conventionalized structure (Gouws 2014). Critics also contend that online dictionaries proliferate discourses of offensiveness and hate (Pettini 2023).

As to its generic structure, the homepage in *Urban Dictionary* includes four components: (1) the heading of the homepage, (2) a horizontal bar for the menu, (3) a search bar and (4) a right-hand side section for the coinage of new words including also small icons for social media sharing. The dictionary exploits the interactive affordance of the internet allowing users to exert different actions like clicking on the search bar, adding an entry, and sending a question (*Urban Dictionary* 2024). With reference to the expression under consideration, the search bar includes two entries, the verb and noun (Israel and Israeled) along with other compound nouns. The noun 'Israel' includes many entries, mostly coined in the years 2000s and more recent versions were added in November 2023. The definitions have clear political orientations and echo divergent views about the state/entity of Israel. On 21 October 2023, for the first time, a new entry in the verbal form, GOT ISRAELED, was added. This entry is defined as the act of allowing someone to share something with you but this person claims the thing as his/her own and expels you. The sign-maker takes the example of sharing a table in a restaurant with a stranger, who ends up expelling you pretending s/he has a meeting. Five entries were coined on the same day (21 October) by different writers with different pseudonyms. Thus, it is difficult to decide which of them is the original entry, especially that the interdiscursive recontextualization of the entry in other media and genres (online magazines and blogs) refer to one entry appearing on the 22nd as the original. For this reason, I will consider all the entries appearing on 21st as the original entries.

Starting from 22 October and during 9 months, the sign has gone viral on social media and has been widely resemiotised inside *Urban Dictionary*. The resemiotisations have shown, in an escalating way, more politicized definitions of the sign. The link of the sign with the colonization of Palestine becomes more direct. The wide reach of this entry and its viral status are mainly due to its iconic nature and its coinage against the backdrop of the Israeli war on Gaza, which may explain the increasing political orientation. The sign is also resemiotised outside *Urban Dictionary* in different mediated forms (blogs, videos, billboards, online journals) by advocates of the Palestinian cause as well as by defenders of Israel.

Grammatically speaking the verbal form of the entry is used in the passive voice 'be Israeled' in the simple, progressive and perfective aspects (11 entries), in the passive voice with the mutative 'get' as in GOT ISRAELED (16 entries) as well as in the active TO ISRAEL (11 entries). GOT ISRAELED is used generically to refer to the verbal realizations of this entry. Starting from May 30th, six different definitions of the verb had been coined, very probably by pro-Israeli writers, to counter the iconisation of the verb but the definition did not achieve the viral status. Since these entries have a different meaning, they are not within the scope of this study.

The generic structure of each entry aligns with the typical structure of dictionary entries outlined by Fuertes-Olivera and Tarp (2022). It includes a headword highlighting the verbal form of the entry although such grammatical information is not explicitly mentioned as in standardized printed and online diction-

aries. The head word is followed by a co-text, which is divided into a definition followed by a usage example in a concrete situation and ends with a sentence example. The generic structure exhibits a few specificities, which are congruent with the communicative function of crowdsourced dictionaries. As illustrated in the sample entry in Figure 1, the top right corner includes social media links to help the user share the entry. Below the sentence example, the pseudo name of the writer and the date of addition are included. At the bottom of the entry, two icons are provided allowing the user to react by upvoting, or downvoting.



Figure 1: A Sample of the dictionary entry ISRAELED included in <https://www.urbandictionary.com/define.php?term=Israeled>

4. Coinage: Mutability and motivation

The entry is considered as a sign that partly subscribes to the structure of the sign in DeSaussure's sign system (1916). A sign, in this framework, is an intrinsic, conventionalized, inherently communicative connection between a cluster of sounds, a signifier and a concept, a signified. The sign under examination is a combination of a signifier in the passive voice with a mutative 'got' (cf. Halliday and Matthiessen 2014: 349) and a signified, 'to be cheated' by someone or a group of people, who take something that used to belong to the victim.

GOT ISRAELED is directly associated with a previous sign in the nominal form (Israel), which is also a controversial sign within *Urban Dictionary*. DeSaussure refers to this shift as the mutability of signs that may affect the signifier, and/or signified and/or the relationship between them (DeSaussure 1916). Despite the phonological resemblance between this sign and the previous nominal form, the mutability occurs at the three levels as shown in Figure 2.

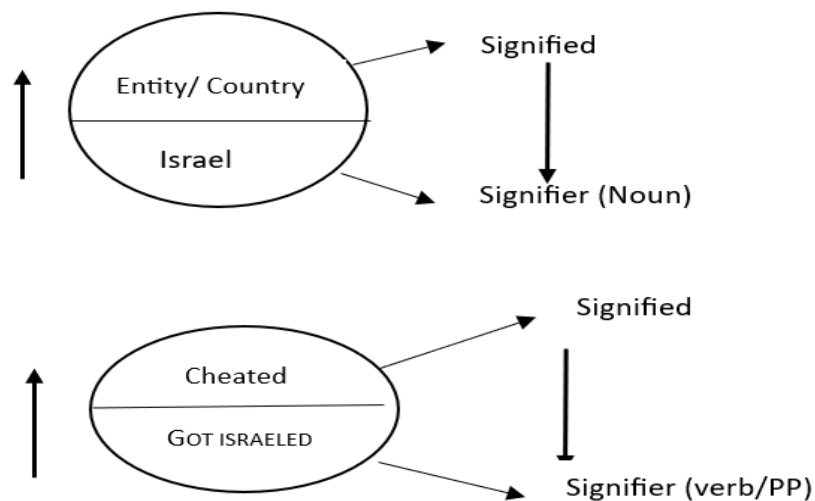


Figure 2: Shift in signifier-signified relationship in GOT ISRAELED

In DeSaussure's framework, language is an inherited immutable system with a fixed relationship between signifier and signified. The individual and linguistic community have little freedom to control or change it. He justifies this fixedness by the fact that "a particular language state is always the product of **historical forces** and these forces explain why the sign is unchangeable, i.e. why it resists any **arbitrary** substitution" (DeSaussure 1916: 72 emphases added). Despite the immutability of signs, DeSaussure argues that language is unable to resist changes at a certain point in time. He posits two factors for sign mutability, which are time and social force. Time allows social force to exert sign mutability but the shift is only possible by the agreement of the community of speakers. His theory provides manifestations of language change (signifier, signified and relationship between them) but its description of time and social force as contributors to sign mutability remains vague and not thoroughly described (Holdcroft 1991).

