# Lexikos  33(2)

# Lexikos  33(2)

## Rufus H. Gouws: 'n Huldeblyk A Tribute

*Redakteurs*
*Editors*

Ilse Feinauer

Amanda Marais

Marius Swart

African Association for Lexicography

BURO VAN DIE WAT

STELLENBOSCH

# Inhoud / Contents

**Artikels / Articles**

# Voorwoord

Ons kollega en vriend Rufus Gouws tree aan die einde van 2023 af ná 'n akademiese loopbaan van meer as 45 jaar. Dit is gebruiklik om ten tyde van so 'n aftrede 'n gedetailleerde lys van die betrokke kollega se pryse, toekennings, publikasies, nagraadse studente en dies meer te maak. Al hierdie inligting is reeds elders uitvoerig geboekstaaf, en in die geval Gouws is so 'n lys nie kort nie. Dit word dus nie hier herhaal nie.

Daar is egter twee aspekte van hierdie leksikograaf en taalkundige se loopbaan wat moontlik ook meer gepas is vir 'n afskeid van 'n jarelange kollega. Eerstens, die begin van die betrokke loopbaan.

Rufus vertel 'n keer in die personeelkamer dat hy tydens sy tyd aan die Universiteit van Suid-Afrika die eerste keer die taak opgelê is om 'n kursus aan te bied in wat vandag as leksikografie bekend staan. In daardie stadium was sy navorsingsfokusarea die Afrikaanse sintaksis. Na afloop van dié gesprek moes hy in 'n verklarende woordeboek gaan naslaan wat presies die woord beteken wat nou, bykans vyf dekades later, bykans 'n sinoniem vir sy werk geword het.

Tweedens behoort daar iets genoem te word oor nalatenskap. Rufus se eie werk en publikasies is reeds besonder indrukwekkend. In hierdie huldigingsuitgawe van *Lexikos* kom egter 'n bykomende dimensie daarvan na vore. Hier is naamlik al die tekens van die kundigheid en ook die menslikheid wat studente en kollegas in verskeie uithoeke van die wêreld geïnspireer het, en in die proses die leksikografiese tuiste inderdaad 'n huis met baie wonings gemaak het — 'n uitdrukking wat Rufus self graag gebruik.

Ons neem afskeid van jou, Rufus, op 'n manier wat hopelik iets vervat van toewyding, die erns en, as ons mag, die liefde wat jy vir die taalkunde, vir Afrikaans en natuurlik vir die leksikografie beliggaam. Jy skryf in 2012 'n artikel met die titel "Who can really be called a lexicographer?" en ons is van mening dat jy dit met meer as voldoende oortuiging beantwoord het. Geluk met 'n verbluffende loopbaan, dankie vir jou kollegialiteit, en net die beste wense vir jou aftrede.

Mag die welverdiende rus, soos jy dit sal stel, bloedstollend lekker wees.

Marius Swart, Ilse Feinauer en Amanda Marais
Redakteurs

# Foreword

Our colleague and friend Rufus Gouws is retiring at the end of 2023 after an academic career spanning more than 45 years. At this time, standard practice would entail making a detailed list of said colleague's prizes, awards, publications, postgraduate students, and the like. All of this information has been recorded extensively elsewhere, however, and in the Gouws case the list is not short. Therefore, we will not repeat it here.

However, there are two aspects of the career of this lexicographer and linguist that might be better suited to saying farewell to a colleague of so many years. Firstly, the start of it all.

In the staff room, Rufus once told us about his time at the University of South Africa, where he was first tasked with presenting a course on what is known today as lexicography. At that point, his research focus area was Afrikaans syntax. After the instruction to present this course, he had to use a dictionary to find the exact meaning of the word that now, five decades later, stands as a synonym to his work.

Secondly, there is the matter of heritage. Rufus's own work and publications are undoubtedly impressive. In this special issue of *Lexikos*, however, another dimension comes to the fore. Here we have the clearest of indications of the expertise and the humanity that have inspired students and colleagues from many parts of the world. The lexicographic world knows no borders, in large part thanks to Rufus.

We are saying farewell, Rufus, in a way that hopefully captures your dedication, your zeal and, if we may, your love for the study of language, for Afrikaans and, of course, for lexicography. In 2012 you wrote an article titled "Who can really be called a lexicographer?". We strongly believe that you have answered this question more than convincingly. Congratulations with your astounding career, thank you for your collegiality, and best wishes in your retirement.

May you enjoy your well-deserved rest, and perhaps also press some flowers.

Marius Swart, Ilse Feinauer and Amanda Marais
Editors

# Redaksionele doelstellings

*Lexikos* is 'n tydskrif vir die leksikografiese vakspesialis en word in die AFRI-LEX-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 leksiko-grafiese gesprek te skep, en in die besonder die leksikografie in Afrika met sy ryk taleverskeidenheid te dien;

(2)    om die gesprek tussen leksikograwe onderling en tussen leksikograwe en taalkundiges te stimuleer;

(3)    om kontak met plaaslike en buitelandse leksikografiese projekte te bewerkstellig en te bevorder;

(4)    om die interdissiplinêre aard van die leksikografie, wat ook terreine soos die taalkunde, algemene taalwetenskap, leksikologie, rekenaarwetenskap, bestuurskunde, e.d. betrek, onder die algemene aandag te bring;

(5)    om beter samewerking op alle terreine van die leksikografie moontlik te maak en te koördineer, en

(6)    om die doelstellings van die *African Association for Lexicography* (AFRILEX) te bevorder.

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

(1)    Bydraes tot die leksikografiese gesprek word in die vaktydskrif *Lexikos* in die AFRILEX-reeks gepubliseer.

(2)    Monografiese en ander studies op hierdie terrein verskyn as afsonderlike publikasies in die AFRILEX-reeks.

(3)    Slegs bydraes wat streng vakgerig is en wat oor die suiwer leksikografie of die raakvlak tussen die leksikografie en ander verwante terreine handel, sal vir opname in die AFRILEX-reeks kwalifiseer.

(4)    Die wetenskaplike standaard van die bydraes sal gewaarborg word deur hulle aan 'n komitee van vakspesialiste van hoë akademiese aansien voor te lê vir anonieme keuring.

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

# 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.

# Voorlopige kriteria vir
# wie en wat die leksikograaf is

Marius Swart, *Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Suid-Afrika (mariusswart@sun.ac.za)*
en
André H. du Plessis, *Buro van die WAT, Stellenbosch, Suid-Afrika; Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Suid-Afrika (ahdp@sun.ac.za)*

**Opsomming:** Daar is verbasend min sistematiese akademiese ondersoek na wie presies leksikograwe is. Rufus Gouws wys hierdie problematiek reeds 'n dekade gelede uit. Gouws (2012) ondersoek nie net die aktiwiteite wat verband hou met Leksikografie nie, maar vra krities watter rolle die persone speel wat binne hierdie vakgebied bedrywig is. Gouws (ibid.) se bespreking lei af dat dit nodig is om ondubbelsinnige kriteria te ontwikkel vir wie 'n leksikograaf genoem kan word. Om hierdie kriteria te help ontwikkel, neem hierdie artikel die werk van Gouws onder die loep in die vorm van 'n gevallestudie met 'n onderliggende tematiese analise. Gouws se eie bydrae op metaleksikografiese vlak, maar ook tot die leksikografiese praktyk en leksikografie-opleiding, spreek vanself in terme van produktiwiteit, kundigheid en relevansie. Die gevallestudie behels 'n literatuuroorsig oor Gouws se navorsingsartikels uit die afgelope 10 jaar sedert die verskyning van die genoemde 2012-artikel. Die artikels is tematies ontleed en 'n stel rolle word afgelei wat algemeen blyk uit hierdie leksikografiese publikasies. Hierdie rolle of aktiwiteite word indringend bespreek, en uit die bespreking word uiteindelik 'n voorlopige stel kriteria afgelei vir wie 'n leksikograaf genoem kan word.

**Sleutelwoorde:** LEKSIKOGRAAF, METALEKSIKOGRAAF, WOORDEBOEKMAKER, PEDAGOOG, TAALKUNDIGE, TAALBURGERSKAP, AKTIVISME

**Abstract: Proposed Criteria for Being a Lexicographer.** There is surprisingly little systematic academic research on who exactly lexicographers are. This issue was identified a decade ago by Rufus Gouws. Gouws (2012) not only investigates the activities of those involved in Lexicography, but critically questions the roles of the active participants. Gouws (ibid.) concludes by emphasising that it is imperative to develop unambiguous criteria for who can be called a lexicographer. This article aims to help develop these criteria by placing the work of Gouws under the microscope in the form of a case study with an underlying thematic analysis. Gouws' own contributions to the meta- and practical lexicography, as well as to dictionary didactics and pedagogy, attests to his productivity, expertise and relevance. This case study includes a literature review of Gouws' research articles from the last 10 years since the publication of the 2012 article mentioned above. The articles are analysed thematically to deduce the general roles and activities of the lexicographer. These roles or activities are discussed incisively, and from this discussion, a set of tentative criteria is proposed for who can be called a lexicographer.

**Keywords:** LEXICOGRAPHER, METALEXICOGRAPHER, DICTIONARY MAKER, EDU-CATOR, LINGUIST, LINGUISTIC CITIZENSHIP, ACTIVISM

## 1.      Inleiding

Die leksikograaf is binne die leksikografie 'n noodwendigheid. Nieteenstaande hierdie logiese gegewe, is daar verbasend min sistematiese akademiese onder-soek na wie presies leksikograwe is, of, op die sosiologiese vlak, na die leksiko-graaf as sodanig. Die leksikograaf word in navorsing geïmpliseer, maar skynbaar selde bestudeer.

Hierdie onderbestudering van die leksikograaf in die leksikografie blyk 'n wydverspreide verskynsel te wees. In die gesaghebbende *Routledge Handbook of Lexicography* (Fuertes-Olivera 2018) is 47 hoofstukke, verdeel in 5 tematiese afdelings. Nie een van hierdie afdelings handel spesifiek oor die leksikograaf nie. Wat ondersoek word, is die basis van die leksikografie, die interdissipli-nêre aard daarvan, woordeboektipes, vernuwende woordeboeke, en die ver-band tussen die leksikografie, wêreldtale en die internet (ibid.). In nie een van die hoofstuktitels word die leksikograaf by name genoem nie.

Nog meer onlangs, in die *Bloomsbury Handbook of Lexicography* (Jackson 2022), is daar 24 hoofstukke in 3 tematiese afdelings. Heelwat word gesê oor die samestelling van woordeboeke of die leksikografiepraktyk as sodanig; die toe-koms van die leksikografie en woordeboeke word ook onder die loep geneem. Daar is egter ook hier geen bydrae wat spesifiek die leksikograaf as noodwen-dige betrokkene ondersoek of bespreek nie.

Hierdie vreemde verskynsel is nie beperk tot die hoofstroom- Engelse vakleksikografiese bronne nie. In De Gruyter se *Lexicographica. Series Maior* kom eweneens nie volumes voor wat spesifiek die leksikograaf sosiologies of sosiaal-maatskaplik ondersoek nie (vgl. die volledige lys hier: https://www.degruyter.com/serial/lexsm-b/html#volumes). In die afgelope 10 jaar het daar 21 boeke in hierdie gesaghebbende boekreeks verskyn, hoofsaaklik in Engels en Duits. Die hoofonderwerpe in hierdie boekreeks is die metaleksikografie (4 volumes), histo-riese ondersoeke (4 volumes), die semantiek of semiotiek (3 volumes) en die woorde-boekdidaktiek (3 volumes). Een boek noem samestellers van mediese woorde-boeke (McConchie 2019), maar verder maak die leksikograaf ook hier nie die onderwerp uit nie.

Rufus Gouws wys hierdie problematiek reeds 'n dekade gelede uit in 'n artikel in *Lexikos*. Volgens Gouws (2012: 218) bestryk mense se betrokkenheid by die leksikografie 'n wye verskeidenheid aktiwiteite. Hierdie mense het miskien te make met nietradisionele aktiwiteite soos die ontwikkeling van rekenaarpro-grammatuur, proefleeswerk en dies meer, asook die maak van nietradisionele — maar in wese leksikografiese — bronne soos ensiklopedieë (2012: 220-221), benewens die tradisionele maak van woordeboeke.

Die vraag is dan (ibid.) of hierdie mense wel "leksikograwe" genoem kan word. Volgens Gouws (2012: 223) kan énige iemand wat "woordeboeke skryf",

'n leksikograaf genoem word. Daar is byvoorbeeld diegene wat praktiese leksikografie beoefen sonder dat hulle enige teoretiese agtergrond het, diegene wat teoreties besig is sonder om praktiese werk te lewer, en 'n groep wat effe meer na een van hierdie kante neig hoewel hulle ook iets van die ander kant doen (Gouws 2006: 75).

Die saak word verder gekompliseer deur die opkoms in formele opleidingsprogramme — Gouws (2012: 219) vra of persone met hierdie opleiding "leksikograwe" is alvorens hulle self prakties aan woordeboekprojekte gewerk het. Hierin word dus 'n tweedeling geïmpliseer: Daar kan iets soos 'n "praktiese leksikograaf" wees, en ook iets soos 'n "teoretiese leksikograaf". Volgens Gouws (ibid.) vorm hierdie onderskeid in wese twee hoofkomponente van die leksikografie.

As daar vir 'n oomblik afgestap word van woordeboeke en soortgelyke leksikografiese hulpbronne, kom 'n volgende verklaring van hierdie problematiek na vore. Gouws (2012: 222) wys daarop dat daar ook mense is wat handboeke oor die leksikografie skryf, akademiese navorsing daaroor doen en publiseer, en teoretiese modelle vir die maak van woordeboeke ontwikkel. Hierbenewens is hierdie mense tipies by onderrig en opleiding betrokke (ibid.). Is hulle leksikograwe? Gouws (ibid.) noem hierdie mense "metaleksikograwe", maar wys ook dadelik op die feit dat dit onduidelik is watter kriteria eintlik geld om iemand 'n "metaleksikograaf" te kan noem, net soos met "leksikograaf" hierbo. Verder is daar ook nog vrae oor die graad van direkte betrokkenheid by 'n woordeboekprojek wat 'n hiërargie impliseer (Gouws 2012: 223-224).

Dit is vervolgens, volgens Gouws (2012: 224-225), nodig om ondubbelsinnige kriteria te ontwikkel vir wie 'n leksikograaf genoem kan word, omdat hierdie gebrek aan kriteria potensieel die ontwikkeling van die vakgebied aan bande kan lê.

Om hierdie kriteria te help ontwikkel, neem hierdie artikel vir Rufus Gouws onder die loep in die vorm van 'n gevallestudie met 'n onderliggende tematiese analise. Die aanname word eerstens gemaak dat Rufus Gouws wel 'n leksikograaf genoem kan word. Gouws se eie bydrae tot die leksikografie, op metaleksikografiese en teoretiese vlak, maar ook tot die leksikografiese praktyk en leksikografie-opleiding, spreek vanself in terme van produktiwiteit, kundigheid en relevansie. Hy publiseer meer as 200 navorsingsartikels en boekhoofstukke, en meer as 20 boeke, benewens die meer as 100 vakkundige referate wat hy al gelewer het en sy groot getalle afgestudeerde nagraadse studente (*LitNet* 2022), baie van laasgenoemde ook op hulle beurt ten nouste in die leksikografie betrokke.

Die gevallestudie behels 'n literatuuroorsig oor Gouws se navorsingsartikels uit die afgelope 10 jaar sedert die verskyning van die 2012-artikel waarin hy vra wie 'n leksikograaf genoem kan word. Alle artikels waarin hy alleen- of mede-outeur was en wat in geakkrediteerde akademiese tydskrifte verskyn het, is in die ondersoek ingesluit. Die artikels is tematies ontleed en 'n stel rolle word daargestel wat algemeen blyk uit hierdie leksikografiese publikasies. Hierdie rolle of aktiwiteite word indringend bespreek, en uit die bespreking

word uiteindelik 'n voorlopige stel kriteria afgelei vir wie 'n leksikograaf genoem kan word. Hierdie kriteria word voorgestel om die gaping te help vul wat Gouws 10 jaar gelede uitgewys het.