Many critics attribute DeSaussure's failure to account for language mutability to the fact that he describes the relationship between signifier and signified as arbitrary and unmotivated. On the contrary, Holdcroft (1991: 53) maintains that, beyond onomatopoeia, many signifiers are "appropriate for the signifieds", at different levels of appropriateness, which are divided into categorical and structural ones. Categorical appropriateness refers to the correspondence between grammatical categories and conceptual ones. For instance, nouns denote entities and verbs denote action. Structural appropriateness refers to the link between sentence structure and the structure of thought.

In the Social Semiotic Theory, this appropriateness is conceptualized within the notion of motivated sign (Hodge and Kress 1988, Kress 1993). In this theory, the motivated sign is the result of an interplay between the social context, the material features of signifiers and the individual's contribution to the production of signs (Björkvall and Karlsson 2011). The social/cultural dimension is the superordinate force in meaning-making that dictates social practices and forms of discourse. The making of meaning resides in the intersection between the agreed upon social meaning attributed to signs and the material features of signifiers (for instance, the grammatical category). In this interplay between the social and the structural, a motivated sign is generated. Grammatical and lexical categories and other semiotic modes are culturally established and materially appropriate to express certain meanings. Kress (1993) illustrates this with the example of the circles (signifier) being apt for car wheels (signified). Social practices execute a force in transforming the material features of signifiers into meaning. The individual sign-maker is aware of the fact that signs have a meaning potential dictated by culture/society and the material features of the sign, and exploits this connection to express the desired meanings. The signs produced are "motivated by the producer's interest and the characteristics of the objects" (Kress 1993: 173).

To illustrate this triadic relationship between the social, material and individual in the sign under examination, the sign makers' choice of the root 'Israel' is motivated by the socio-political context of the colonization of Palestine and its appropriateness is accentuated against the backdrop of the war on Gaza. The use of the signifier in the past participle is also materially appropriate to express the signified, the act of being cheated. This sign aligns with Kress's description of motivated sign as transparent in which the reader can easily see why the signifier GOT ISRAELED is used to express the signified. Accordingly, the writers of the entries have drawn upon the choices available in grammar and dictated by context to produce signs. Kress (1993: 174) describes this as a selection process of what to be presented and how to be presented through the use of apt signifiers, which should have "criterial characteristics in respect to the object to be presented, that is, it becomes, the signified". The selection process also reflects the sign maker's relation to an object or event at a very particular context of the situation to reflect his/her intentions, purposes, world views and political positions. The sign makers, who are obviously positioned in favor of the Palestinian cause, have exploited the criterial characteristics manifested in Jewish settlement and the genocidal acts against the people of Gaza to transform the name 'Israel' into a verb in the past participle form as an apt signifier for the signified 'be cheated'. In the selection of the apt signifier, a metaphorical relation is selected by the sign-maker, who believes that the metaphor has the appropriate characteristics for the signified (Kress 1993). Taub (2004: 9) illustrates this criterial aptitude in connection with the coinage of new signs deriving from existing signs. She argues that the use of the new signs reflects the users' beliefs about the original sign. She exemplifies that 'dogged persistence' reflects our attitude

towards 'dogs' as patent and persistent.

The concept of motivated sign, which underscores the interplay between the social anchoring of signs, the individual contribution in their production and the signs' material appropriateness explained above, can account for sign mutability. According to Kress and Van Leeuwen (2001), changes in the linguistic system like the mutability of signs are motivated by the larger scale social, economic and political practices and are situated within the domain of ideology. Sign mutability explains the ideological positioning of the sign-maker. The ideological embedding of sign mutability can be a concretization of DeSaussure's notion of social force. This force is, however, not necessarily extended over a long period of time. Kress and Van Leeuwen (2001: 36) correlate the force of change in society with the intensity of change in modes of representation saying that "the intensity of the dynamic for change in practices increases the intensity of the dynamic of change in the various articulatory modes, and in the discourses which they articulate". Seen from this perspective, we can say that changes of attitudes and growing empathy towards the Palestinian cause have induced this coinage.

This section draws upon the Social Semiotic principle of motivated sign according to which signs are motivated by social forces and the sign-maker's interest. The latter chooses the materially and socially appropriate signifiers to express the intended meaning (Van Leeuwen 2005). In this process, ideology stands as a mediating element that can explain the choice of signs and their mutability. From this perspective, there is no neutral arbitrary sign and all signs are subject to criticism. Therefore, CDA rests on the principle of motivated sign to undertake a critical reading of signs. (Kress 1993). The following section will highlight how the iconic nature of GOT ISRAELED facilitates its ideological work adopting Peirce's principle of 'hypo-iconicity in iconic signs (Farias and Queiroz 2006).

5. Iconicity: Resemblance and metaphoric distortion

In the essay entitled 'on a new list of categories' in *Syllabus*, Peirce divides signs into *icons*, *symbols* and *indexes* (Peirce 1931). Indexes relate signs with the ideas they represent with a factual relation, for instance smoke as representation of fire. Symbols connect signs with the idea they represent by inferred characteristics as in white color symbolizing life. Icons relate signs with the objects they represent by shared qualities such as, phonological resemblance in *onomatopoeia*. Unlike DeSaussure, Peirce asserts that icons are not marginal signs but central to communication (Chu 2020). Peirce highlights the importance of icons saying "every assertion must contain an icon or a set of icons, or else must contain signs whose meaning is only explicable by icons (Peirce 1931 CP 1.158).

Peirce introduces hypo-iconicity as a conceptual framework to describe the essential features of iconic signs. These features are direct resemblance, schematization and metaphoric distortion. In resemblance, the sign has a direct straightforward resemblance with an object in real life (Chu 2018, 2020). Meta-

phoric distortion is a distancing in the expression of meaning between the sign and the object. The sign is placed in a new environment, which invites the reader to exert an interpretation of the sign but this interpretation does not prevent the reader from establishing and reconstructing the historical contextual connection with the object (Chu 2018, 2020). Schematization refers to the transformation of the icon into a pictural diagrammatic form to gain some degree of abstraction. Since the icon in this study is verbal, this feature of schematization is not applicable. This aligns with Chu's (2020) statement that the three features are not necessarily present together in iconic signs. Hypo-iconicity is a scale that describes how iconic signs operate within a spectrum from direct resemblance with the object/idea they represent to diagrammatic abstraction and creative/metaphorical distortion (Chu 2018, 2020).

In the present sign, the features of direct resemblance and metaphorical distortion are applicable and enhance its critical potential by inviting the reader to draw connections. At the resemblance level, by the combined influence of the phonological resemblance with the name, 'Israel' and the resemblance in the shared qualities of cheating according to the sign maker, the reader can make direct connections between the signifier GOT ISRAELED, the signified (taking something from the original owner) and the object/idea in real life 'Israel'. This triadic connection is accentuated by the fact that the sign is a closed one. This is illustrated in the paradigmatic connections this sign can/cannot establish with other signs. As example 1 shows, this sign imposes selection restriction constraints making paradigmatic substitutions of ISRAELED with other lexical items unacceptable as in b and c.

- (1) a- I got israeled.
b- *I got loved.
c- *I got stolen.
d- I got cheated/robbed.

These selection restrictions narrow down the signified to a sudden, negative and painful offense executed by an israeling force upon the 'subject', who is normally a human being taking a *goal* role. Due to this closedness, signifier and signified become mutually evocative. By virtue of the phonological resemblance between the sign and the name Israel, the reader can establish direct straightforward cognitive connections between the sign and Israel. Taub (2004: 21) stresses the link between resemblance and the reader's cognitive processing and critical reading saying that "there is no such thing as 'resemblance' or 'similarity' in the absence of an observer who makes a comparison: resemblance is not an objective fact about two entities but is a product of our cognitive processing" (Taub 2004: 21). These cognitive connections are likely to stereotype Israel as treacherous, which is obvious even among advocates of Israel. Steinberg (in Lebovic 2024) says in *Times of Israel* "the usage suggests that Israel inherently represents unjust appropriation and that offensive behavior typifies Israel. It is an antisemitic use of our people's ancient name".

At the level of metaphoric distortion, the idea of GOT ISRAELED is put in a casual, apolitical environment through the use of several techniques in the definition of the entry and in the example section. In the definition, apolitical participants are used, such as 'someone', 'they', and 'you'. In the example section, a casual usage example is given, in which someone is sitting in a table in a restaurant or a coffee shop and allows a person to share the table with him/her but at the end, this person asks him/her to leave because s/he has a meeting. The use of the table metaphor gives a pseudo-objective aspect to the lexicographic entry. The reader, however, can easily connect the sign with the political meaning and establish the inevitable connection with the historical context of the colonization of Palestine due to the resemblance feature of the sign explained above.

In this section, through applying the conceptual framework of hypo-iconicity, the analysis has shown that the sign displays two essential features of iconic sign, which are resemblance and metaphoric distortion. These features boost the critical potential of the sign and allow the reader to build direct straightforward connections between the sign and the political reality of the colonization of Palestine by Israel. In this study, iconicity is used as a distinct term from iconisation. The latter refers to the strategy of pooling attitudes around iconic signs, which is also a strategy of meaning-making aiming to expose the writers' political views and align the reader around them. This will be explained in details in the following section.

6. Iconisation: Ideation and bonding icon

Iconisation is a semiotic tool and an evolving research direction in SFL that examines the role of signs in "solidarity-building and affiliation" (Stenglin 2012: 125). It refers to the process through which a sign (symbol, icon or index) becomes a bonding icon, "a social emblem, which communities rally around or reject" (Stenglin 2012: 125). A bonding icon realizes the bonding through merging ideational and interpersonal meanings (Stenglin 2004). The ideational meanings connect to the *field* components of register in SFL (cf. Halliday and Matthiessen 2014), such as what the sign is both concretely and symbolically, the social activities associated with it and the underlying social processes (Stenglin 2008, 2012). Such processes can explain how signs are iconized. At the interpersonal level, the bonding is materialized through infusing attitudes to ideational meanings, which results in aligning people into communities and social networks with shared feelings and attitudes (Knight 2010). Turning a sign into a bonding icon is accompanied by a strong interpersonal charge with the amassing of interpersonal attitudes. Martin (2010) claims that ideational meanings are discharged while Stenglin (2012) contends that the ideational is weakened while the interpersonal is intensified and crystallized. The intensification of the interpersonal is mediated through several semiotic resources in verbal and multimodal forms (Wignell, Tan and O'Halloran 2016). Stenglin (2008) explores the expansion of the interpersonal in iconisation through the Appraisal System to gauge the expansion of attitudes

(affect, appreciation and judgement) as developed by Martin and White (2005).

To illustrate the iconisation process, flags, for instance, are bonding icons merging ideational and interpersonal meanings. Ideationally, the sign is associated with concrete component parts evoking its historical and cultural anchoring and is associated with different social activities such as hoisting flags during official ceremonies. Interpersonally, values get charged into this sign such as, pride, belonging, patriotism as well as homesickness, affiliating people into communities of shared values called citizens or countrymen.

The sign under examination is a typical bonding icon that merges ideational and interpersonal meanings. Ideationally, despite the surface apolitical definition discussed in the previous section, the sign recontextualizes the colonization of Palestine, the ethical illegitimacy in this colonization, Jewish settlements as well as the issue of violence associated with them. This ideational recontextualization has induced an interpersonal charging and an intensification of attitudes. The large community of internet users are rallied around a set of values such as the condemnation of the offender and sympathy with the victim. This amassing of attitudes has three manifestations:

- The wide resemiotisation of the sign inside the dictionary and in other mediated forms, playing two roles: being simultaneously a manifestation of iconisation and a semiotic tool fostering this process.
- The large-scale reactions to the sign through social media sharing and the large upvoting in *Urban Dictionary*. Zappavigna (2019) refers to this as an ambient affiliation through which the sign bonds internet users virtually.
- The use of semiotic resources in the definition to underscore the writers' attitude and the attempts to align the reader with it.

The current literature contends that a bonding icon triggers alternative reactions: communities either endorse an icon or reject it. In both situations, the icons are "supercharged and radiate specific values for communities to rally around" (Stenglin 2022: 6). *Bond-icons* evoke the communities' positive attitudes towards an icon, while *anti-cons* evoke rejection (Carr 2023, Stenglin 2022, Knight 2010). In both cases, bonding icons transcend differences and lead to unity and conformity. Stenglin (2012) argues that the Olympic torch is a bonding icon uniting people around the values of excellence, friendship and respect associated with the Olympic values to the point that the Nazi origin of the torch is totally eradicated. To my knowledge, there is a scarcity of research on a bonding icon triggering divergent attitudes across communities. The present paper contends that communities are rallied around the bonding icon according to two opposing directions. Pro-Palestine voices endorse the sign and align around the values it promotes. Pro-Israeli voices reject the sign through downvoting it in *Urban Dictionary*, and organizing social media campaigns for this purpose. The divergence of attitudes towards a bonding icon is expressed thanks to the affordances of online dictionaries and the crowdsourced ones, in particular.