## 2.    Die leksikograaf as woordeboekmaker of leksikografiese praktisyn

In die ontwikkeling, vestiging en voortdurende verandering van die leksikografiepraktyk staan die mense wat in hierdie praktyk werk sentraal, hoewel hulle, soos later bespreek, nie naastenby die nodige akademiese bestudering ondergaan nie. Sonder die maker van 'n leksikografiese produk, kan daar nie 'n produk of praktyk wees nie. Gouws (2016a; 2017: 135) wys dat die leksikografiepraktyk veel ouer as die teorie is en dat die ontwikkeling van die praktyk gekenmerk word deur ingrypende aanpassings en verskeie beduidende oorgange, soos die oorgang van primitiewe woordeboeke op kleitablette en perkament na gedrukte boeke en uiteindelik gevorderde elektroniese produkte.

Een oorgang waarvan die impak nooit geminag mag word nie, is, volgens Gouws (2017: 135), die oorgang vanaf die tematiese na die alfabetiese ordening in woordeboeke. Hierdie oorgang "was […] nie net 'n oorgang met betrekking tot 'n ordeningsmetode nie, maar ook 'n oorgang in die manier waarop woordeboeke as houers van kennis gesien, beplan en saamgestel is" (ibid.). In die aanvang van die leksikografiepraktyk het woordeboekmakers slegs gepoog om die taal vas te vang en weer te gee, terwyl die tematiese geordendheid van vroeë woordeboeke, veral in die Middeleeue, ook kennis oorgedra het (Gouws 2016b: 358). Die skolastici wat die praktyk aanvanklik ontgin het, wou "kennis organiseer, sorteer en klassifiseer, en hulle het gepoog om die sogenaamde *omne scibile* (= alles wat wetenswaardig is) vas te vang" (Gouws 2016b: 358). Hierdie oordrag van wetenswaardige kennis is steeds 'n belangrike mandaat in die moderne leksikografiese praktyk, maar die taalkundige en leksikografiese behoeftes van die gebruiker is die eerste prioriteit. Die evolusie van die praktyk het, saam met die uiteindelike vestiging van die metaleksikografie en die onlangse digitale revolusie, vir, *inter alia*, woordeboektipologieë, spesifieke teikengebruikers en sinvolle raamwerke of riglyne vir die saamstel van woordeboeke gesorg (vgl. Gouws 2015a, 2016a, 2018b). Dit is hierdie dinamiese, veranderlike aard van die praktyk, sowel as die teorie, wat veroorsaak dat die leksikograaf telkens aanpasbaar moet wees (Gouws 2018c: 225) en nie los van die leksikografiese praktyk kan staan nie.

Gouws (2020b: 4) beklemtoon dat 'n goeie woordeboek gekenmerk word deur die getroue uitoefening en nakoming van 'n deurdagte, sistematiese woordeboekplan, veral een wat op gevestigde metaleksikografiese modelle of beginsels berus[1]. Dit beteken dat die leksikograaf, ongeag die medium, telkens bewus moet wees van die besondere kenmerke van die beplande woordeboek, soos die funksies, inhoud en struktuur, asook die teikengebruikers wat daardie woordeboek moet raadpleeg, en hulle behoeftes (Gouws 2018c: 226; 2022b: 97). 'n Leksikograaf wat bewus is van veral woordeboekstrukture, en die verskeie

maniere om hulle te gebruik, sal aldus Gouws (2022a: 113) 'n beter woordeboek skep wat data toegankliker vir die gebruiker maak (vgl. ook Gouws 2018c).

Praktiese leksikograwe neem té gereeld nie kennis van die diverse strukture wat in woordeboeke gebruik (kan) word nie (Tarp en Gouws 2020). Daar word dikwels net op enkele strukture soos die makro- of mikrostruktuur gefikseer (ibid.). Die aanbieding van en toegang tot data in 'n leksikografiese produk is egter 'n bepalende faktor in die geslaagdheid daarvan, nie net vir die gebruiker nie maar ook vir die leksikograaf. Woordeboeke is mensgerigte gebruiksinstrumente wat so ontwerp moet wees dat dit die mens wat dit uiteindelik gebruik, se behoeftes kan bevredig (Tarp en Gouws 2020: 472; Bothma en Gouws 2022). Dit is fundamenteel dat 'n leksikograaf noukeurig moet wees in die sifting, opneem en aanbieding van data in 'n woordeboek, juis sodat die gebruiker nie die skuld kry as 'n hulle nie die nodige inligting kan opspoor of onttrek uit die woordeboek nie (Tarp en Gouws 2020: 494-495). Byvoorbeeld, selfs iets so elementêr soos die aanbieding van sinonieme in tweetalige woordeboeke moet van so 'n aard wees dat die gebruiker binne die bepaalde konteks maklik en vinnig die gepaste sinoniem moet kan opspoor (Gouws 2013a: 347). Gebrekkige data-aanbieding of toegangsroetes kan 'n gebruiker frustreer, in die verleentheid laat, of die raadplegingsproses onnodig vertraag (vgl. Gouws 2018a). Dit geld nie net vir die data wat in die sentrale teks, die data tipies tussen die alfabettrajekte van A tot Z, opgeneem is nie. Gouws (2018a: 186-187) stel daarom tereg die volgende:

> [I]n die benutting van leksikografiese korpora, die saamstel van die databasis en die ontrekking van data aan die databasis om as aanduiders in woordeboekartikels of inskrywings in buitekomponente te dien, speel die leksikograaf die bepalende rol. Die leksikograaf maak naamlik die keuse, weliswaar gebaseer op die vasgestelde behoeftes van 'n geïdentifiseerde teikengebruiker. In gedrukte woordeboeke is daar 'n vaste en statiese verhouding tussen gebruiker en leksikograaf met 'n eenrigtingverskaffing van data. [...] Vir woordeboeke in die dinamiese aanlyn omgewing word 'n vergelykbare verhouding gehandhaaf van die leksikograaf wat die verskaffer van data is deur keuses uit 'n korpus te maak en met behulp van 'n spesifieke databasisprogram die data 'n neerslagplek in woordeboekartikels of in buitekomponente te laat vind.

Verder voer Klosa en Gouws (2015: 171) aan dat die leksikograaf nie net op die woordeboekdata en die aanbieding van daardie data kan fokus nie, maar dat die skakels na relevante eksterne bronne of data 'n noodsaaklikheid word. Gouws (2021b: 61) eggo hierdie gedagte ten opsigte van woordeboekportale: "die toegang tot en benutting van ander woordeboeke in dieselfde soekportaal stel gebruikers aan bykomende leksikografiese data bekend, maar 'n nog belangriker aspek is die toegang tot nieleksikografiese bronne soos die internet." Begrip van en kundigheid oor leksikografiese sowel as nieleksikografiese data is daarom uiters belangrik in die samestelling van enige leksikografiese produk — dit is 'n grondbeginsel waarvolgens die praktiese leksikograaf moet werk. Bothma en Gouws (2022: 54) noem dat dit dikwels die ekstra-leksikografiese

situasies is waarin gebruikers hulle bevind wat maak dat daar verskillende raadplegingsprosedures gevolg word. Soos genoem, moet die leksikograaf bewus wees van die raadplegingsprosedures van gebruikers maar ook van die navorsingsmoontlikhede wat 'n bepaalde produk kan bied. Die leksikograaf moet juis in die digitale era waar gebruikers maklik oorweldig kan word deur data-oorlading (vgl. Gouws en Tarp 2017), noukeurig met eksterne materiaal omgaan om uiteindelik die minder kundige of bekwame gebruiker ter wille te wees.

Gouws en Tarp (2017: 391) wys daarop dat die voortdurende vooruitgang van moderne tegnologie en die oorgang na elektroniese media, nuwe werktuie en werksmetodes inlei vir die skep, aanbieding of ontwikkeling van leksikografiese produkte of platforms. Die feit dat daar in die onlangse verlede, eintlik steeds, te veel e-woordeboeke is wat nie na wense is nie, dui daarop dat die leksikograaf nie meer net op die erkende of tradisionele leksikografiese proses of tendense steun nie (vgl. Gouws 2014a; Tarp en Gouws 2020). Gouws (2018c: 226) noem dat wanneer die aanlyn leksikografie betrek word, die bestaande leksikografiese praktyke en teorieë wat hoofsaaklik vir gedrukte woordeboeke geskik is steeds gebruik word, maar dit word telkens aangepas of verfyn. Die moderne praktyk dwing die leksikograaf om meer as net woordeboekmaker te wees: Hy of sy moet ook iets weet van koppelvlakontwerp, elektroniese korpusse, taaltegnologie, die bruikbaarheidsbenadering ten opsigte van rekenaars en selfone, en programmatuur (vgl. Bergenholtz et al. 2015, asook afdeling 4). Aangesien tegnologiese innovasie eksponensieel ontwikkel, word hierdie kundigheid en die diversiteit van die leksikograaf se vaardigheid al hoe belangriker (Gouws en Tarp 2017: 391). Die leksikograaf moet bewus wees van die moderne gebruiker se vaardighede en verwagtinge, want gebruikers in die "digitale samelewing" is gewoond daaraan om inligting amper onmiddellik en op verskillende maniere te bekom (Bothma en Gouws 2020: 49). Die feit dat toekomstige gebruikers grotendeels sogenaamde "digital natives" gaan wees, beklemtoon weereens dat die (moderne) leksikograaf meer bedrewe as hulle voorgangers moet wees, veral om die aanbiedingsruimte en toegangsmoontlikhede van die aanlyn of elektroniese omgewing werklik te benut (ibid.). Gouws (2018a: 180) se punt dat woordeboeke werktuie moet wees waarmee lede van veral Generasie Z inligting kan bekom wat aan hulle leksikografiese behoeftes voldoen, is daarom 'n leuse vir die huidige en toekomstige leksikografiese praktyk.

As leksikograwe weet dat hulle teikengebruikers die nodige raadplegingsvaardighede het om hulle woordeboeke te kan gebruik, sal, volgens Tarp en Gouws (2020: 494), 'n woordeboekkultuur in swang wees. Hierdie idee is nie noodwendig in die praktyk verwesenlik nie. Daar bestaan telkens 'n skeiding tussen die leksikograaf en die gebruiker, juis omdat daar verwag word dat die leksikograaf verantwoordelik moet wees vir die maak van die produk en die teikengebruiker vir die gebruik daarvan, ongeag hoe goed of swak die produk en die gebruiker se vaardighede is (Tarp en Gouws 2020: 494). Volgens Gouws (2016b: 361) blyk dit in die (onlangse) leksikografiese praktyk dat die teiken-

gebruikers van 'n woordeboek ook die geleentheid gegun word om deel te neem aan die leksikografiese proses, veral "vir die daarstelling van 'n woordeboek wat op hulle behoeftes gerig is." Bergenholtz et al. (2015: 5) suggereer dat die leksikograaf self ook 'n ervare woordeboekgebruiker is en daarteenoor dat ervare en ingeligte gebruikers hulle ook van die taak van die leksikograaf kan vergewis.

Die leksikografie hoef dus nie altyd beperk te word tot die werk van opgeleide spesialiste nie. Die gebruiker as leksikografiese deelnemer kan die praktyk sowel as die vakgebied verryk. Volgens Gouws (2016b: 361) kan gebruikers besluit om 'n bepaalde woordeboek te maak, of die behoeftes van 'n bepaalde taalgemeenskap of gebruikersgroep noop iemand wat nie 'n opgeleide leksikograaf is nie, om aan die samestelling van 'n woordeboek te begin werk (vgl. ook Gouws 2018a: 189). Dit is sigbaar in projekte soos die *Urban Dictionary* (https://www.urbandictionary.com/) wat volledig deur gebruikers aanlyn saamgestel en onderhou word, of die *N|uu–Nama–Afrikaans–Engels-Woordeboek* (Sands en Jones 2022) wat grootliks deur of met behulp van die N|uu- en Nama-taalgemeenskap saamgestel is. Selfs 'n leksikografiese baanbreker soos die *Woordeboek van die Afrikaans Taal* sou nie sonder die hulp van sy taalgemeenskap en gebruikers kon ontstaan en voortbestaan nie (vgl. Gouws 2021a). Gouws (2016b: 361) noem dat hierdie soort woordeboeke (veral in Afrikaans) as "leksikografiese monumente" gesien kan word en dat hierdie bydraes tot die leksikografiese praktyk nooit onderskat mag word nie.

## 3.     Die leksikograaf as teoretikus en akademiese navorser

Soos reeds bepaal is die leksikografiese praktyk veel ouer as die teorie. Die praktyk het in 'n pre-teoretiese era ontstaan met die nagevolg dat die teorie, oftewel die metaleksikografie, as deel van die breër vakgebied, telkens die praktyk moes inhaal, of dat die teorie grotendeels slegs deur die praktyk beïnvloed is (Klosa en Gouws 2015: 143). Dit is byvoorbeeld sigbaar in die terminologie wat in die praktyk geskep is of die tipologiese onderverdeling van woordeboeke wat uit die praktyk ontstaan het en vandag steeds met vrug in die teorie gebruik word, hoewel dit soms onakkuraat of onkrities gehandhaaf word (Gouws 2015a: 345). Gouws (2016a: 104) noem verder dat daar, met dié dat die praktyk veel ouer as die teorie is, talle aanpassings was wat uitsluitlik op die praktyk gerig en deur praktiese behoeftes bepaal is. In ander gevalle het die leksikografieteorie, of minstens bepaalde vroeë teoretiese insigte, wel 'n rol gespeel. Samuel Johnson se preskriptiewe benadering tot sy *A Dictionary of the English Language* (1755) is, volgens Gouws (2016a: 105), 'n vroeë voorbeeld van waar die teorie uiteindelik die praktyk begin beïnvloed het. Met die vestiging van die leksikografie as dissipline in eie reg (nie as bv. onderafdeling van die taalkunde nie — vgl. Zgusta 1971 en Wiegand 1996), het die metaleksikografie ook erkenning en vergestalting gekry. Daar is wel uiteraard steeds oorvleueling tussen die taalkunde en leksikografie, maar die begrippe of konsepte word

dikwels anders geïnterpreteer of toegepas, en die leksikograaf moet telkens besluit watter interpretasie geldig is (Bergenholtz en Gouws 2013b: 1-2).

Volgens Wiegand et al. (in Gouws 2020b: 3) is die metaleksikografie die oorkoepelende teoretiese domein van die leksikografie wat die teoretiese grondslag vir alle leksikografiese aktiwiteite en prosesse omvat. Gouws (2018a: 180) lys enkele verskillende afdelings van die metaleksikografie, naamlik historiese woordeboeknavorsing, navorsing oor woordeboekkritiek, navorsing oor woordeboekgebruik en sistematiese woordeboeknavorsing. Die metaleksikografie betrek dus, onder andere, navorsing oor al die onderskeie leksikografiese aktiwiteite en prosesse, navorsing oor en opnames van die geskiedenis van die vakgebied, die bestudering van woordeboekgebruikers, asook die kritiese beskouing en beoordeling van die praktyk as geheel en leksikografiese produkte wat daaruit voortvloei (vgl. ook Bergenholtz en Gouws 2016; Gouws 2018b, 2020a).

Dit is ook in die metaleksikografie waar twee van die bekendste en mees gesaghebbende leksikografiese teorieë ontwikkel, verfyn en toegepas word, nl. die Algemene Leksikografieteorie (vgl. Wiegand 1996) en die Funksieteorie (vgl. Tarp 2013). Teoretiese modelle vir navorsing oor of met woordeboeke, of vir die samestelling van bepaalde leksikografiese produkte, berus gereeld op dié twee teorieë en hierdie teorieë word ook tipies aangepas om 'n verskeidenheid leksikografiese probleme te probeer oplos (vgl. Faaß et al. 2014, Gouws 2014a, 2014b, 2018a, en Gouws en Prinsloo 2021a, 2021b). Vanweë hierdie wesenlike rol wat teorie, navorsing en kritiese beoordeling en beskouing in die leksikografie speel, kan daar geargumenteer word dat allerlei teoretici, kritici en navorsers ook leksikograwe kan wees.

In afdeling 5 en 6 word die leksikograaf as pedagoog en taalkundige bespreek. Hierdie rolle oorvleuel in 'n mate met die leksikograaf se rol as teoretikus, veral ten opsigte van die nodige teoretiese kennis en kundigheid rondom die spesialiteitsgebied van die leksikografie. Gouws (2018b) noem tereg dat daar in leksikografiekursusse, tipies op universiteit, 'n aanslag moet wees wat wyer strek as die eng taalkundige blik op woordeboeke. Verder stel Gouws (ibid.) weer dat beplanning en samestelling van woordeboeke nie sonder die nodige teoretiese basis en kundigheid suksesvol kan wees nie. Klosa en Gouws (2015: 164) is daarvan oortuig dat metaleksikograwe verantwoordelik is vir die formulering van teoretiese benaderinge wat die praktyk uiteindelik kan bevoordeel. Dit is ook vanuit die metaleksikografie dat die belang van die identifisering van 'n teikengebruikersgroep en die vasstelling van die leksikografiese behoeftes en naslaanvaardighede van daardie groep as voorvereistes gestel word om uiteindelik suksesvolle woordeboekgebruik te help verseker (Gouws 2016a: 110). Dit geld ook vir woordeboekstrukture (vgl. o.a. Gouws 2018a) en die aanbieding van data m.b.v. daardie strukture (vgl. Gouws en Tarp 2017). Woordeboeke met 'n teoretiese basis is daarom, volgens Gouws (2014b: 483), beter gebruiksinstrumente, en hierdie woordeboeke is streng gerig op die bevrediging van gebruikers se behoeftes.

Dit is egter problematies dat daar steeds woordeboeke of ander leksiko-grafiese produkte ontstaan wat geen teoretiese onderbou het nie, of dat daar produkte is wat té gespesialiseerd is en so ook die (leke)gebruiker in die steek laat (vgl. Gouws 2020a, 2020b). Die belang van navorsing oor en die beplanning en implementering van 'n woordeboekkultuur kan vir die gebruik van en die begrip oor leksikografiese produkte nie onderskat word nie — dit is 'n formele komponent van die metaleksikografie (Gouws 2020b: 5). Die wisselwerking tussen die teorie en die praktyk kan ook daarom nie misken word nie, maar die onus lê by die diegene wat hul met die formulering van die leksikografieteorie bemoei. Gouws (2018a: 179) beklemtoon dat vernuwende denke daarop gerig moet wees om gebruikers en potensiële gebruikers daarvan bewus te hou dat woordeboeke steeds voorkeurbronne vir die oplossing van bepaalde probleem-tipes kan wees.

Soos die praktyk verander en aanpas, so moet die metaleksikografie ook daarby aanpas en voortuitgang maak. Die aanvang van die digitale era het in hierdie verband ernstige implikasies vir die metaleksikograaf (vgl. Gouws 2021a: 1159). Die vroeë ontwikkeling van die aanlyn leksikografie het ongelukkig die pre-teoretiese tydperk van gedrukte woordeboeke nagevolg. Die samestellers van die eerste elektroniese en aanlyn woordeboeke was só ingenome met die moontlikhede van die e-omgewing dat hulle telkens vergeet het van die nodige teoretiese onderbou vir hulle produkte (Klosa en Gouws 2015: 143). Gedrukte woordeboeke, anders as elektroniese woordeboeke, is reeds enigsins verouderd wanneer hulle verskyn (Gouws 2022b: 119). Die teorieë en modelle wat vir e-woordeboeke geld, hoef nie dieselfde roete te volg nie. Al is die gevestigde metaleksikografie op gedrukte woordeboeke gerig, kan metaleksikograwe besluit of hulle hierdie teorieë en modelle vir die e-omgewing kan aan- en toepas, en of daar sal moet besin word of 'n nuwe teoretiese rigting ingeslaan moet word wat slegs geld vir e-produkte (Gouws 2022a: 114). Klosa en Gouws (2015: 143) voel sterk dat die metaleksikografie en metaleksikograwe, ongeag die verkose benadering, pro-aktief moet wees binne die e-omgewing, juis sodat leksikogra-fiese teorie nie weer net "a follower and not a leader" word nie. Gouws (2022a: 112) verduidelik dat daar wel onlangs vordering gemaak is op hierdie front, maar dat aan die een kant, metaleksikograwe nog nie oral die nuwe moontlikhede wat die e-omgewing aanbied, aangegryp het nie, en dat daar aan die ander kant nog nie sinvol oor die innoverende ondernemings in die praktyk nage-dink is nie.

Bothma en Gouws (2020: 53) stel dat dit vanuit 'n leksikografiese perspek-tief duidelik is dat die dinamika van die e-omgewing metaleksikograwe dwing om noukeurig te besin oor die aard en omvang van die relevante leksikogra-fiese prosesse. Dit is reeds sigbaar wanneer die strategieë wat leksikograwe wat e-woordeboeke saamstel en bestudeer, onder die loep geneem word (vgl. Gouws 2014a). 'n Voorbeeld hiervan is die ontwerp en publisering van selfoonwoorde-boeke. Gouws (2016a: 108) demonstreer dat aanlyn woordeboeke wat saam-gestel word met die oog op 'n slimfoon as werktuig se struktuur sal verskil van

dié van 'n woordeboek wat met behulp van 'n rekenaar geraadpleeg word. Hier word bestaande strukture aangepas en verfyn vir 'n baie spesifieke soort werktuig en gebruiker.

Die metaleksikograaf se taak is juis om hierdie nuwe aanpassings in die teorie verder te ontwikkel en omskryf, sodat toekomstige of soortgelyke produkte bevredigend en suksesvol gebruik kan word. Die voortdurende verandering van tegnologie vereis dat die moderne metaleksikograaf ook soos hulle praktiese teenhanger interdissiplinêr te werk moet gaan. Kundigheid oor bv. die rekenaarwetenskap of van bruikbaarheidsbenaderinge is belangrik vir die suksesvolle aanpassing en toepassing van leksikografiese modelle, benaderinge en teorieë in die moderne en toekomstige leksikografiese milieu (vgl. Tarp en Gouws 2020). Die metaleksikografie moet nie medium-spesifiek wees nie. Klosa en Gouws (2015: 144) sluit hierby aan deur aan te voer dat die suksesvolle aanpassing van die (algemene) leksikografieteorie noodsaaklik is vir die uitdagings en vereistes van die e-praktyk, en sodat die metaleksikograaf in 'n posisie is om nuwe en vindingryke maniere aan te beveel vir die beplanning en samestelling van elektroniese leksikografiese produkte.

## 4.    Die leksikograaf as medewerker

'n Wesenskenmerk van die kontemporêre leksikografie is waarskynlik dat diegene wat by die maak van woordeboeke betrokke is, ook nie in isolasie werk nie. Die maak van leksikografiese hulpbronne berus nie by een individu nie, maar kan spanne mense insluit met verskillende kennisvelde, vaardighede en fokuspunte. Tarp en Gouws (2020: 496) verwys na kenners soos inligtingsingenieurs, programmeerders, ontwerpers, gebruikstoetsers en so meer, en bepleit 'n situasie waar hierdie samewerking die norm word. Spanwerk en samewerking is in die digitale era meer as ooit tevore 'n voorvereiste vir suksesvolle leksikografiese werk (Tarp en Gouws 2019: 266).

Gouws (2022b: 97) beklemtoon ook die belang van hierdie wesenlik interdissiplinêre interaksie tussen die metaleksikografie en velde soos die rekenaarwetenskap. Leksikograwe werk saam met vakkundiges uit hierdie veld (ibid.) om woordeboeke te maak, maar hierdie samewerking is ook in die beplanningsfase, oftewel metaleksikografies, belangrik. In die metaleksikografie word onder meer op woordeboekelemente se inhoud en leksikografiese funksie gefokus, wat hierdie interaksie en samewerking se belang beklemtoon (Gouws 2022a: 112). Selfs iets soos grafika is byvoorbeeld van belang — as 'n leksikografiese instrument — en daarom moet die leksikograaf ook by die ontwerp daarvan inspraak hê sodat dit nie bloot aan uitgewers oorgelaat word nie (Gouws et al. 2014: 37).

Dit is egter nie net eenrigtingverkeer waardeur 'n metaleksikografiese oriëntasie kan bydra tot die woordeboekproduk en produksieproses nie. Gouws (2018a: 188) wys uit dat daar ook op grond van ontwikkelinge in die verwante dissiplines, soos die inligtingwetenskap, nuwe werkswyses ontwikkel kan word binne die metaleksikografiese omgewing. Dit kan bydra tot die ontwik-

keling van nuwe stelsels en praktyke (Bothma en Gouws 2020: 54).

Hierdie interdissiplinariteit is in die geval van moderne leksikografiese produkte 'n noodwendigheid (Tarp en Gouws 2020: 476). Benewens mede-kundiges, is daar boonop in die hedendaagse aanlyn omgewing ook die moontlikheid vir deurlopende en oombliklike gebruikerskommentaar en -terugvoer. Hierdie terugvoer kan vinnig ingesamel word en selfs tydens die maak van die woordeboek reeds in ag geneem word (Gouws 2016b: 361).

Hierdie interdissiplinêre tweerigtingverkeer en uiteindelike multidissiplinariteit dra daartoe by dat gebruikers se behoeftes in 'n verskeidenheid kommunikatiewe situasies bevredig kan word. Aangesien die leksikografie primêr gemoeid is met inligtingshulpbronne met data vir gebruikers in kommunikatiewe situasies (Bergenholtz et al. 2015: 3), dien hierdie prosesse en samewerking duidelik en ondubbelsinnig die oorhoofse doel van die leksikografie.

## 5.    Die leksikograaf as pedagoog

Volgens Gouws (2018b) is die woordeboekdidaktiek 'n belangrike element van 'n omvattende woordeboekkultuur, en moet woordeboekgebruik binne so 'n kultuur op 'n praktiese wyse onderrig word, aangesien woordeboeke in wese praktiese gebruiksinstrumente is. Die doelwit hiermee is nie net die bevordering van praktiese woordeboekgebruik as sodanig nie, maar ook die beplanning van 'n geskikte kurrikulum (ibid.).

Leksikografiese onderrig, net soos die leksikografie self en die aard van die leksikograaf, is 'n veelvlakkige saak. Dit kan hoofsaaklik verdeel word in die onderrig van nuwe leksikograwe en die onderrig van woordeboekgebruikers.

In die eerste geval, die onderrig van nuwe leksikograwe, moet 'n leksikografiekursus ten eerste 'n oorsig bevat van ontwikkelinge in die praktiese sowel as die teoretiese leksikografie (Gouws 2018b). Daar moet van 'n algemene leksikografieteorie gebruik gemaak word, wat volgens Gouws (2018b) ook belangrik is by die skakeling van die leksikografie-opleiding as sodanig met verwante vakgebiede soos die breëre taalkunde en die inligtingwetenskap.

'n Groot komponent van Gouws se werk fokus egter ook op die tweede deel, naamlik die onderrig van (veral nuwe) woordeboekgebruikers as onderdeel van die ontwikkeling van 'n nasionale (en omvattende) woordeboekkultuur. Hierdie onderrig moet eerstens wyer strek as die tradisionele gedrukte woordeboek (of elektroniese weergawes van hierdie soort produk), want daar is 'n groot verskeidenheid aanlyn hulpbronne wat as nietradisionele woordeboeke getipeer kan word. Volgens Gouws (2016a: 118) moet woordeboekvaardighede ook opleiding in die gebruik van hierdie andersoortige aanlyn naslaanbronne bewerkstellig; verder moet dit nie aan 'n enkele medium gekoppel (hoef te) word nie, en hier is die waarde van 'n algemene leksikografieteorie weereens ooglopend (Gouws 2018b). Sodanige opleiding moet verder veral ook tred hou met die gebruiker se behoeftes: Gouws et al. (2014: 38) beklemtoon byvoor-

beeld dat daar nie sommer net gebrekkige inligting in 'n woordeboek gegee kan word omdat dit byvoorbeeld op grondslagfaseleerders gemik is nie.

In hierdie soort ideale situasie wat deur Gouws et al. (2014) geskets word, moet die persoon wat ander opvoed in woordeboekgebruik, byvoorbeeld onderwysers en dosente, ook bekend wees met die eiesoortige kenmerke of eienskappe van die woordeboeke ter sprake. Hulle leer studente en skoolkinders nie soseer die inhoud van die woordeboek nie, maar eintlik 'n stel navorsings- en naslaanvaardighede wat uiteindelik aanleiding gee tot suksesvolle woordeboekgebruik, en potensieel tot meer geslaagde navorsing as sodanig. Om hierdie rede is dit onontbeerlik dat hierdie opvoeders en dosente sélf ook goed onderlê moet wees in die leksikografie (Gouws et al. 2014).

Selfs so vroeg as in die grondslagfase waar leerders heel moontlik die eerste keer met woordeboeke kennis maak, is dus 'n groter onderliggende proses in werking wat berus op die kundigheid van die opleier. Gouws et al. (2014: 40-41) sit uiteen hoe 'n benadering wat kinders reeds hier op 'n goed-gefundeerde en kundige manier aan woordeboeke bekendstel, nie net oor woordeboeke as sodanig gaan nie — dit gaan weliswaar oor die woordeboek as 'n integrale deel in die lewenslange leerproses.

Dit blyk ooglopend uit Gouws se gepubliseerde navorsing dat die rol van pedagoog 'n belangrike element van die leksikograaf se taak is, hetsy vir die opleiding van 'n volgende generasie praktiese of teoretiese leksikograwe, of vir die opleiding van die gebruikers en die vestiging van 'n woordeboekkultuur.

## 6.     Die leksikograaf as taalkundige en linguis

In een van sy mees onlangse artikels skryf Gouws (2022b) onderhoudend oor ruimtelike en temporele deiktiese verankering in die leksikografie. Hierdie soort navorsing beklemtoon die feit dat die leksikograaf ook tipies 'n spreekwoordelike klap van die taalkunde weg het, hetsy taalwetenskaplik (algemeen) of taalkundig (van 'n spesifieke taalkunde), wat vir die doeleindes van hierdie rolbeskrywing saamgegroepeer kan word. Woordeboeke bestaan uit materiaal wat as taalkundige data geklassifiseer kan word, maar volgens Gouws (2022b: 98) sluit dit hierbenewens ook in die benutting, in leksikografiese vorm, van die taalkunde se begrippe-apparaat, veral in die "meta-taal" (2022b: 99) van woordeboeke.

Dit is uiteraard belangrik dat diegene wat hulle met die leksikografie besig hou, bepaalde taal- en taalkundekennis moet hê — Gouws (2022b: 98) noem samewerking tussen die linguistiek en die leksikografie "noodsaaklik". Daar moet byvoorbeeld sinchronies en diachronies na taalgebruik gekyk (kan) word (Gouws 2013b: 138). Sekere begrippe uit die taalkunde staan ook sentraal in die leksikografie, waarvan polisemie en homonimie sekerlik van die bekendstes is. Gouws (2018b) beskou hierdie soort inhoude as onontbeerlike kennis vir leksikografiestudente. Hierdie soort begrippe word ook in die leksikografie geproblematiseer en verryk (Bergenholtz en Gouws 2017) en voed daardeur weer die

taalkundekennis — die leksikografie voed uiteindelik ook die linguistiek.