To uncover the semiotic resources used by meaning-makers for the iconi-

sation of the sign, the analysis draws upon a toolkit from SFL to map the construction and aligning of attitudes. According to this theory, every situation of communication is an act of meaning (Halliday and Matthiessen 2014), which fulfills simultaneously three metafunctions: an ideational, an interpersonal and a textual. In the context of iconisation where ideational and interpersonal meanings are compressed, the two first functions are particularly relevant. They are defined as follows:

- The ideational function refers to the construal of experience through a configuration of participants, process types and optional circumstances. The participants are the represented participants who perform different participant roles depending on the process type. For instance, in the sentence, 'John hits Mary', 'John' and 'Mary' are the represented participants assuming the roles of *actor* and *goal* respectively. This meaning is mediated through the *material process*, 'to hit'. In counterpart, in 'John loves Mary', 'John' and 'Mary' are the represented participants assuming the roles of 'senser' and 'phenomenon' respectively. This meaning is mediated through the *mental process*, 'to love'.
- The interpersonal function is explored from the principle of interactive participants, which refers to the relation between interactants, who are the participants in a conversation or the writer, on the one hand and the reader, on the other (Kress and Van Leeuwen 1996/2021). Different resources mediate this relation in SFL such as, the articulation of positive and negative attitudes, which can be graduated by *intensification* or *weakening* (Martin and White 2005).

Table 1 introduces the resources used to map the iconisation of the sign under examination. It proposes a two-levelled analysis to gauge the coupling of ideational and interpersonal meanings.

Table 1: Ideational and interpersonal systems for the analysis of iconisation

	Ideational	Interpersonal
Participants and participant roles	Represented participants Process types and participant roles (Halliday and Matthiessen 2014)	Interactive participants (Kress and Van Leeuwen 1996/2021)
Ideational expansion and interpersonal charging	Metaphoric distortion (Peirce 1931)	Graduation (Martin and White 2005)

6.1 Represented participants and interactive participants

At a first level, ideationally, focus is laid on the identification of the represented participants, the participants introduced in the definition of the entry and in the

example. Emphasis is also put on the roles of represented participants (actor, goal, sayers, sensors) and mediating process types (*material*, *mental*, and *verbal*). Interpersonally, emphasis is put on the interactive participants, the writers of the entry, and the readers and how they connect to each other.

6.1.1 Represented participants, their roles and process types

The represented participants can be divided into two broad categories: the israeling participant (oppressor) and the israeled (victim). The oppressor is associated with the third person in the singular and plural, 'someone' is used in 20 instances and 'they' in 17. The victim is frequently associated with the first person singular and the second person plural; 'I' is employed 19 times and 'you' 17. This distribution is meant to identify with the victim, who is either the speaker or the reader. On the contrary, the israeling participant is distanced through the use of the third person to reflect the authors' negative attitude to them. The shift from the third person singular 'he/someone' to the plural 'they' indicates that the israeling act has evolved from the act of a single person to the work of a group of people. In addition to the use of personal pronouns, the two categories of participants are referred to with names. In seven entries, names with apolitical orientation (Elizabeth, Peter, John, Robert and Nick) are used to give a casual dimension to the entries. As the resemiotisations progress in time, proper names with obvious political, cultural and symbolic connotations are used. Names such as, Benjamin, Jacob, Hananya versus Ishmail, Ali, Maryem and Hamad obviously recontextualize the conflict between Jews and Arabs. In this conflict, the writers of the entries are sided by the Palestinians, who are described as the 'rightful owner'. On the contrary, nouns and names such as, 'trespasser' and 'Barbara' (a name in Greek mythology to refer to a foreigner or a stranger) are used to refer to the israeling participant.

Table 2 shows that the israeling participant has functioned more frequently as the subject of the sentence assuming active participant roles mediated by verbal, material and mental process types (143 versus 19). The material processes with an actor participant role are the most frequent (70 material processes). They are realized by the verbs 'steal' and 'take' (29 processes) aiming to construct the idea of unlawful taking. This culminates in the use of 'to israel' in the active voice, which is exclusively used in the resemiotisations (12 processes). The clear straightforward association of this with the war on Palestine is made through the use of the verb 'to kill' (3 times) as in "when you resist they kill you for it and take everything else you own" (ISRAELED, 5 November 2023). The Ideological anchoring of the entry is also evident through associating GOT ISRAELED with the victimhood narrative in Jewish communities (Markiewicz and Sharvit 2021). This is mentioned in five entries as in "to act like the victim 24/7, but is able to steal things that belong to someone else" (ISRAELED, 5 November 2023).

The verbal processes come second in frequency with 63 occurrences. The verbs 'ask' and 'claim' are the most frequent (14 and 15 occurrences respectively).

These processes are chronologically sequenced to mean that the israeling act started with asking someone to share something with him/her and then claiming it as one's property as in "when someone asks to share something of yours and then claims it's become theirs and takes it for themselves" (GOT ISRAELED, 28 October 2023).

As opposed to the frequent occurrences of the israeling participant in a subject position with actor and sayer participant roles, the israeled participant assumes these roles only 19 times, which are often associated with the verbs 'share', 'allow' and 'give' (six verbs). The political nature of the entries is also highlighted through the use of the verbs of 'refuse' and 'resist' as in "when you resist, they kill you for it and take everything else you own" (ISRAELED, 5 November 2023). As Table 2 shows, the israeled participant is more frequently associated with the goal participant role in both subject and object positions. In all the occurrences of subject as goal, the israeled participant is used with the verb 'to israel' in the passive voice or with the mutative 'got'.

Table 2: Participant roles in the entries

Participant roles	Israeling participant	Israeled participant
Subject as actor/sayer/ carrier	143	19
Subject as goal	0	30
Object as goal	2	30

6.1.2 Interactive participants

Writers of the entry are portrayed as articulate and biased participants in the different entries by means of several communication strategies. In 19 instances, the writer is identified as the victim as in "I was sitting at my table I think I just GOT ISRAELED". The use of pseudonyms such as 'Israeled', 'an Israeled man' indicate the direct involvement of the writer. Another manifestation of the writers' bias is the use of pseudonyms with clear political affiliations such as 'October Revolution' 'FreeFelestin'.