Die taalkundige omgewing binne die woordeboek is vervolgens ooglopend belangrik vir die woorde daarin. Die leksikograaf moet egter nie net na woorde as sodanig kyk nie, maar ook na ander taalbousels (vgl. Bergenholtz en Gouws 2013a). Bergenholtz en Gouws (2013b: 2) verwys byvoorbeeld na items wat uit meer as een woord bestaan, sowel as items op die subwoordelike vlak. Om woordeboekprodukte te skep wat op hierdie veelvlakkige wyse nuttig kan wees, is 'n stuk beplanning, en wel beplanning aan die hand van taalkundige kennis, noodsaaklik. Daar moet byvoorbeeld gepaste klassifikasiemeganismes wees om hierdie soort items op die regte manier te hanteer, eerstens in die databasis waaruit die woordeboek saamgestel word, en ook in die woordeboek as sodanig (Bergenholtz en Gouws 2013b: 17).

Volgens Gouws (2014b: 502) is die uiteindelike onttrekking van inligting uit die woordeboek ten nouste gekoppel aan goeie beplanning in die samestelling van die databasis waaruit die data kom. Hierdeur word dan ook nie net 'n gedekontekstualiseerde woord weergegee nie, maar verder 'n stuk kotekstuele leiding verskaf. Gouws (2015b: 171, 180) beklemtoon die belang daarvan om tipiese gebruik te illustreer binne die parameters van die bepaalde taal se sintaksis. Vir hierdie aspekte van die woordeboekmakery is eweneens grondige taalkundekennis nodig.

Die taalkundige omgewing *buite* die woordeboek is egter ook nie vir die leksikograaf van min belang nie. Volgens Gouws (2020b: 4) moet 'n uitdrukking se betekenis en gebruik behoorlik ingebed wees by 'n kulturele konteks, anders is daar die gevaar dat die gebruiker se onttrekking van inligting uit die woordeboek suboptimaal kan wees. Die leksikograaf moet dus nie net na die taalkundige eienskappe van 'n sekere taalvorm kyk nie (ibid.), hoewel dit belangrik is, maar ook na die kultuuromgewing waarbinne dit voorkom en na die moontlike buitetalige inhoud wat daardeur oorgedra (kan) word. (Vgl. in hierdie verband ook afdeling 7.)

Dit gaan dus oor sowel taalkennis as interpretatiewe kennis. Volgens Bergenholtz et al. (2015: 2) is kommunikatiewe kennis en vaardighede uiteraard belangrik, maar daarbenewens is daar ook operatiewe kennis en vaardighede, en interpretatiewe kennis waardeur bepaalde nietalige tekens verstaan kan word. Al hierdie faktore is vir die leksikograaf van belang, en kan deur 'n taalkundige onderbou ondersteun word. Met woordeboeke wat in die sentrale woordelys leksikografiese bewerkings bied wat die soort data daarstel waaraan byvoorbeeld taalpraktisyns 'n behoefte het (Gouws 2021b: 54) word die gebruiker se behoefte vervul op grond van die leksikograaf se oorwegings op taalkundige vlak.

'n Verdere belangrike funksie van taalkundige opleiding en kennis, is dat dit (aldus Gouws 2018b) kan bydra tot die "belangrike intellektuele aktiwiteit van woordeboekkritiek". Woordeboekkritiek, in die vorm van sowel resensies as portuurbeoordeling, hou volgens Bergenholtz en Gouws (2015: 107) aansienlike waarde vir outeurs in, waaronder die belangrikste sekerlik 'n stuk gehalte-

versekering is. Die woordeboekkritiek omvat 'n breë leksikografiese aktiwiteit wat nie net die woordeboekproduk self insluit nie, maar ook metaleksikografiese literatuur (Bergenholtz en Gouws 2016: 62). Hierdie kritiek situeer die betrokke werk binne 'n spesifieke wetenskaplike gemeenskap (Bergenholtz en Gouws 2015: 109) en dra dus ook by tot die daarstel en ontwikkeling van die veld as geheel. Hiervoor is kundigheid nodig; kundigheid wat heel spesifieke kriteria skep en operasionaliseer (ibid.) — en dit is veral en spesifiek taalkundige kundigheid. Daar is reeds in afdeling 3 verwys na die belang van die metaleksikografie vir die leksikograaf. Die koppeling met hierdie afdeling blyk ook duidelik.

Vir die leksikograaf om los van enige taalkunde te probeer werk, is soos vir 'n soogdier om onder die water te probeer asemhaal, of vir 'n vis op land.

## 7.    Die leksikograaf as taalburger en aktivis

Die leksikograaf se werk word toenemend nie beperk tot dit wat direk woordeboekintern of binne die leksikografiepraktyk gedoen word nie, maar kan ook 'n aktivistiese dimensie aanneem. Gouws (2016b: 137) verwys na woordeboeke wat "monumente [word] ter erkenning van en ter ere van 'n taal". Die werk wat op die oppervlak beskou vir die gebruiker gedoen word en oor die gebruiker handel, het die newe-effek dat dit ook die taal, en per implikasie die taalgemeenskap as sodanig, tot voordeel strek.

In die domein van taalbeplanning is statusbeplanning een van die belangrike konseptuele vlakke waarop gefokus word. Dit betref die relatiewe posisie van een taal ten opsigte van 'n ander binne die tradisionele oriëntasie van 'n nasiestaat wat sekere tale ampstaalstatus gee. Dit slaan egter ook op die status van tale relatief tot mekaar, en relatief tot die gebruiksdomeine waarin hulle voorkom. Gouws (2016b: 358, 360) verwys na die belang van byvoorbeeld vakwoordeboeke in die status van Afrikaans as 'n taal wat vir wetenskapsbeoefening geskik is, en derhalwe ook vir ander gebruike, sowel op hoëfunksie- as die meer algemene domein. Aan woordeboeke word immers in die oë van die gebruikerspubliek "gesag" (ibid.) toegeken. In die gekontesteerde Suid-Afrikaanse taalwerklikheid met durende en skreiende taalongelykhede moet die belang van die leksikograaf as bemiddelaar in 'n konteks van ongelyke magsverhoudinge dus nie geringgeskat word nie.

Hier is die rol van die leksikograaf dus ook 'n sosiaal betrokke een. In die geval van 'n omvattende woordeboek moet daar byvoorbeeld na die "hele geografiese omgewing" (Gouws 2022b: 102) van die teikengebruikersgroep gekyk word, en waar dit 'n sinchrone fokus het, moet die "verskillende variëteite" (Gouws 2022b: 110) van die betrokke taal in ag geneem en weerspieël word. Te gereeld is en was woordeboeke uitsluitend. Volgens Gouws (2022d: 50) is daar byvoorbeeld histories in verskeie Afrikaanse woordeboeke weens lengtebeperkings slegs op die standaardvariëteit gefokus, tot nadeel van ander variëteite — 'n situasie wat daartoe aanleiding gegee het dat so 'n woordeboek nie optimaal

bruikbaar is in die leer- en onderrigkonteks nie. Hierbenewens gee dit ook aanleiding tot ernstige etiese vraagstukke.

Hierdie sosiale betrokkenheid impliseer ook 'n sensitiwiteit vir gebruikers se vermoëns, en nie net hul behoeftes nie. Gouws (2018b) bepleit die erkenning van die nosie van 'n "woordeboekkultuur" binne die terrein van woordeboeknavorsing, waardeur die leksikografie ook by veranderende omstandighede kan aanpas. So 'n kultuur kan nie gevestig word in 'n konteks waar leksikografiese produkte nie voldoende op die gebruiker toegespits is nie (Tarp en Gouws 2020: 488). Die nosie van 'n "woordeboekkultuur" is natuurlik nie onproblematies nie. Gouws (2016b: 361) sistematiseer die nosies van "woordeboekkultuur" en "gebruikersvriendelikheid" as twee punte van 'n spektrum. Onder die noodsaak van 'n "woordeboekkultuur" word implisiet gesê dat die gebruiker by die leksikografie moet aanpas, terwyl "gebruikersvriendelikheid" daarna verwys dat die leksikografie by die gebruiker moet aanpas (ibid.). Leksikograwe misken volgens Gouws (2020: 3) nog te veel die noodwendig interaktiewe verband wat daar tussen leksikografie en kultuur behoort te bestaan wanneer hulle woordeboeke maak.

In die maak van woordeboeke lewer die leksikograaf immers nie slegs linguistiese kommentaar nie; dit is volgens Gouws (2022d: 40) ook 'n geleentheid waar dominante ideologieë, onderliggende houdings, denkwyses en dies meer weerspieël word. Woordeboeke affekteer mense se gedrag, en woordeboeke het daarom mag (Gouws 2022d: 59). Met ander woorde, die leksikografiese produk kan, op dieselfde manier as die heersende kultuur, inklusief of eksklusief wees, en dit is iets waarvan leksikograwe moet kennis neem, met potensiële gevare waarteen gewaak moet word. Van die vroegste Afrika-woordeboeke is byvoorbeeld deur (eksterne) agente soos sendelinge, koloniale administrateurs of taalkundiges geskep (Gouws 2022d: 44), en nie deur die gemeenskappe aan wie hierdie normeringsbronne uiteindelik oorgedra is nie. Dit is nie die ideaal nie, omdat 'n reeds ongelyke magsituasie dan ook in die woordeboek uitgebeeld, genormaliseer en uiteindelik geperpetueer word.

Leksikograwe is ook bloot menslik, en daarom moet hulle bewus daaraan aandag gee om onbevooroordeeld en objektief te wees in dit wat Gouws (2022d: 41) sistematiseer as die seleksie, aanbieding en behandeling van uitdrukkings. Die leksikograaf is inherent besig om die twee pole op die reeds vermelde spektrum tussen "maak gebruikers vir die woordeboek" en "maak woordeboeke vir die gebruikers" met mekaar te probeer versoen.

Die reikwydte van die leksikografie, en dus die leksikograaf, strek egter ook verder. Volgens Gouws (2022c) dien 'n taal wat as wetenskapstaal gevestig is, nie slegs sy eie taalgemeenskap nie, maar dra dit ook by tot die internasionale statuur van daardie taal. Afrikaans beskik oor 'n "gevestigde leksikografiese vaktaal" (ibid.) — in geen geringe mate nie te danke aan die werk van metaleksikograwe wat in Afrikaans aktief is, waaronder Rufus Gouws seer sekerlik die leiersfiguur is. Dit is 'n stuk geslaagde taalaktivisme met 'n wye reikwydte, gebou op 'n stewige vakkundige basis.

Hierdie taalaktivisme plaas Afrikaans midde-in 'n vakkundige diskoers wat, volgens Gouws (2022c) ook Bulgaars, Engels, Frans, Hongaars, Italiaans, Portugees, Russies en Spaans insluit. Wanneer hierdie lysie tale beskou word, is dit maklik om spore te sien van die soort epistemiese ongeregtigheid wat Stroud en Kerfoot (2021: 20) bekla:

> Both within and outside of universities worldwide, and South Africa in particular, languages and languaging practices continue to be ranked and regulated in ways that privilege communication and knowledge production through European languages — with insidious consequences such as educational failure.

Afrikaans staan egter tussen hierdie tale, en die Afrikaanse vakleksikografie maak sodoende die Afrikaanse taalgemeenskap in die breë deel van die groter vakkundige wêreld, en vestig ook daardeur die middele en die vermoë om leksikografiese produkte vir en deur die hele taalgemeenskap te laat skep.

Hoewel woordeboeke in die verlede ook as instrumente van beheer, afgrensing en onderdrukking misbruik is (vgl. Stroud en Kerfoot 2021: 24), is die soort werk wat gedoen is en word om Afrikaans 'n leksikografiese register te gee, 'n goeie voorbeeld van die "forms of meaning-making able to transcend the constraints of fixed understandings of legitimised language" wat Stroud en Kerfoot (2021: 34) bepleit as 'n wenslike vorm van taalburgerskap. 'n Leksikografiese register in Afrikaans is uiters bruikbaar in die "re-borderings, processes of equivocal translation" (Stroud en Kerfoot 2021: 37) wat gesien word as onontbeerlik in 'n taalburgerskapsbenadering tot groter epistemiese geregtigheid vir gebruikers van gemarginaliseerde vorme, oftewel variëteite, van Afrikaans.

Hoewel Stroud en Kerfoot (2021) krities is oor nosies van tale as opvoedkundige hulpbronne, en waarsku teen die inherent koloniale wêreldbeeld wat ingebed is in sienings van tale as streng afgebakende afsonderlike entiteite en die implikasie dat tale geïntellektualiseer moet word alvorens hulle tot hoëfunksiegebruiksdomeine kan toetree, is dit ooglopend dat die taalaktivisme wat beoefen word deur Afrikaans metaleksikografies relevant te maak, oorspoeleffekte het wat vir die totale Afrikaanse gemeenskap van waarde is en nuwe vorme van semiotisering en uitdrukking moontlik maak. Weereens is dit die prototipiese definisie van taalburgerskap; 'n taalburgerskap wat hier beliggaam word deur die leksikograaf.

In hierdie verband is Gouws (2016c) se artikel oor ANNA belangrik, en 'n interessante geval om te ondersoek. Dié artikel handel oor Prisma se *Groot woordeboek Afrikaans en Nederlands.* Met die eerste oogopslag is dit maklik moontlik om te dink dat hier al weer 'n hele woordeboek is wat die historiese verband tussen Afrikaans en Nederlands oorbeklemtoon, ten koste van die erkenning van die ander invloede in die taal se ontstaansgeskiedenis. Dit is egter opmerklik dat hierdie woordeboek "veral op die verskille tussen die twee tale [fokus]" (Gouws 2016c: 59). In hierdie klein opmerking sit weereens 'n groot stuk taalburgerskap opgesluit. Daar word erken en beklemtoon dat daar

vele *verskille* tussen die twee tale is, wat 'n noemenswaardige punt maak oor populêre sienings dat Afrikaans bloot 'n dogtertaal van Nederlands is.

Gouws (2016c: 75) skets naamlik 'n historiese trajek waar Afrikaans aanvanklik as wordende jong taal *deels* deur middel van koördinering met Nederlands gestandaardiseer is, maar daardie koördinering het juis 'n nuwe taal tot stand help bring. Op dieselfde manier is daar hedendaagse leksikografiese produkte waar Afrikaans hierdie rol vervul vir van die oudste en mees bedreigde tale van Suid-Afrika. In die geval van Gouws (2016c: 75) gaan dit naamlik oor "Ju|'hoan, 'n noordelike Khoesantaal, en N|uu, 'n nie-Bantoe-kliektaal wat deel van die !Ui-familie is, as niegestandaardiseerde tale wat vir hulle vroeë leksikografiese toetrede afhanklik is van die mede-optrede van 'n gestandaardiseerde taal". Afrikaans verskaf hier die leksikografiese skouers waarop "van die mees bedreigde tale" (Gouws 2016c: 76) kan staan deurdat dit die enkele sprekers van hierdie tale se "verkeerstaal" (Gouws 2016c: 77) is.

Dit vergestalt leksikografie as taalburgerskap en aktivisme, wat durende epistemiese ongeregtighede by die spreekwoordelike horings pak.

## 8.      Gevolgtrekking

Die bovermelde ses onderverdelings is verkry deur middel van 'n grondige tematiese analise van Rufus Gouws se artikels wat oor die afgelope dekade in geakkrediteerde vaktydskrifte gepubliseer is. Gouws (2012: 224-225) se versugting is na ondubbelsinnige kriteria vir wie 'n leksikograaf genoem kan word.