Writers establish direct bonds with the reader. This is illustrated in 17 entries, where he/she addresses the reader directly and identifies him/her as the victim as in "when you allow someone to use a thing you own or sit in a place reserved for you and then they claim it's theirs all along" (ISRAELED, 21 October 2023). The reader can express his/her attitude towards these entries thanks to upvoting and downvoting. These reactions reflect that the sign triggered divergent attitudes among internet users (almost 600000 reactions from 21 October 2023 to 11 October 2024). It is noticeable that the total number of upvoting exceeds downvoting (387405 versus 201628).

To sum up, the ideational choices (participant roles and process types) indicate that the israeling participant is introduced as the major participant in the

construal of experience, who is taking what originally belongs to the israeled participant. However, through the use of personal pronouns, the writers identify with the israeled participant and distance themselves from the israeling one. This strategy leads to the interpersonal charging of the sign with positive attitudes towards the victim and negative ones towards the oppressor. Interpersonally, this charging is translated into positioning the victim as teller of the story. This highlights the causal connection between the two metafunctions; the reality of colonization has amassed positive interpersonal attitudes such as sympathy and compassion towards the victim. This is congruent with the analysis of interactive participants, which also highlights the interpersonal charge of the sign manifested in the reactions of the readers.

6.2 Metaphoric distortion and graduation

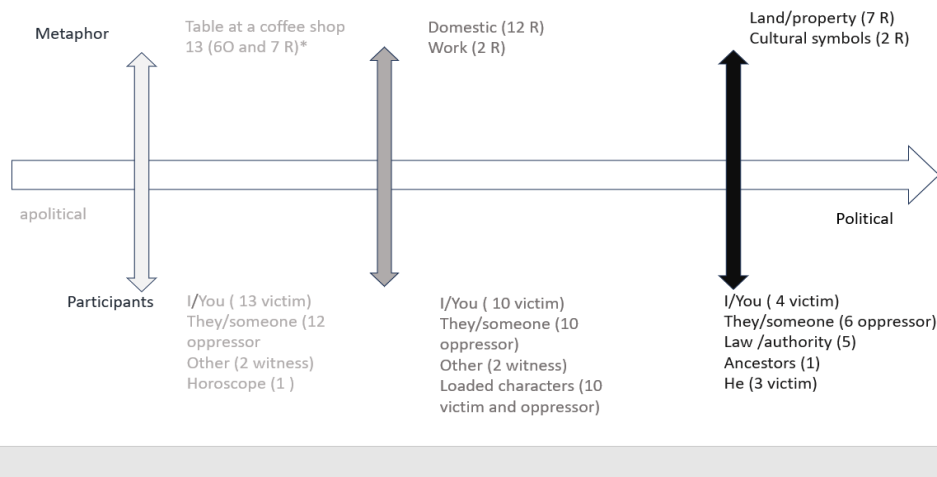
At a second level, the analysis of metaphoric distortion reveals the ideational expansion of the sign through the escalating recontextualization of reality as the resemiotisations unfold. The graduation analysis underscores the interpersonal charging of the bonding icon.

6.2.1 Metaphoric distortion

Metaphoric distortion is a central feature of iconic signs. The analysis focusing on the original entries (those written on 21 October 2023) done in section 5 has shown that the 'table metaphor', a casual example used as a metaphoric explanation of GOT ISRAELED, does not prevent the reader from establishing straightforward connections with the colonization of Palestine. This section concentrates on the metaphoric distortions in the resemiotisations to underscore the progressive recontextualizations of this colonization as the resemiotisations progress in time. The analysis of metaphoric distortion covers shifts at three areas:

- Nature of metaphor across resemiotisations
- Participants' identity
- Use of process types

As indicated in Figure 3, the use of metaphors shifts towards a more political orientation as the resemiotisations unfold. The original entries use a casual metaphor of a table in a restaurant or a coffee shop. It is accompanied by a humorous/sarcastic tone as in "is this an episode of Seinfeld or am I getting israeled?" (ISRAELED, 21 October 2023) or "my horoscope said that I should be sitting at this table" (ISRAELED, 22 October 2023). As resemiotisations progress, metaphors relating to the domestic and work domains are used such as israeling one's t-shirt, pen, ice-cream, bicycle or project. The political orientation becomes more evident in the last resemiotisations. The metaphors used become those of israeling land/property and cultural symbols such as Falafel and tangerines.



*O= original entry R= resemiotisation

Figure 3: Metaphoric distortion across entries

This incremental political shift is also mirrored in the use of participants. In the original entries, the participants are consistently used with 'I' and 'you' representing the israeled and 'someone' and 'they' representing the israeling participants. In the resemiotisations with the domestic/work metaphors, in addition to the participants mentioned above, others like mother, sister and colleagues are used. Most importantly, pairs of participants with cultural connotations are used, such as Ali versus Jacob and Ishmail versus Barbara. These pairs not only reflect the cultural orientation of the entry but also point out the conflict between the two cultures. In the property-related metaphors, the use of participants becomes less consistent ('he' is used for israeling and israeled participants). More importantly, intervening participants denoting power and authority are used to show that the israeling act is helped by power as in "when a person tells you that your property is theirs ... and demands you just give this property to them, and if you refuse, they take it by force and the law will somehow be on their (israeling) side" (ISRAELED, 22 October 2023). The political dimension is, in particular, emphasized in one resemiotisation, which depicts a conversation between a settler and the original owner of the land, in which the former claims the land belonged to his ancestors "my ancestors were here 3000 years ago. If I do not take your house, someone else will" (ISRAELED, 28 October 2023). This statement is actually identical to a previous utterance in a real-life video reporting a conversation between a Jewish settler and a Palestinian woman in Cheikh Jarrah area in 2021 (AJ+ 2021).

The evolving political orientation is also evident in the use of process types.

Though verbs like 'claim' 'take' 'steal' and 'kill' are employed in the original metaphor, their number increases in the domestic and property metaphors (for instance 'claim' evolves from four to seven and 'steal' from one to seven). The verb 'act like a victim' was scarcely used with the table metaphor as opposed to six uses in the other metaphors.