Die belang van hierdie soort kriteria moenie geringgeskat word nie. Daar is reeds bestaande wetgewing oor die gebruik van die Suid-Afrikaanse amptstale (Wet 12 van 2012), asook oor die professionalisering van die taalpraktyk deur die instelling van 'n taalpraktisynsraad (Wet 8 van 2014), en duidelike stappe in die rigting van groter regulering van die taalpraktyk deur middel van 'n proses van akkreditasie vir sulke praktisyns (vgl. Adams 2023), wat noodwendig effekte op die leksikografie ook sal hê.

Vanuit die bovermelde tematiese analise kom daar ses belangrike rolle vir die leksikograaf na vore, deurdat die leksikograaf tipies in etlike van hierdie ses rolle optree, wat hier in die volgorde van die afdelings hierbo gelys word:

— Leksikografiese praktisyn
— Teoretikus en akademiese navorser
— Medewerker
— Pedagoog
— Taalkundige en linguis
— Taalburger en aktivis

Hoewel die rol van die leksikograaf as pedagoog nie geringgeskat moet word nie, is dit die een van die ses rolle wat waarskynlik die mins noodsaaklike is om afsonderlik te definieer en as kriterium te stel. Hierdie rol word in die

ondersoekte literatuur beklemtoon, hoewel dit onses insiens nie absoluut nood-
saaklik is nie en eintlik meer indirek figureer, of in baie gevalle implisiet inge-
sluit word by die eerste en tweede rolle. Leksikografiese pedagogie vind nie
slegs in formele opleidingsprogramme plaas nie, maar ook in die praktyk-
omgewing waar praktiserende leksikograwe byvoorbeeld hul juniors deurlo-
pend oplei. Die metaleksikografie-vereiste waarna hieronder verwys word,
ondervang ook grotendeels die vereiste dat die leksikograaf op die gebruiker
ingestel moet wees. 'n Bereidwilligheid tot hierdie ingesteldhede kan wel as
belangrike onderliggende etiese beginsel uitgelig word. Dit is tegnies heeltemal
moontlik om die ander vyf rolle te vervul sonder om direk by die woordeboek-
pedagogiek betrokke te wees, maar dit is onwaarskynlik dat die werkende lek-
sikograaf nie in elk geval informeel met opleiding doenig sal wees nie.

Daar bly dus vyf rolle oor wat die leksikograaf nie behoort te kan ontwyk
as daar sinvol en logies nagedink word oor wat dit behels om by die leksiko-
grafie betrokke te wees in Suid-Afrika, maar ook internasionaal, nie. Geen
woordeboek word van begin tot einde gemaak sonder om saam met ander
mense te werk nie (rol 3), hetsy op praktiese of metaleksikografiese vlak. Dit
maak nie saak wat die aard van die leksikografiese bedrywigheid is nie, daar
gaan met taal gewerk word, en daarom is taalkundige kennis belangrik (rol 5).
Dit blyk ook duidelik dat "teorie ter wille van teorie" nie voldoende of wenslik
is nie, terwyl die praktiese leksikografie eweneens die rigtinggewende rol van die
metaleksikografie moet in ag neem; daar is dus 'n onlosmaaklike en veronder-
stelde verband tussen die praktiese en die teoretiese leksikografie (rolle 1 en 2).

Rol 6 beweeg sterker op die terrein van 'n stuk leksikografiese etiek. Etiese
beginsels is logieserwys dwarsdeur die leksikografie van belang, maar die akti-
vistiese kant van die leksikografie word in die huidige era toenemend belang-
rik, gedagtig aan die maatskaplike werklikhede van die tyd. Woordeboeke is
die weerspieëling van die taalgemeenskap se verhoudinge met mekaar en met
ander, van benoemingskonvensies, voornaamwoorde, taboes, en van die kul-
tuur as sodanig. Dit is reeds onontbeerlik, en sal toenemend selfs dringender
raak, dat die leksikograaf hiervoor sensitief moet wees en met groot omsigtig-
heid daarmee te werk moet gaan. Die leksikograaf moet bewus wees van die
noodwendig politiese aard van enige werk met taal in die hedendaagse samele-
wing, en daarom bewus aktivisties optree. Daar kan nie geskuil word agter die
wanindruk dat daar "slegs met taal" of "slegs met woordeboeke" gewerk word
nie. Om met taal en met woordeboeke te werk, is om 'n effek op mense en op
die wêreld te hê, en die ware leksikograaf neem hiervan kennis.

Hieruit kan afgelei word dat daar vyf minimumvereistes waaraan die heden-
daagse leksikograaf behoort te voldoen. Die leksikograaf is iemand wat:

— prakties by die maak van leksikografiese produkte betrokke is;
— kennis neem van ontwikkelinge in die meta- of teoretiese leksikografie,
   hetsy deur formele opleiding daarin of deur ander kanale;
— 'n stuk taalkundige of taalwetenskaplike agtergrond het;

— kundigheid het in verwante velde soos die inligtingwetenskap, of gerede-like toegang het tot sulke kundigheid deur medewerkers; en

— bewus is van hulle eie posisie en mag as taalpolitiese agent, en dit eties operasionaliseer aan die hand van 'n regstellende ingesteldheid.

Die bovermelde vyf kriteria is 'n eerste poging om die vraag "Wie is die leksi-kograaf?" te beantwoord vanuit tematiese data uit die werk van een van die grootste leksikograwe van ons tyd, en sekerlik die grootste een in hierdie taal.

## Eindnoot

1. Gouws se eie bydraes as praktiese leksikograaf ondersteun hierdie gedagte, vgl. o.a. sy werk aan die *Verklarende Handwoordeboek van die Afrikaanse Taal* (Odendal en Gouws 2005; Luther et al. 2015), die *Basiswoordeboek van Afrikaans* (Gouws et al. 1994) en die *Longman Grondslagfasewoordeboek Afrikaans/English* (Gouws et al. 2010).

## Bronnelys

**Adams, N.** 2023. UWC Professor Assumes Chairmanship of the South African Language Practitioners Council. [Aanlyn]. Beskikbaar: https://www.uwc.ac.za/news-and-announcements/news/uwc-professor-assumes-chairmanship-of-the-south-african-language-practitioners-council. [07/09/2023].

**Bergenholtz, H.J. en R.H. Gouws.** 2013a. The Presentation of Word Formation in General Monolingual Dictionaries. *Lexikos* 23: 59-76.

**Bergenholtz, H.J. en R.H. Gouws.** 2013b. A Lexicographical Perspective on the Classification of Multiword Combinations. *International Journal of Lexicography* 27(1): 1-24.

**Bergenholtz, H.J. en R.H. Gouws.** 2015. Proposals for the Writing of Peer Reviews in Lexicography. *Hermes. Journal of Language and Communication in Business* 54: 107-114.

**Bergenholtz, H.J. en R.H. Gouws.** 2016. On the Metalexicographic Genre of Dictionary Reviews, with Specific Reference to *LexicoNordica* and *Lexikos*. *Lexikos* 26: 60-81.

**Bergenholtz, H.J. en R.H. Gouws.** 2017. Polyseme Selection, Lemma Selection and Article Selection. *Lexikos* 27: 107-131.

**Bergenholtz, H.J., T.J.D. Bothma en R.H. Gouws.** 2015. Phases and Steps in the Access to Data in Information Tools. *Lexikos* 25: 1-30.

**Bothma, T.J.D. en R.H. Gouws.** 2020. e-Dictionaries in a Network of Information Tools in the e-Environment. *Lexikos* 30: 29-56.

**Bothma, T.J.D. en R.H. Gouws.** 2022. Information Needs and Contextualization in the Consultation Process of Dictionaries that are Linked to e-Texts. *Lexikos* 32(2): 53-81.

**Faaß, G., S.E. Bosch en R.H. Gouws.** 2014. A General Lexicographic Model for a Typological Variety of Dictionaries in African Languages. *Lexikos* 24: 94-115.

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

**Gouws, R.H.** 2006. Leerdergedrewe leksikografie. *Tydskrif vir Taalonderrig* 40(1): 72-84.

**Gouws, R.H.** 2012. Who can Really be Called a Lexicographer? *Lexikos* 22: 217-225.

**Gouws, R.H.** 2013a. Contextual and Co-textual Guidance Regarding Synonyms in General Bilingual Dictionaries. *International Journal of Lexicography* 26(3): 346-361.

**Gouws, R.H.** 2013b. Die leksikografiese aanbieding en behandeling van vaste uitdrukkings. *Lexikos* 23: 135-149.

**Gouws, R.H.** 2014a. Article structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.

**Gouws, R.H.** 2014b. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3): 481-504.

**Gouws, R.H.** 2015a. Wie is die teikengebruikers van eentalige aanleerderwoordeboeke? *Tydskrif vir Geesteswetenskappe* 55(3):343-355.

**Gouws, R.H.** 2015b. The Presentation and Treatment of Collocations as Secondary Guiding Elements in Dictionaries. *Lexikos* 25:170-190.

**Gouws, R.H.** 2016a. Op pad na 'n omvattende woordeboekkultuur in die digitale era. *Lexikos* 26: 103-123.

**Gouws, R.H.** 2016b. Enkele minder bekende Afrikaanse woordeboekmonumente. *Tydskrif vir Geesteswetenskappe* 56(2-1): 355-370.

**Gouws, R.H.** 2016c. Oor ANNA en ander andershede in die Afrikaanse leksikografie. *Tydskrif vir Nederlands en Afrikaans* 23: 58-80.

**Gouws, R.H.** 2017. Van tematiese na alfabetiese na tematiese ordening in woordeboeke — wisselwerking tussen teorie en praktyk. *Spil plus* 53: 133-148.

**Gouws, R.H.** 2018a. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.

**Gouws, R.H.** 2018b. Woordeboekdidaktiek in 'n omvattende woordeboekkultuur. *Literator* 39(2): 1-8.

**Gouws, R.H.** 2018c. Expanding the Data Distribution Structure. *Lexicographica* 34: 225-238.

**Gouws, R.H.** 2020a. Special Field and Subject Field Lexicography Contributing to Lexicography. *Lexikos* 30: 143-170.

**Gouws, R.H.** 2020b. Metalexicography, Dictionaries and Culture. *Lexicographica* 36: 3-9.

**Gouws, R.H.** 2021a. Op pad na 'n omvattende aanlyn woordeboek — met spesifieke verwysing na die *Woordeboek van die Afrikaanse Taal. Tydskrif vir Geesteswetenskappe* 61(4-1): 1156-1176.

**Gouws, R.H.** 2021b. 'n Verbeterde leksikografiese dataverspreiding- en inligtingsonttrekkingstruktuur. *Spil Plus* 61: 49-64.

**Gouws, R.H.** 2022a. Extended Article Comments in Online Dictionaries. *Lexicography* 9(2): 111-138.

**Gouws, R.H.** 2022b. Ruimtelike en temporele leksikografiese deiktiese verankering. *Lexikos* 32(3): 95-122.

**Gouws, R.H.** 2022c. Afrikaans is geen bonsai. *Netwerk24.* [Aanlyn]. Beskikbaar: https://www.netwerk24.com/netwerk24/stemme/aktueel/rufus-gouws-afrikaans-is-geen-bonsai-20220611. [16/06/2023].

**Gouws, R.H.** 2022d. Dictionaries as instruments of exclusion and inclusion: some South African dictionaries as case in point. *Lexicographica* 38: 39-61.

**Gouws, R.H., I. Feinauer en F. Ponelis (Reds.).** 1994. *Basiswoordeboek van Afrikaans.* Pretoria: Van Schaik.

**Gouws, R.H., L. Potgieter en S. Burgess (Reds.).** 2010. *Longman Grondslagfasewoordeboek Afrikaans/ English.* Kaapstad: Maskew Miller Longman.

**Gouws, R.H. en D.J. Prinsloo.** 2021a. Lexicographic Data Boxes. Part 1: Lexicographic Data Boxes as Text Constituents in Dictionaries. *Lexikos* 31: 330-373.

**Gouws, R.H. en D.J. Prinsloo.** 2021b. Lexicographic Data Boxes. Part 3: Aspects of Data Boxes in Bilingual Dictionaries and a Perspective on Current and Future Data Boxes. *Lexikos* 31: 402-433.

**Gouws, R.H., D.J. Prinsloo en M. Dlali.** 2014. A Series of Foundation Phase Dictionaries for a Multilingual Environment. *Spil* 43: 23-43.

**Gouws, R.H. en S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

**Jackson, H. (Red.).** 2022. *The Bloomsbury Handbook of Lexicography*. Londen: Bloomsbury.

**Klosa, A. en R.H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.

*LitNet.* 2022. Aankondiging: Rufus Gouws wen die Jan H Maraisprys vir 2022. [Intyds]. Beskikbaar: https://www.litnet.co.za/aankondiging-rufus-gouws-wen-die-jan-h-maraisprys-vir-2022/. [25-05-2023].

**Luther, J., F. Pheiffer en R.H. Gouws.** 2015. *Handwoordeboek van die Afrikaanse Taal*. Sesde uitgawe. Kaapstad: Pearson Education.

**McConchie, R.W.I.** 2019. *Discovery in Haste. English Medical Dictionaries and Lexicographers 1547 to 1796.* Berlyn/Boston: De Gruyter.

**Odendal, F.F. en R.H. Gouws (Reds.).** 2005. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Vyfde uitgawe. Kaapstad: Pearson Education.

**Sands, B. en K. Jones (Reds.).** 2022. *Nǀuuki Namagowab Afrikaans English ǂXoakiǂxanisi/Mîdi di ǂKhanis/ Woordeboek/Dictionary*. Stellenbosch: African Sun Media.

**Stroud, C. en C. Kerfoot.** 2021. Decolonizing Higher Education: Multilingualism, Linguistic Citizenship and Epistemic Justice. Bock, Z. en C. Stroud (Reds.). 2021. *Language and Decoloniality in Higher Education. Reclaiming Voices from the South*: 19-46. Londen: Bloomsbury Academic.

**Tarp, S.** 2013. Lexicographic Functions. 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:* 460-468. Berlyn: De Gruyter.

**Tarp, S. en R.H. Gouws.** 2019. Lexicographical Contextualisation and Personalization: A New Perspective. *Lexikos* 29: 250-268.

**Tarp, S. en R.H. Gouws.** 2020. Reference Skills or Human-centered Design: Towards a New Lexicographical Culture. *Lexikos* 30: 470-498.

**Wiegand, H.E.** 1996. A Theory of Lexicographic Texts. An Overview. *SA Journal of Linguistics* 14(4): 134-149.

**Zgusta, L.** 1971. *Manual of Lexicography.* Berlyn: De Gruyter.

# Structuring a Collection of Lexicographic Data for Different User and Usage Situations

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

**Abstract:** According to Fuertes-Olivera and Tarp (2020), lexicography is being currently shaped by three related tendencies: (a) the increasing use of disruptive technologies; (b) the necessity of finding new business models that can finance new lexicographic projects; and (c) the existence of growing competition from other information sources, e.g. Google. These trends have a particular influence on specialized dictionaries, defined here as tools that cover areas outside general cultural knowledge and its corresponding Language for General Purpose (LGP). This article adopts the view that lexicographers can deal with the abovementioned tendencies by preparing structured collections of lexicographic data with details that can be easily converted into information and stored in DWSs that allow multiple combinations and possible retrievals. This option is suitable for the tendencies, as it uses adequate technologies, e.g. ways of profiling or individualizing searches, defends a new business model based on the creation of lexicographic data that can feed many different tools, e.g. *Write Assistant* (Fuertes-Olivera and Tarp 2020), and offers better and more precise information than that of Google and other information tools.