6.2.2 Graduation

Graduation is a system within the Appraisal Theory of Martin and White (2005: 135-155), which maps evaluation in language. It covers, alongside with *graduation*, *attitude* and *engagement*. Graduation is anchored within SFL's interpersonal metafunction and includes the resources for grading evaluation through *up-scaling* and *down-scaling*. As shown in Figure 4, graduation functions along two axes of evaluation, which are *force* and *focus*. Scalability along the axis of force covers grading according to *intensification* and *quantification*. Graduation along the axis of focus refers to grading according to *prototypicality*. The first sub-system of graduation includes evaluation relative to the degree of intensity as well as quantity. Assessment of degree of intensification can operate at the level of quality as in 'very kind' and 'quite helpful' and at the level of 'processes' 'this greatly surprised me' and 'you slightly disturbed me'. Graduation in quantification covers amount (size and number) and extent (time and space) as in 'a short break' and 'a large field'. Graduation according to prototypicality (focus) delineates the degree the phenomenon matches the prototypical features of a semantic category as in 'a true lover' (sharpening) and 'kind of lover' (softening).

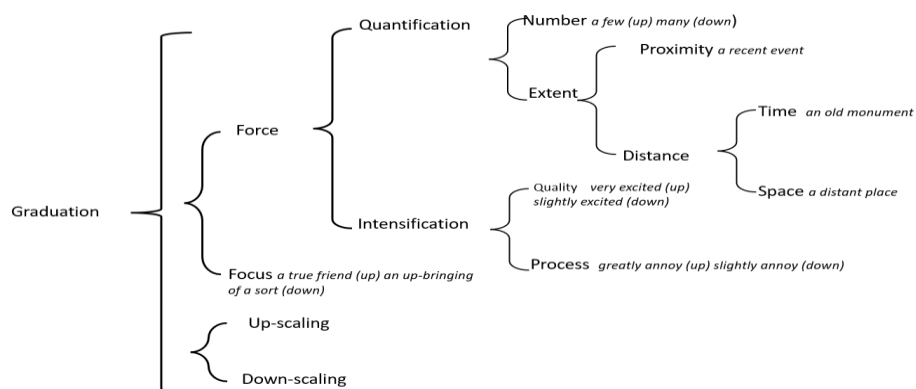
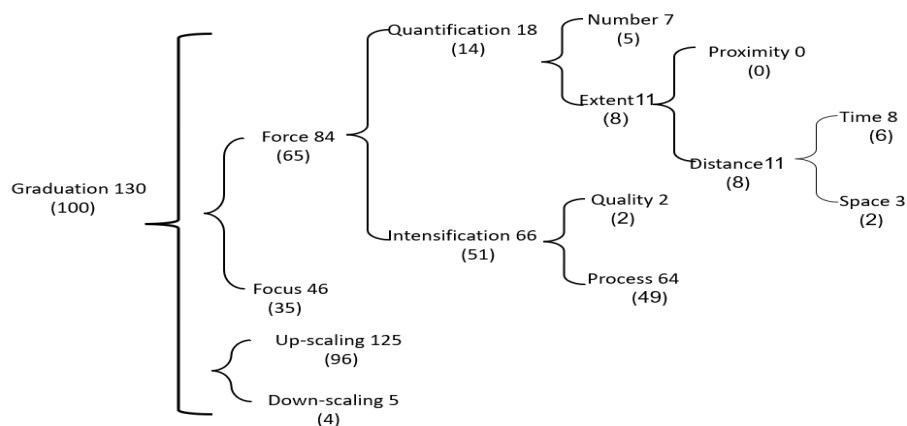


Figure 4: System network for graduation adapted from Martin and White (2005: 154)

Since graduation is concerned with grading experiences and attitude, this framework can be useful in understanding the iconisation of signs. Graduation can

reveal the resources the authors used in charging the sign with values in order to orient the reader. The entries include 130 graduation instances at an average of four graduations per entry. As shown in Figure 5, the direction of grading is predominantly up-scaling. Force-based graduation is more frequent than focus-based one (84 versus 46).

Graduation in terms of force using the sub-systems of intensity and quantification interacts with the up-scaling of attitude, which "frequently acts to construe the speaker/writer as maximally committed to the value position being advanced and hence as strongly aligning the reader into that value position" (Martin and White 2005: 152). Intensification of qualities by grading adjectives and adverbs are typically attitudinal. They are, however, rarely used in the entries under examination, which may be explained by the fact that clearly inscribed attitudes are incongruent with the genre of dictionary entries. On the contrary, intensification by processes is very frequent with 66 instances. This intensification is not typically attitudinal: up-scaled verbs are used to convey the invoked attitude of the authors towards the action of both israeling participants as in 'kick out the rightful owner' "get you out by force" (ISRAELED, 21 October 2023). Grading of processes is also used to convey the degree of powerlessness of the israelied as in "you can't resist it whatsoever" (To israel, 25 October 2023).



() The figures between parentheses refer to the percentage of each graduation resource.

Figure 5: Frequency of graduation resources in the entries

Quantification resources are used to convey the extent of the israeling act in terms of taking the victim's property as in "and take **everything else** you own" (ISRAELED, 5 November 2023). Quantification covering extent in time is used to highlight the religious anchoring of the israeling act "when someone asks for something from you and then **overtime** starts claiming that thing was **always**

theirs" (ISRAELED, 5 November 2023 emphasis added). The authors draw connection with the Jewish propaganda of promised land as "the act of taking something from someone else for some reason thousands of years ago" (ISRAELED, 28 October 2023). The use of these quantification resources is concentrated in the entries recontextualizing the religious conflict.

Graduation under focus sub-system amounts to 46 instances representing 35% of all graduation resources. They are attitudinal terms construing the speakers' attitude towards a situation or a fact, which are explicitly attitudinal as in "it can mean causing **harm** and **destruction** in a **tyrannical way**" (TO ISRAEL, 25 October 2023 emphasis added) or in "a **sickly exaggerated desire** to have something that belongs to someone else" (ISRAELING, 24 October 2023 emphasis added). These graduation resources "indicate maximal investment by the authorial voice in the value position (either negative or positive) being advanced and hence to strongly align the reader into the value position being advanced." (Martin and White 2005: 139)

To sum up, at the ideational level, metaphoric distortion is used to illustrate the shift into more politicized metaphors. This shift underscores the authorial voice, which is realized with the use of graduation resources meant to accentuate the interpersonal charge of the bonding icon. This finding aligns with Stenglin (2008) who emphasizes the coupling of ideational and interpersonal meanings to charge values into the Olympic flame. She demonstrates this through combining the ideational activities of lighting the Olympic flame with the attitudes of affect, appreciation and judgement evoked by these activities. The present study uses the coupling of metaphoric distortion and graduation because these systems are more congruent with the genre of dictionary entries where invoked attitudinal choices are genre-appropriate.