**Keywords:** LEXICOGRAPHIC DATA, TECHNOLOGY, SPECIALIZED LANGUAGE, SPECIALIZED LEXICOGRAPHY, E-LEXICOGRAPHY

**Opsomming: Die strukturering van 'n versameling leksikografiese data vir verskillende gebruikers- en gebruiksituasies.** Volgens Fuertes-Olivera en Tarp (2020) word die leksikografie tans gevorm deur drie verwante tendense: (a) die toenemende gebruik van ontwrigtende tegnologieë; (b) die noodsaaklikheid om nuwe sakemodelle wat nuwe leksikografiese projekte kan finansier, te vind; en (c) die voorkoms van groeiende kompetisie van ander inligtingsbronne, bv. Google. Hierdie tendense het 'n bepaalde invloed op gespesialiseerde woordeboeke, wat hier gedefinieer word as hulpmiddels wat gebiede buite die algemene kulturele kennis en ooreenstemmende Taal vir Algemene Doeleindes (TAD) dek. Hierdie artikel steun die beskouing dat leksikograwe bogenoemde tendense kan hanteer deur gestruktureerde versamelings leksikografiese data met detail voor te berei wat maklik in inligting omskep kan word en in woordeboekskryfstelsels (WSSs) geberg kan word wat veelvoudige kombinasies en moontlike afvoere moontlik maak. Hierdie opsie is toepaslik vir die tendense, aangesien dit voldoende tegnologieë gebruik, bv. metodes vir profielsamestelling of geïndividualiseerde soektogte, steun bied aan 'n nuwe sakemodel wat gebaseer is op die skep van leksikografiese data wat baie verskillende hulpmiddels kan voed, bv. *Write Assistant* (Fuertes-Olivera en Tarp 2020), en beter en meer presiese inligting as dié van Google en ander inligtingshulpmiddels kan aanbied.

**Sleutelwoorde:** LEKSIKOGRAFIESE DATA, TEGNOLOGIE, GESPESIALISEERDE TAAL, GESPESIALISEERDE LEKSIKOGRAFIE, E-LEKSIKOGRAFIE

## 1.    Introduction

A review of recent publications on lexicography and terminology reveals that we are in the middle of a *Cambrian explosion* (Fuertes-Olivera 2016), a concept typically used for referring to situations that are similar to what occurred around 540 million years ago, when new life forms began to multiply and everything changed in a way that made life completely different (Siegele 2014). For instance, Fuertes-Olivera and Tarp (2020) mention three recent tendencies that are shaping lexicographers' work and point to the future. The first tendency is the increasing use of disruptive technologies in almost all aspects of the discipline. The second is the growing obsolescence of the business model which, by and large, has financed the compilation of reference works for the last five hundred years. The third tendency is growing competition from other information sources, especially Google. This article addresses how these trends can affect the creation of online lexical resources for specialized languages, which is the topic of this study. By doing so, I present my solution and see it in line with Gouws's main academic endeavour, which has focused on adapting theory to practice and practice to theory (Gouws 2011), and being always "open" to new ideas and methods (Domínguez Vázquez and Gouws 2023).

Firstly, we will review the concept of specialized language, and especially that of specialized online dictionaries (section 2). Next, we will focus on the concept of *lexicographic data*, which we envisage as central for the future of lexicography (section 3). In section 4, we will illustrate a proposal for structuring lexicographic data in novel ways with the aim of offering alternatives that might lead to a rebirth of lexicography. For reasons of space, we will concentrate on different consultation possibilities, especially those that can be connected with the creation of specialized information tools from the same Dictionary Writing System (DWS). A final conclusion will summarise the main points discussed and will reflect on future developments in online lexicography.

## 2.    Specialized languages

Cabré (1998: 118-121) indicates that specialized languages are sub-assemblies of the common language with their own rules and specific units. They are not easily understood by non-specialists, have to be consciously learned, and undergo change, variation and other processes that are also found in the common language, i.e. the unmarked linguistic code all humans learn and use, for instance, in face-to-face conversations and similar daily encounters. The above broad definition has typically led researchers to focus on the rules and units that are restricted, i.e. not used in the common language, and have created ref-

erence works that only (or mostly) deal with these rules and units.

Regarding the different specialized rules and units described so far, we will focus on lexical units, i.e. single words or chains of words that form the basic elements of the vocabulary of specialized languages. These represent the backbone of specialized dictionaries, i.e. reference tools that cover areas outside general cultural knowledge and its corresponding Language for General Purpose (LGP). Hence, they include disciplines such as technology, natural and social sciences, health sciences and humanities, to name just a few. In addition, they should not be considered a very specific type of dictionary with rigid and well-established characteristics, but rather a series of lexicographic works with different characteristics, sizes and terms, such as "glossary", "terminological database", "lexicon", "technical dictionary", "encyclopaedia", "vocabulary", "domain ontology", "knowledge bank", and so on (Fuertes-Olivera and Tarp 2014: 7-8). Under this general definition, specialized dictionaries describe the various specialized languages and substances of these disciplines to provide direct, specific access to their cognitive elements.

Such works may be either in paper format or online, but some are only paper whereas others are only digital. Regarding specialized online dictionaries, Fuertes-Olivera (2016: 227-228) has commented on two findings that are shaping the discipline. The first is that only a limited number of these are really different from their printed counterparts, that is, in terms of the use they make of internet technologies, e.g. for favouring individualisation (i.e. customisation or profiling); this is a defining factor when differences between printed and online dictionaries are analysed. The *Accounting Dictionaries* (Fuertes-Olivera and Niño Amo 2018) to a certain extent exploit available technologies from information science like user profiling, filtering and adaptive hypermedia, and also frequently link to the internet, where already existing data are reused in order to satisfy users' specific needs (Bothma 2011).

The second finding is that the design and construction of specialized online dictionaries should combine the use of lexicographic theories and methods with the active participation of IT experts as well as experts in the field; these should be experts, for instance, in medicine if the dictionary describes medical concepts and language. Such cooperation is of paramount importance, and will lead to the design of a high-quality online dictionary, which, according to Fuertes-Olivera (2016: 229-230)

1.   is an information tool that contains (or can contain) many types of lexicographic data — e.g. dictionary articles, systematic introductions, hyperlinks, and so on (see section 3, below);
2.   offers the data in a structured way (see section 4, below);
3.   is always for specific consultation with the aim of converting its structured data into information quickly and easily;
4.   contains specialized relevant data categories for the user: terms, definitions, example sentences, instructions, non-verbal signs, and so on;

5.  offers reliable data, i.e. the validity and suitability of the data offered have been checked;
6.  makes use of internet technologies, especially those favouring individualisation;
7.  consists of three related components: a lexicographic database, a search system and a graphical user interface (GUI);
8.  is updated continuously.

As the above characteristics show, online dictionaries cover both the language and facts of one or more domains. Regarding facts, i.e. knowledge of a particular domain, Hashimzade et al. (2014: 11-16) explicitly indicate that only experts with proper knowledge can write definitions of economic concepts in dictionaries directed at experts and semi-experts, who are the typical users of specialized dictionaries. As for the language of a particular domain, gone are the days when specialized dictionaries had very few data types, typically lemma and/or equivalent, definition, and sub-domain, e.g. example 1 from the *Diccionario McGraw-Hill de Química bilingüe español–inglés/English–Spanish*:

**mezcla azeotrópica** *azeotropic mixture* [QUIM]

Disolución de dos o más líquidos, cuya composición
   no cabía por destilación. También conocida por
   azeótropo.

**Example 1:**   Dictionary entry in *Diccionario McGraw-Hill de Química*

Instead, modern specialized (online) dictionaries aim at offering a very precise description of specialized lemmas. Fuertes-Olivera and Tarp (2014: 199-200), for instance, indicate that the DWS for the *Accounting Dictionaries* contains up to 21 lexicographic data categories for each accounting lemma (Table 1, below) so as to assist their target users: (1) translators and language staff; (2) accounting experts and semi-experts; and (3) students and laypersons interested in Danish, English and Spanish matters related to accounting.

**Table 1:**   Data types and categories in the accounting lexicographic database

| Data Type | Rationale |
|---|---|
| Lemma | Self-evident: all dictionaries describe lemmas. |
| Homonymy index (superscript in the dictionary) | Included when homographs have different inflectional paradigms or are countable or uncountable, respectively. It assists all user types in all situations: it restricts the meaning of lemmas. |
| Polysemy index (Arab numbers in the dictionary | Included when homographs have different definitions. It assists all user types in all situations. It restricts the meanings of the same lemma. |

| | |
|---|---|
| Language code to lemma | Included for indicating language and language variant of the lemma. It assists all user types in all situations. |
| Grammatical data addressed to lemma | Offers inflections, countability, active and passive forms of verbs. It assists all user types in two communicative situations: production and translation. It offers linguistic profiles of lemmas. |
| Equivalent | One Danish, English or Spanish equivalent in several dictionary combinations: Danish–English (Danish Lemma; Danish definition; English equivalent); English–Danish (English lemma; English definition; Danish equivalent); English–Spanish (English lemma; English definition; Spanish definition; Spanish equivalent); Spanish–English (Spanish lemma; Spanish definition; Spanish equivalent). It assists all user types in communicative situations. |
| Language code to equivalent | Included for indicating language and language variant of the equivalent. It assists all user types in communicative situations. |
| Grammatical data addressed to equivalent | Offers inflections, countability, active and passive forms of verbs. It assists all user types in two communicative situations: production and translation. It offers linguistic profiles of equivalents. |
| Definition of lemma | Explains the meaning(s) of lemmas. Always one definition per sense. Definitions are crafted in the language of the lemma. There is one exception: in the English–Spanish dictionary, definitions are in English and Spanish. It assists all user types in all situations. |
| Collocations | Short and long phrases but not full sentences. They are crafted in the language of the lemma. They *mainly* assist users in two communicative situations: production and translation. In addition, the Spanish dictionary also includes collocations that assist users in cognitive situations (this is a useful method for dealing with culture-bound subject fields). |
| Translation of collocations | In some dictionary combinations, collocations are translated: English collocations are translated into Danish and Spanish, and Danish and Spanish collocations are translated into English. They assist all user types in two communicative situations: production and translation. |
| Language code to translation of collocations | Included for indicating language and language variant of the lemma and equivalent. It assists all user types in two communicative situations: production and translation. |
| Examples | Full sentences showing lemmas in use. They assist all user types in cognitive situations and two communicative situations: production and translation. |
| Translation of examples | They are full sentences showing equivalents in use. They are in some dictionary combinations: English examples are translated into Danish and Spanish, and Danish and Spanish examples are translated into English. They assist all user types in cognitive situations and two communicative situations: production and translation. |

| Synonyms and antonyms addressed to lemmas or equivalents | Assist all user types in two communicative situations: production and translation. |
|---|---|
| Language code to synonyms and antonyms | Included for indicating language and language variant of the synonym and antonym. It assists all user types in two communicative situations: production and translation. |
| Source | Hyperlinks to external texts. It mainly assists experts in cognitive situations. It can also assist translators and semi-experts in two communicative situations: production and translation. |
| Lexicographic notes | Usage, and/or contrastive notes that are addressed to lemma or equivalent, especially for indicating factual differences between lemma and/or equivalent, particular usages, cultural details, etc. They mainly assist all user types in cognitive situations. |
| Grammar notes attached to lemma and/or equivalent | Indicate language profiles of some lemmas and/or equivalents. They mainly assist users in two communicative situations: production and translation. |
| Proscriptive notes | Used for recommending lemmas. They mainly assist all user types in communicative situations. |
| Cross-references | Hyperlinks to internal texts. They assist all user types in cognitive situations and two communicative situations: translation and production. |

Source: Adapted from Fuertes-Olivera and Tarp 2014: (199-200)

Table 1 also speaks of a new trend in online lexicography, namely, that the time is ripe for creating lexicographic data (section 3) that pay scant attention, if any, to dictionary typologies. Bowker (2018) seems to concur with this idea by explaining the process of convergence of lexicography and terminology (i.e. specialized lexicography in this paper), which she views as inevitable and illustrates in Table 2:

**Table 2:**   Evolution and convergence of characteristics associated with lexicography and terminology

|  | **Lexicography** | **Terminology** |
|---|---|---|
| **Practitioner** | mainly lexicographers, but with greater involvement from the general public (via crowdsourcing) | mainly terminologists, but with greater involvement from the general public and subject matter experts (via open and closed crowdsourcing) |
| **Object of study** | mainly words, but also some terms | mainly terms, but also some general language words or expressions |
| **Domain** | mainly general language, but also some specialised language | mainly the language of a specialised domain, but also some general language |

| **Point of view** | mainly descriptive | mainly normative/prescriptive in the public and academic sectors, but incorporating more descriptive elements in commercial settings |
| --- | --- | --- |
| **Approach** | mainly semasiological | increasingly semasiological, but retaining some onomasiological elements where useful |
| **Organisation** | mainly alphabetical, but sometimes incorporating thematic elements | mainly thematic, but allowing alphabetic searching |
| **Main information provided** | words, meanings, examples, usage information (e.g. collocations, frequency, phraseology), a range of linguistic information (e.g. part of speech, pronunciation) | preferred term, variants, context and usage information (e.g. collocations, frequency, phraseology), meaning, conceptual relations |
| **Intended users** | lay people, professional and academic audiences | public sector (for language planning), domain experts, scientific/technical writers, translators (for bi- or multilingual resources), commercial enterprises |

Source: Bowker (2018: 148)

### 3.      Lexicographic data in the era of the internet

We define lexicographic data as any data that have been prepared or accepted by lexicographers and stored in a DWS with the aim of helping humans and/or machines convert them into information *in a straightforward manner*. This definition merits several comments. Firstly, any data in a DWS implies that the system must allow the inclusion of such data in any format, especially:

— Words such as **azeotrope**
— Running text, e.g. in definitions and systematic introductions
— Symbols, e.g. that of the dollar: **$**
— Sounds, e.g. that of *lion* (Figure 1):
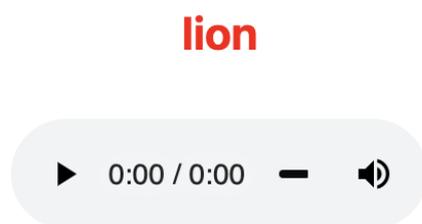


**Figure 1:**   Pronunciation of *lion* in *howjsay*

— Films, e.g. how to make pancakes in *YouTube*
— Pictures, e.g. that of a lion (Figure 2)



**Figure 2:** Picture of a *lion*

Secondly, the data should be prepared or accepted by lexicographers. This comprises selection, analysis and acceptance of the data to be included in the DWS. Fuertes-Olivera and Tarp (2014), among others, have shown that the process of selection based on restricted corpus data is not recommended in specialized lexicography. Instead, we believe that for selecting, analysing and accepting data we can rely on the internet as a corpus, and make use of big data analytics for extracting, say, the initial lemma list of the dictionary (Fuertes-Olivera et al. 2018), and of Google minitexts for preparing definitions, examples, grammar, etc. of each lemma (Fuertes-Olivera 2012; Fuertes-Olivera et al. 2018; Tarp and Fuertes-Olivera 2016). For instance, IT staff at Ordbogen.com, a Danish language technology company, tracked around one million daily searches in English and Spanish. They found that approximately 80% of these can be matched, i.e. the same search is identified in the logfiles of different dictionaries

and may, therefore, be interpreted in order to identify the most popular articles in the dictionaries under scrutiny. After two months of work with the logfiles of the searches, which amount to more than 60 million, IT staff at Ordbogen A/S were able to produce two lists of 20,000 English words and 16,000 Spanish words. They comprise the words most commonly searched in the period under analysis and were used by the editor of the *Diccionarios Valladolid-UVa* for compiling the initial lemma lists of this project (Fuertes-Olivera et al. 2018). Something similar can be done for selecting the initial lemma list of specialized dictionaries.

Regarding Google minitexts, they offer at a glance data that can be easily understood and used, especially if the retrieved data are analysed by an expert in the field. For instance, googling *azeotrope* retrieves around 786,000 hits. The first six hits (Figure 3) offer enough data for one to learn

(a)   that *azeotrope* is a countable noun used in chemistry;
(b)   its synonym is *constant boiling point mixture*;
(c)   *azeótropo* and *mezcla aceotrópica* are its Spanish equivalent, one can be included as equivalent and the other as synonym of the equivalent;
(d)   and that it is "a mixture of two or more liquids whose proportions cannot be altered or changed by simple distillation" because it "exhibits the same concentration in the vapor phase and the liquid phase".

https://en.wikipedia.org › wiki › Azeotrope       ⋮
### Azeotrope - Wikipedia
An **azeotrope** or a constant heating point mixture is a mixture of two or more liquids whose proportions cannot be altered or changed by simple distillation.
Raoult's law · Heteroazeotrope

https://en.wikipedia.org › wiki › Azeotrope_tables       ⋮
### Azeotrope tables - Wikipedia
This page contains tables of **azeotrope** data for various binary and ternary mixtures of solvents. The data include the composition of a mixture by weight (in ...

https://chem.libretexts.org › ... › Non-ideal Solutions       ⋮
### Azeotropes - Chemistry LibreTexts
Sep 23, 2020 — An **azeotrope** is a mixture that exhibits the same concentration in the vapor phase and the liquid phase. This is in contrast to ideal solutions ...

https://www.sciencedirect.com › topics › chemistry › azeo...       ⋮
### Azeotropic Mixture - an overview | ScienceDirect Topics
An **azeotrope** is a mixture of two or more liquid components under constant boiling, and distillation processes are performed as if they were a pure compound (see ...

**Figure 3:**   Initial hits when googling **azeotrope**

Thirdly, humans and machines, e.g. *Write Assistant* (Fuertes-Olivera and Tarp 2020), should convert lexicographic data into information in a single cognitive process or click. This is a really crucial point in our definition of lexicographic data. In these circumstances, most data in, say, existing Spanish dictionaries are not lexicographic, as they cannot be understood due to several flaws in their treatment and presentation, especially in terms of the use of abbreviations, recursive definitions and lack of adequate information. For reasons of space, we will illustrate this defining condition with two examples of current practices that should be abandoned:

— Recursive definitions should be discarded because they are useless. For instance, Spanish dictionaries define "action nouns", i.e. deverbal nouns that refer to an action or an event with the formula "acción y efecto de" (action and effect of) plus the verb they refer to (example 2):

**dilución**

acción y efecto de diluir[1]

**Example 2:**   Definition of *dilución* in Spanish dictionaries, e.g. in DLE

Such a definition says nothing and, therefore, is not an example of lexicographic data. As indicated previously, lexicographers can google *dilución* (English: dilution), study the hits found and act accordingly. For instance, we have found that *dilución* has five meanings (example 3): one of them refers to the action of making a solid more dilute by dropping it into a liquid; another meaning refers to the solid that has been diluted (definitions 1 and 2 of example 3). Both are used in general language. In addition, it has a figurative meaning, also used in general language, which refers to the process

of weakening some abstract process, e.g. the action of making something weaker in form, content, value and so on (definition 4 of example 3), and two further meanings; a literal one used in chemistry, referring "to the process of decreasing the concentration of a solute in a solution, usually simply by mixing with more solvent like adding water to the solution" (Wikipedia: Dilution (equation) (definition 3 of example 3), and a figurative one found in economics, referring to reducing the value of a shareholding due to offering more shares without increasing assets (definition 5 of example 3). These five meanings, as well as many more data about the same, are all stored in the DWS of the *Diccionarios Valladolid-UVa* (Fuertes-Olivera 2019). These are examples of lexicographic data, as all of them illustrate the meanings of the word in the different contexts and domains found in real usage, and help potential users (humans and/or machine) to immediately disambiguate meanings and usages:

**dilución**

1. en sentido literal, disolución de un cuerpo sólido en un líquido (literally, making a solid more dilute by dropping it into a liquid)
2. en sentido literal, sustancia que resulta del proceso de disolución de algo por medio de un líquido (literally, solid that comes out of the process of dilution)
3. en sentido literal, reducción de concentración de una sustancia química en una disolución; es un procedimiento para preparar una disolución menos concentrada a partir de otra más concentrada; se usa en química (literally, process of decreasing the concentration of a solute in a solution; it is used in chemistry)
4. en sentido figurado, proceso realizado de forma consciente con el que se intenta un debilitamiento de algo (figuratively, conscious process used for weakening something)
5. en sentido figurado, disminución del valor teórico de las acciones de una empresa debido a la emisión de nuevas acciones sin prima de emisión, es decir, a la par o a un valor inferior al valor de mercado, o a la conversión de bonos u obligaciones en acciones; se usa en economía (figuratively, reducing the value of a shareholding due to offering more shares without increasing assets; it is used in economics)

**Example 3:**    Definitions of *dilución* in the DWS of the *Diccionarios Valladolid-UVa*

—  Dictionaries should always contextualize each meaning and usage. For example, Spanish dictionaries do not typically differentiate between literal and figurative meanings. Sometimes they offer a meaning and indicate that the lemma in question has figurative meanings but do not explain them. This is also of no use, so it should be discontinued. The difference between "literal" and "figurative" really matters. The vocabulary stock of specialized languages tends to be made up of three main processes of word formation: (a) derivation, e.g. "zero derivation" in English, such as *audit* (a noun,

a verb and an adjective; see examples (5, 6 and 7, below); (b) compounding, especially the creation of extended units of meanings and prefabricated chunks of words, such as *account day*, *define contribution scheme*, *defined contribution pension plan*, and so on; (c) and figurative extensions of general meanings, e.g. definition 5 of *dilución* in example 3. As shown in example 3, above, we believe it necessary to differentiate between literal and figurative meanings when defining words that have both types of meanings. This has led us to write in the DWS of the *Diccionarios Valladolid-UVa* the expression "en sentido figurado" (English: figuratively) at the start of the definition and "se usa en economía" (English: it is used in economics) at the end of the definition, as shown in example 3. These expressions are relevant, as will be shown in the next section.

## 4.     Structured lexicographic data collections

Our definition of lexicographic data merits two main comments. Firstly, dictionaries must always use simple and easy-to-understand language and signs. This means that all language and signs that need more than one step to be understood should be eliminated. Secondly, the lexicographic treatment of lemmas should be as complete as possible. Both ideas will lead to the use of DWSs that should be both complex and dynamic, i.e. they can contain as much data as possible but will permit punctual consultation of the data really needed in a particular usage situation and by a specific user type.

We believe that this can be achieved by creating structured lexicographic data collections, i.e. repositories of data that have four defining characteristics. Firstly, they are created by human analysts, who must study the sociolinguistic context of words and offer detailed descriptions of all their relevant characteristics. These analysts can make use of all the (computer) tools and traditions they consider appropriate for offering deep and logical analyses of the words studied. For instance, the English noun *black swan* originated in day-to-day conversation for referring to *Cygnus stratus*, a term used in zoology, which describes a "species of swan which breeds mainly in the southeast and southwest regions of Australia" (Wikipedia: Black swan). The noun describes the physical and salient characteristics of a swan that was unknown until Europeans landed in Australia in the 17th century. In 2007, Nassim Taleb elaborated on a "black swan theory" in the domains of economics and political science. He referred metaphorically to a *black swan* as an unpredictable or unforeseen event, e.g. the September 11 attacks. This meaning has also re-surfaced again in general language to refer to something extremely rare, which can have positive or negative consequences. Hence, googling *cisne negro* (English: *black swan*) will inform human analysts of the animal *Cygnus stratus,* and two related metaphorical meanings, i.e. an extremely rare event, which is used in general language and may have positive or negative outcomes, and the specialized meaning in economics and political science, where it refers to an event that comes as a sur-

prise, has a major effect, and "is often inappropriately rationalized after the fact with the benefit of hindsight" (Wikipedia: Black swan theory). It is interesting to highlight that Spanish dictionaries ignore the metaphorical meanings of *cisne negro*, although this word is found in daily use in Spanish newspapers, magazines and TV programs. For instance, a search of "cisne negro" in Google retrieves 1,660,000 hits, most of which refer to above-mentioned metaphorical meanings and a film released by Hollywood in 2010.

Secondly, a description of the words must be placed in slots, i.e. allotted places in dictionary writing systems, each of which must be reserved for a particular data type, e.g. one slot for indicating that the noun is countable or uncountable, one for the different forms of the conjugation of the verb, one for its meaning, and so on. We believe that we need around 30 slots for each language: (1) lemma; (2) word class; (3) index number for homonymy; (4) index number for polysemy; (5) inflections; (6) grammar; (7) proscription notes, i.e. notes informing on sociolinguistic aspects of the word, e.g. recommended spellings; (8) meaning; (9) synonyms; (10) antonyms; (11) related words; (12) grammar notes; (13) usage notes; (14) diaphasic variant; (15) diastratic variant; (16) diatopic variant; (17) diachronic variant; notes on (18) synonyms, (19) antonyms and/or (20) related words; (21) notes on any variants; (22) sentence examples; (23) chunks of texts showing specific usages of the word, e.g. collocations; (24) figures; links to external sources, e.g. (25) to corpus data, (26) encyclopedic articles, e.g. Wikipedia; (27) symbols, equations, and so on; also, there should be three empty slots for possible use during the process of compilation.

To the best of our knowledge, existing specialized dictionaries tend not to include most of the above data types. For instance, we have investigated the accounting dictionaries retrieved when googling "accounting dictionary" and found a collection of reference works, all of which only offer the definition(s) of words, such as *audit* (example 4):

AUDIT is the inspection of the accounting records and procedures of a business, government unit, or other reporting entity by a trained accountant for the purpose of verifying the accuracy and completeness of the records. It could be conducted by a member of the organization (internal audit) or by an outsider (independent audit). A CPA audit determines the overall validity of financial statements. A tax audit (IRS in the U.S.) determines whether the appropriate tax was paid. An internal audit generally determines whether the company's procedures are followed and whether embezzlement or other illegal activity occurred.

**Example 4:**     Entry for *audit* in *Ventureline*

In our view, structured lexicographic data collections should contain better descriptions, such as, for example, our description of *audit* in the DWS of the *Diccionarios Valladolid-UVa* (examples 5, 6 and 7):

*audit*
noun: an audit, the audit, audits

definitions:

1.  An audit is an official, methodological examination or review by an expert

Synonym: examination

Collocations:
*   annual audits
*   audit and quality assurance

2.  An audit is an examination of the financial report and statements of an enterprise by an independent auditor after which the auditor gives an opinion which is expressed in an audit report.

Synonym: auditing

Collocation:
*   Plan an audit
*   Perform an audit
*   The scope of the statutory audit
*   Etc.

Example:
*   An audit also includes assessing the accounting policies used and significant estimates made by the Board of Directors, as well as evaluating the overall Annual Report presentation.
*   We conducted our audit in accordance with International Standards on Auditing.

**Example 5:**    Dictionary entry for *audit* as a noun in the DWS of the *Diccionarios Valladolid-UVa*

*audit*
verb: audits, audited, has audited, auditing, is audited, are audited, was audited, were audited, been audited, being audited

Definitions

1.  To audit is to independently examine and subsequently express an opinion on the financial statements of an organisation.

Collocations:
*   have the annual report audited
*   audit the accounts independently

Example:
- We have audited the accompanying balance sheet
of the ABC Company as of 31 December 2016.

2.  To audit is to check something officially.

Collocations:
- audit the accounts

**Example 6:**    Dictionary entry for *audit* as a verb in the DWS of the *Diccionarios Valladolid-UVa*

*audit*
adjective

definition:
The placement of 'audit' in front of a noun specifies the meaning of the noun to pertain to the field of auditing: audit services, for example, are services performed with a view to auditing financial statements.

Collocation:
- special audit consideration
- an audit requirement
- draw conclusions from audit observations

**Example 7:**    Dictionary entry for *audit* as an adjective in the DWS of the *Diccionarios Valladolid-UVa*

Thirdly, there are three types of slots: "open slot"; "restricted slot" and "expanding slot". The open slot, e.g. the definition slot, is prepared for storing running text without limited space. The restricted slot is reserved for roll-down menus that display sets of previously defined categories of data, e.g. "informal Spanish". For instance, in a Spanish dictionary, there may be 23 geographical labels for referring to an "americanismo" (English: Americanism), a geographical variant typically used for referring to a meaning that is used outside Spain, e.g. in Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippines, Puerto Rico, United States, Uruguay, Venezuela and West Sahara. This second type of slot allows the lexicographer to include "and data", e.g. a word can be used in Argentina, Bolivia, and Mexico, and "either or data", i.e. if the word is "informal" it is not formal. Finally, the "expanding slot" is reserved for data types that have different forms, e.g. Spanish verbs can have up to 56 different forms of the same verb, which are stored in sub-slots, each reserved for a specific form. In other words, the slot for verb conjugations can have as many as 56 sub-slots, although not all of them are needed for all verbs. It is the lexicographer's task to analyse each

verb and decide which of the different forms are needed in the DWS. In the *Diccionarios Valladolid-UVa,* we have limited the number of sub-slots to eight, as these will allow the retrieval of the verb, no matter which form is used in the search engine. This is important because in Spanish dictionaries search strings such as "comíamos" (English: we ate) retrieves nothing (Figure 4):
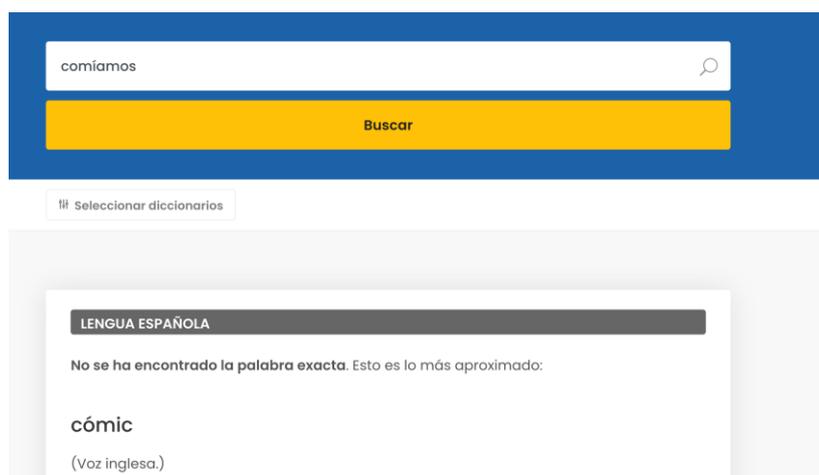


**Figure 4:**    Searching "comíamos" in *Diccionarios.com*

The three types of slots should allow different types of connections among them, especially for enabling different types of retrieval. Table 3 shows the technical document used for preparing a grammar and spelling dictionary with the data types stored in the DWS of the *Diccionarios Valladolid-UVa*:

**Table 3:**    Technical document for preparing a grammar and spelling diction-
ary of Spanish

| search in the fields + search sequence | Field | sequence in the dictionary+text | shown as a list if more than 10 results | explication text |
|---|---|---|---|---|
| 1. lemma field (including inflection forms) 2. *lemma* 3. Fuzzy search | 1.    Lemma | 1 | | |

| | | | | |
|---|---|---|---|---|
| | 2.   Style marker to lemma | | | |
| | 3.   Homonym number | 2 | | |
| | 4.   Polyseme number | 8 (but only if polysemy by a homonym OR if a grammar remark by a polyseme) | | |
| | 5.   Meaning | 9 (but only if homonymy OR if a grammar remark by a polyseme) | | |
| | 6.   Lexical remark | | | |
| | 7.   Lexical remark for text production | | | |
| | 8.   Word class and expression class | 3 | | |
| | 9.   Grammar, inflexion class 1 | 4 | | |
| | 10. Grammar, inflexion class 99 | 5 | | too, but not recommended <***> |
| | 11. Valency | 6 | | |
| | 12. Grammar/spelling remark | 7 | | |
| | 13. First reference | 10 | | This variant is not recommended, use instead *** |
| | 14. Second reference(s) | | | |
| | 15. Collocation(s) | | | |
| | 16. Example(s) | | | |
| | 17. Word formation(s) | | | |
| | 18. Synonym(s) | | | |
| | 19. Style marker to synonym | | | |
| | 20. Antonym(s) | | | |
| | 21. Style marker to antonym | | | |
| | 22. Synonym remark | | | |
| | 23.  Internet link | | | |
| | 24. Dictionary grammar | | | FIELD NOT USED |
| | 25. Memo field | | | |

Source: Henning Bergenholtz (personal communication) and IT staff at Ordbogen A/S.