7. Conclusion

This paper has proposed to link Lexicography and CDA by demonstrating that the semantic and lexical descriptions of new dictionary entries should be supplemented with a critical reading. As shown by the study, the ideology-infused entries constitute interpersonally-charged semiotic acts introducing the sign-maker's world view and inducing reactions from the reader. This was illustrated by GOT ISRAELED, a new entry in the crowdsourced *Urban Dictionary* that was heavily resemiotised inside the dictionary reaching 36 entries.

The sign was analyzed at three levels: coinage, iconicity and iconisation drawing upon Semiotic and Social Semiotic frameworks. The first level of analysis shows that the sign is a motivated sign showing the correspondence between structural features, social anchoring and sign makers' intentions. At the second level, Peirce's (1931) principles of hypo-iconicity underscore the critical potential of the sign through drawing straightforward links with the colonization of Palestine. Third, this paper has exploited the Social Semiotic principle of iconisa-

tion to describe the transformation of the sign into a bonding icon rallying people around the writers' world view (Stenglin 2012).

This study has contributed to the current literature on the relatively new area of critical lexicography by introducing a framework for the analysis of signs with a semiotic action. This framework illustrated in Figure 6 represents a two-levelled analysis drawing upon tools from Critical Lexicography (Chen 2019), Semiotics (DeSaussure 1916, Peirce 1931) and Social Semiotics (Halliday and Matthiessen 2014, Martin and White 2005). The first level is a contextual analysis emphasizing the generic structure of the entry, the context of genre production and explores the specific political context of the entry coinage.

The second level of analysis comprises three steps showing that the potential of signs to exert action can be viewed on a spectrum ranging from the coinage of motivated signs to the design of signs building real and distorted connections with reality (iconicity). At the highest level of this spectrum is iconisation, in which the sign maker exploits ideational and interpersonal resources to turn the sign into a bonding icon. This level is reached when the sign is largely resemiotised in different mediated forms. The aims of resemiotisations are to construct a set of values around the iconized sign.

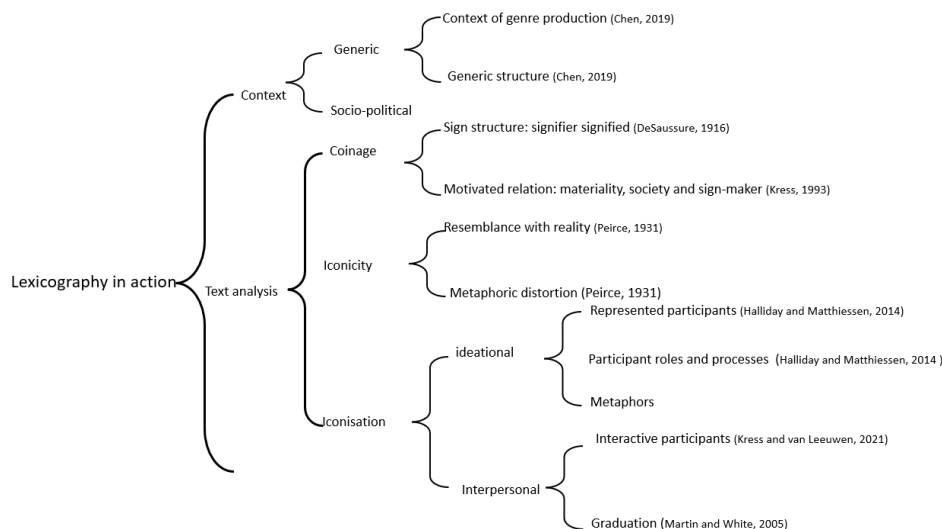


Figure 6: A Framework for the analysis of lexicographic entries in action

This framework has been used to analyze dictionary entries but it is applicable to other signs in language and other semiotic systems. Multimodality can promote semiosis in action at the levels of coinage and iconicity as well as iconisation because it can help the expansions of the sign's meaning. This can be a promising topic for future research.

References

- AJ+. 2021. If I Don't Steal [Your Home] Someone Else Will. Israeli Settler Justifies Forcible Takeover. [video]. YouTube, 5 May 2021.
<https://www.youtube.com/watch?v=KNqozQ8uaV8>.
- Bergenholtz, H. and R.H. Gouws. 2012. What is Lexicography? *Lexikos* 22: 31-42.
- Bertulesi, C. 2023. Chinese Lexicography and the Critical Analysis of Lexicographical Discourse. Zuccheri, S. (Ed.). 2023. *Studies on Chinese Language and Linguistics in Italy*: 95-116. Bologna: Bologna University Press.
- Björkvall, A. and A.-M. Karlsson. 2011. The Materiality of Discourses and the Semiotics of Materials: A Social Perspective on the Meaning Potentials of Written Texts and Furniture. *Semiotica* 187(1-4): 141-165.
- Carr, G. 2023. *With Respect to Consent: The Language of Sex Education*. Unpublished Ph.D. Dissertation. Sydney: University of Sydney.
- Chen, W.G. 2015. Bilingual Lexicography and Recontextualization: A Case Study of Illustrative Examples in a New English-Chinese Dictionary. *Australian Journal of Linguistics* 35(4): 311-333.
- Chen, W.G. 2019. Towards a Discourse Approach to Critical Lexicography. *International Journal of Lexicography* 32(3): 362-388.
- Chu, Y. 2018. Hypoiconicity in the Architecture of Suzhou: Authentic Resemblance, Diagrammatic Reduction, and Metaphoric Displacement. *Social Semiotics* 30(1): 114-132.
doi:10.1080/10350330.2018.1531548.
- Chu, Y. 2020. On the Hypoiconic Structure of Cartoons. *Social Semiotics* 32(2): 262-278.
<https://doi.org/10.1080/10350330.2020.1756587>.
- Coleman, J. 2004. *A History of Cant and Slang Dictionaries*. Oxford: Oxford University Press.
- DeSaussure, F. 1916. *Course in General Linguistics*. (Edited by Bally, C. and A. Sechehaye in collaboration with A. Riedlinger. 1966. Translated, with an Introduction and Notes by Wade Baskin.) New York/Toronto/London: McGraw-Hill.
- Fairclough, N. 1995. *Critical Discourse Analysis: The Critical Study of Language*. London: Longman.
- Farias, P. and J. Queiroz. 2006. Images, Diagrams, and Metaphors: Hypoicons in the Context of Peirce's Sixty-six Fold Classification of Signs. *Semiotica* 162(1-4): 287-308.
- Fuertes-Olivera, P.A. and S. Tarp. 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132.
- Gouws, R.H. 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Halliday, M.A.K. and C.M.I.M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar*. Fourth edition. Oxon: Routledge.
- Hodge, R. and G. Kress. 1988. *Social Semiotics*. Cambridge: Polity.
- Holdcroft, D. 1991. *Saussure: Sign, System and Arbitrariness*. Modern European Philosophy. Cambridge: Cambridge University Press.
- Hornscheidt, A. 2008. A Concrete Research Agenda for Critical Lexicographic Research within Critical Discourse Studies: An Investigation into Racism/Colonialism in Monolingual Danish, German, and Swedish Dictionaries. *Critical Discourse Studies* 5(2): 107-132.
- Kachru, B.B. 1995. Afterword: Directions and Challenges. Kachru, B.B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta*: 417-424. Tübingen: Max Niemeyer.

- Kirkness, A.** 2004. Lexicography. Davies, Alan and Catherine Elder. 2004. *The Handbook of Applied Linguistics*: 54-81. Oxford: Blackwell.
- Knight, N.K.** 2010. Wrinkling Complexity: Concepts of Identity and Affiliation in Humour. Bednarek, M. and J.R. Martin (Eds.). 2010. *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*: 35-58. London: Continuum.
- Kress, G.** 1993. Against Arbitrariness: The Social Production of the Sign as a Foundational Issue in Critical Discourse Analysis. *Discourse and Society* 4(2): 169-193.
- Kress, G.** 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*, London: Routledge.
- Kress, G. and T. van Leeuwen.** 1996/2021. *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Kress, G. and T. van Leeuwen.** 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. New York: Hodder Arnold.
- Lebovic, M.** 2024. Is *Urban Dictionary*'s New Entry 'to Get Israeled' an Antisemitic Trope? *The Times of Israel*, 11 January 2024.
<https://www.timesofisrael.com/is-urban-dictionarys-new-entry-to-get-israeled-an-antisemitic-trope-adl-weighs-in>
- Ledin, P. and D. Machin.** 2018. Multi-modal Critical Discourse Analysis. Flowerdew, J. and J.E. Richardson (Eds.). 2018. *The Routledge Handbook of Critical Discourse Studies*: 60-76. London/New York: Routledge.
- Linell, P.** 1998. Discourse across Boundaries: On Recontextualizations and the Blending of Voices in Professional Discourse. *Text* 18(2): 143-157.
- Liu, X.** 2015. Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos* 25: 210-232.
- Markiewicz, T. and K. Sharvit.** 2021. When Victimhood Goes to War: Israel and Victim Claims. *Political Psychology* 42(1): 111-126.
- Martin, J.R.** 2010. Semantic Variation: Modelling Realisation, Instantiation and Individuation in Social Semiosis. Bednarek, M. and J.R. Martin (Eds.). 2010. *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*: 1-34. London: Continuum.
- Martin, J.R. and P.R.R. White.** 2005. *The Language of Evaluation. Appraisal in English*. New York: Palgrave Macmillan.
- Moalla, D.** 2023. The Resemiotisation of Chokri Belaid's Talk before his Assassination: Roles of Sign-maker Interest and Re-reading Forces. *Social Semiotics* 33(4): 808-840.
DOI: 10.1080/10350330.2021.1949245.
- Peirce, C.S.** 1907. Pragmatism. Houser, N. and C.J.W. Kloesel (Eds.). 1998. *The Essential Peirce. Vol. 2: Selected Philosophical Writings (1893-1913)*: 398-433. Bloomington: Indiana University Press.
- Peirce, C.S.** 1931. *The Collected Papers of Charles Sanders Peirce*. Hartshorne, C., P. Weiss and A.W. Burks (Eds.). 1966. Cambridge: Harvard University Press. [Reference to Peirce's papers will be designated CP followed by volume and paragraph number.]
- Pettini, S.** 2023. The Language of Ethnic Conflict in English Online Lexicography: Ethnophaulisms in "powered by Oxford" Lexico.com. *Lexikos* 33: 299-317.
- Ro, C.** 2019. How Linguists Are Using *Urban Dictionary*. *Daily JSTOR*, 13 November 2019.
<https://daily.jstor.org/how-linguists-are-using-urban-dictionary/>
- Stenglin, M.** 2004. *Packaging Curiosities: Towards a Grammar of Three-dimensional Space*. Unpublished Ph.D. Dissertation. Sydney: University of Sydney.

- Stenglin, M.** 2008. Interpersonal Meaning in 3D Space: How a Bonding Icon Gets its 'Charge'? Unsworth, L. (Ed.). 2008. *Multimodal Semiotics: Functional Analysis in Contexts of Education*: 50-66. London: Continuum.
- Stenglin, M.** 2012. 'Glocalisation': Exploring the Dialectic between the Local and the Global. Bowcher, W.L. (Ed.). 2012. *Multimodal Texts from Around the World: Cultural and Linguistic Insights*: 123-145. Houndmills, Basingstoke, Hampshire/New York: Palgrave Macmillan.
- Stenglin, M.** 2022. Binding and Bonding: A Retrospective and Prospective Gaze. Caldwell, D., J.S. Knox and J.R. Martin (Eds.). 2022. *Applicable Linguistics and Social Semiotics: Developing Theory from Practice*: 309-324. London: Bloomsbury.
- Strand, T.** 2013. Peirce's Rhetorical Turn: Conceptualizing Education as Semiosis. *Educational Philosophy and Theory* 45(7): 789-803.
- Taub, S.F.** 2004. *Language from the Body. Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press.
- Urban Dictionary.** 2023–2024. *Israeled*.
<https://www.urbandictionary.com/define.php?term=Israeled>
- Urban Dictionary.** 2024. *Homepage*.
<https://www.urbandictionary.com/define.php?term=homepage>
- Van Leeuwen, T.** 2005. *Introducing Social Semiotics*. London/New York: Routledge.
- Wignell, P., S. Tan and K.L. O'Halloran.** 2016. Violent Extremism and Iconisation: Commanding Good and Forbidding Evil? *Critical Discourse Studies* 14(1): 1-23.
- Zappavigna, M.** 2019. Ambient Affiliation and #Brexit: Negotiating Values about Experts through Censure and Ridicule. Koller, V., S. Kopf and M. Miglbauer (Eds.). 2019. *Discourses of Brexit*: 48-68. London: Routledge.