Table 3 is in line with Gouws's concept of "mother dictionary" (2014). It offers flexibility and allows lexicographers to create a central data collection, which

can be displayed differently with the aim of catering to different user needs in different usage situations. Table 3, then, tells IT experts how they should connect the different slots (and sub-slots) in order to retrieve only the specific data type needed in the usage situation envisaged, which is a user searching for grammar and/or spelling information. The label "search in the field + search sequence" informs them that the search string should be prepared for three types of searches: the lemma as it is, including its inflected forms, part of the lemma (*lemma*), and fuzzy search. This will help users retrieve all possible search strings. The label "field" enumerates the possible slots needed for this type of specialized dictionary. For example, it will need a slot for storing a note on grammar and/or spelling (number 12 in Table 3). The labels "sequence in the dictionary + text" and "explication text" indicate which lexicographic data must be retrieved, i.e. which slots are to be retrieved, in which order and possible wording of any rule. For example, searching *a ver* in a grammar or spelling dictionary of Spanish, will retrieve the following (example 8):

- Lemma: *a ver*
- Expression (i.e. the word class). It is used as an adverb.
- Spelling variant: *veamos*
- Meaning:
  1. expresión que se usa para indicar que se siente expectación, ganas o curiosidad por que ocurra determinada situación que se menciona justo después de la expresión, o por observar de qué forma sucede (it works as a formula for showing interest in someone or something)
  2. expresión que se emplea para pedir a alguien que enseñe o muestre algo que se desea observar o conocer (it asks someone to show something he or she has)
  3. expresión que se emplea para invitar a otras personas a adoptar una actitud de espera con el fin de conocer o comprobar qué sucede con determinada situación sobre la que hay incertidumbre (it is used for inviting someone to adopt a particular stance in an unknown situation)

**Example 8:**    Lexicographic data retrieved when searching *a ver* in the DWS of a grammar and or spelling dictionary of Spanish

Finally, some of the slots must be equipped with "add-on" buttons, whose function should also be explained in technical documents such as that of Table 3. Both methods will allow the lexicographic data stored in a DWS to be accessed in multiple ways. For instance, the add-on button in the homonym index slot of *audit* (examples 5, 6 and 7 above) indicates that we have three different words, each of which is linked separately to the meaning slot, which also includes an add-on button for adding as many meanings as necessary. Each add-on button of the meaning slot controls the synonyms, collocations and examples of each meaning. This leads to the joint retrieval of each meaning with its own lexico-

graphic data, as shown in examples (5), (6) and (7).

## 5.    Conclusion

According to Fuertes-Oliver and Tarp (2020), lexicography is currently being shaped by three related tendencies: (a) the increasing use of disruptive technologies; (b) the necessity of finding new business models that can finance new lexicographic projects; and (c) the existence of growing competition from other information sources, e.g. Google. These three trends are especially influencing specialized dictionaries, here defined as tools that cover areas outside general cultural knowledge and its corresponding Language for General Purpose (LGP).

This paper has illustrated how the creation of online lexical resources for specialized languages can deal with the above-mentioned tendencies, by making a very precise definition of lexicographic data which must be stored in structured collections. Lexicographic data is any data in any format that has been prepared and/or revised by lexicographers and can be converted into information rapidly and straightaway, e.g. without further look-ups.

Structured lexicographic data collections are repositories which are created by humans who have studied the sociolinguistic contexts of each word, offering detailed descriptions of meanings and usage and placing them in different types of slots, each of which can be linked in different ways so as to allow potential users to retrieve the data they need in each particular usage situation.

If properly implemented, these ideas indicate that the future of online lexical resources for specialized languages depends on implementing data collections that can be restricted during the process of retrieval but not during the process of compilation. In short, lexicographers should store in the DWS as much data as possible and work with IT experts to create DWSs that allow multiple combinations and multiple retrievals. This option can handle the three tendencies, as it uses novel technologies, e.g. ways of profiling or individualizing searches, favours a new business model based on the creation of lexicographic data that can feed many different tools, and offers better and more precise information than that provided by Google and other information tools.

## Acknowledgments

## References

**Bothma, Theo J.D.** 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, Pedro A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.

**Bowker, L.** 2018. Lexicography and Terminology. Pedro A. Fuertes-Olivera (Ed.). 2018: 138-151.

**Cabré, M.T.** 1998. *La terminologie: théorie, méthode et application. Traduit du Catalan, adapté et mis à jour par Monique C. Cormier et John Humbley.* Ottawa: Les Presses de l'Université d'Ottawa.

**Diccionario McGraw-Hill de Química.** Parker, S.P., J. Weil, B. Richman, E.J. Fox, J. Faulk and F. Jr. Kotowski (Eds.). 1991. *Diccionario McGraw-Hill de Química bilingüe español–inglés/English–Spanish.* Mexico: McGraw-Hill.

**Diccionarios.com.** Diccionarios con la garantía Larousse and Vox. https://www.diccionarios.com/. Last access: July 28, 2022.

**DLE** = *Diccionario de la Lengua Española*. Real Academia Española. https://dle.rae.es/. Last access: July 28, 2022.

**Domínguez Vázquez, María José and Rufus H. Gouws.** 2023. The Definition, Presentation and Automatic Generation of Contextual Data in Lexicography. *International Journal of Lexicography* 36(3): 233-259.

**Fuertes-Olivera, Pedro A.** 2012. Lexicography and the Internet as a (Re-)source. *Lexicographica* 28: 49-70.

**Fuertes-Olivera, Pedro A.** 2016. A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. *International Journal of Lexicography* 29(2): 226-247.

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

**Fuertes-Olivera, Pedro A.** 2019. Designing and Making Commercially Driven Integrated Dictionary Portals: The *Diccionarios Valladolid-UVa*. *Lexicography* 6(1): 21-41.

**Fuertes-Olivera, Pedro A. and M. Niño Amo.** 2018. The Accounting Dictionaries. Pedro A. Fuertes-Olivera (Ed.). 2018: 455-472.

**Fuertes-Olivera, Pedro A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography.* Berlin/Boston: De Gruyter.

**Fuertes-Olivera, Pedro A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.

**Fuertes-Olivera, Pedro A., S. Tarp and P. Sepstrup.** 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Diccionarios Valladolid-UVa. *Lexikos* 28: 152-176.

**Gouws, Rufus H.** 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography:* 17-29. London/New York: Continuum.

**Gouws, Rufus H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.

**Hashimzade, N., G.A. Myles and G.D. Myles.** 2014. Can Authority Be Sustained while Balancing Accessibility and Formality. *Hermes. Journal of Language and Communication in Business 27*(52): 11-24.

**Howjsay.** https://howjsay.com/. Last access: July 28, 2022.

**Siegele, L.** 2014. A Cambrian Moment. *The Economist*, January 18, 2014: 1-14 (Special Report).

**Tarp, S. and Pedro A. Fuertes-Olivera.** 2016. Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-UVa. *Lexikos* 26: 273-295.

**Ventureline** = *Ventureline. Accounting Terms.* https://www.ventureline.com/accounting-glossary/. Last access: July 28, 2022.

**Wikipedia.** Black swan theory: https://en.wikipedia.org/wiki/Black_swan_theory. Last access: July 28, 2022.

**Wikipedia.** Dilution (equation): https://en.wikipedia.org/wiki/Dilution_(equation). Last access: July 28, 2022.

# Ways out of the Dictionary: Hyperlinks to Other Sources in German and African Online Dictionaries

Annette Klosa-Kückelhaus, *Leibniz Institute for the German Language, Mannheim, Germany (klosa@ids-mannheim.de)*

**Abstract:** This study examines a number of German and African online dictionaries to see how they make use of the possibility of linking to external sources (e.g. other dictionaries, encyclopaedias, or even corpus data). The article investigates which hyperlinks occur at which places in the word articles and how these are presented to the dictionary users. This is done against the background of metalexicographic considerations on the planning of outer features and the mediostructure in online dictionaries as well as different categorizations of hyperlinks in online reference works. The results show that retro-digitized dictionaries make virtually no use of hyperlinks to external sources. Genuine online dictionaries, on the other hand, do, but often in a form that needs improvement, since, for example, explanations of dictionary-external links are not always found in the user guide and their design is different even within a dictionary.

**Keywords:** ONLINE DICTIONARIES, RETRO-DIGITIZED DICTIONARIES, MEDIOSTRUCTURE, HYPERLINKS, AFRICAN LANGUAGES, GERMAN

**Opsomming: Roetes uit die woordeboek uit: Hiperskakels na ander bronne in aanlyn Duitse en Afrika-woordeboeke.** In hierdie artikel word 'n aantal aanlyn Duitse en Afrika-woordeboeke bestudeer om waar te neem hoe die moontlikheid om met eksterne bronne (bv. ander woordeboeke, ensiklopedieë, of selfs korpusdata) verbind te word, in hierdie bronne gebruik word. Daar word ondersoek ingestel na watter hiperskakels op watter plekke in die woordartikels voorkom en hoe hulle aan die woordeboekgebruikers aangebied word. Dit word gedoen teen die agtergrond van metaleksikografiese beskouings oor die beplanning van buitekenmerke en die mediostruktuur in aanlyn woordeboeke sowel as verskillende kategoriserings van hiperskakels in aanlyn naslaanwerke. Die resultate dui daarop dat retro-gedigitaliseerde woordeboeke feitlik glad nie gebruik maak van hiperskakels na eksterne bronne nie. Aan die ander kant maak ware aanlyn woordeboeke wel daarvan gebruik, maar dikwels in 'n vorm wat verbetering benodig, aangesien verklarings van woordeboek-eksterne skakels, byvoorbeeld, nie altyd in die gebruikersgids gevind word nie en hul ontwerp binne een woordeboek kan verskil.

**Sleutelwoorde:** AANLYN WOORDEBOEKE, RETRO-GEDIGITALISEERDE WOORDE-BOEKE, MEDIOSTRUKTUUR, HIPERSKAKELS, AFRIKATALE, DUITS

## 1.    Introduction

Hyperlinks to other sources are one type of outer feature in electronic dictionaries. These can be links to other dictionaries, to other reference sources, or to language tools (Klosa and Gouws 2015: 167). External information is made available to users through hyperlinks either to make more information available to them, or because no or no better information compiled within the dictionary project is available to the user (cf. Heid, Prinsloo and Bothma 2012: 269).

In this paper, such hyperlinks are examined on the basis of lexicographic practice in several (German) dictionaries of the dictionary portal *OWID* at IDS Mannheim[1] as well as in some African online dictionaries[2]. All of the dictionaries are freely available. The dictionaries published in the *OWID* (*Online-Wortschatz-Informationssystem Deutsch*) portal contain a number of different link types; there are, for example, links to grammatical online information systems, lexical online information systems, other online dictionaries (outside and inside *OWID*), or online encyclopaedias.[3] In the African dictionaries we also find hyperlinks to websites that provide users with background information about the publisher, etc.[4]

Different ways of categorizing hyperlinks in online dictionaries are presented as part of the discussion on the outer features of online dictionaries. The main part is a detailed analysis of hyperlinks in the dictionaries mentioned above. In the analysis, the metalexicographic model, which was developed in Klosa and Gouws (2015), will be examined for outer features in electronic dictionaries, and the necessity and helpfulness of links in specific situations of dictionary use will be discussed.

The description of practices in German and African dictionaries should support lexicographers when they are designing new dictionaries because "[i]n the planning of online dictionaries the lexicographers need to respond to the functions and genuine purpose of the envisaged dictionary when identifying existing sources that could be linked to the dictionary" (Klosa and Gouws 2015: 170).

## 2.    Categorization of hyperlinks

Hyperlinks are part of the mediostructure ("Mediostruktur", cf. Wiegand and Smit 2013), or, more appropriately expressed for the electronic medium, the reference structure ("Verweisstruktur", cf. Müller-Spitzer 2007: 169-180), of an online dictionary. Engelberg, Müller-Spitzer and Schmidt (2016: 158) introduce the distinction between link indicator ("Linkanzeiger") and link relation ("Linkrelation") to account for the fact that all hyperlinks in online dictionaries exist in both the data and presentation layers.

Link indicators can be seen and activated by users on the dictionary's web interface. Blumenthal, Lemnitzer and Storrer (1988: 359-360) classify link indicators into different groups: (a) the link consists of a link symbol (e.g. an arrow) and a representation of the link target (e.g. name of the web page); (b) the link

consists only of a representation of the link target (e.g. name of the web page); (c) the link consists only of a link symbol (e.g. an arrow). These types can be found in the dictionaries studied here and can be understood as an "invitation or suggestion by the lexicographer to perform a user action of following up a cross-reference" (Wiegand and Smit 2013: 221). They enable lexicographers to "make further information available without too much trouble for the users" (Apel 2014: 496).

There are two sub-types of hyperlinks: structural links ("Strukturlinks") and content links ("Inhaltslinks") (see Engelberg, Müller-Spitzer and Schmidt 2016: 159). Structural links allow users to navigate through an article, e.g. by opening tabs or additional windows. Content links allow users to find additional language-related content (e.g. translation equivalent, sound file with pronunciation). In the analysis of hyperlinks in German and African online dictionaries, the focus will be on content links; the different link indicators used are also examined. According to Mann (2010: 29), link indicators are colour (often the typical hyperlink (blue)), underlining, or the change of the mouse cursor when hovering over a hyperlink. They can all be found in the dictionaries examined in this study and are often designed according to the general design standards of websites.[5]

Regarding the hyperlink origin and destination, Storrer (2013: 1247) differentiates between internal and external links. When clicking on internal links, dictionary users are guided from one article to another within one dictionary or move from one position to a different position within one article. When clicking on external links, they are guided to articles in other (external) dictionaries. Blumenthal, Lemnitzer and Storrer (1988: 357-358) classify links in terms of their origin and destination and distinguish between article-internal links, article-to-article links, links from articles to outer texts, links from outer texts to articles, links from articles to outside the dictionary (e.g. source texts), and links from outer texts to outside the dictionary. In the analysis, the focus will be on links outside the dictionary, i.e. dictionary-external hyperlinks.

## 3.    Dictionary-external hyperlinks as part of the outer features

Klosa and Gouws (2015: 148) discuss the differences between outer features in printed and online dictionaries and suggest substituting the phrase "outer texts" with "outer features" (see also Müller-Spitzer 2013: 368), as not all elements to be found on dictionary websites belong to the category of texts. Many of the outer features to be found in online dictionaries are also not directly connected to the genuine purpose of a dictionary (cf. Klosa and Gouws 2015: 144-145) and they present data that are not strictly lexicographic data (cf. Klosa and Gouws 2015: 147).

Some outer features can guide users to dictionary-external sources (cf. Klosa and Gouws 2015: 149). These external sources include other dictionaries (e.g. other general dictionaries or a thesaurus), other reference sources (e.g. encyclopae-

dias or search for images) and language tools (e.g. translation tools or spell checkers) (see Klosa and Gouws 2015: 167 for examples in general online dictionaries of English). It should be noted that outer features in general and dictionary-external hyperlinks more specifically have to allow users to open other sources that are language-related and can be used in a situation of langue-related information need (cf. Klosa and Gouws 2015: 168). Therefore, hyperlinks to social media, information on products on websites etc. do not belong to this category. Overall, lexicographers should be fully aware of the fact that there are many external online sources that might prove valuable for dictionary users and which they do not need to compile within their own project (cf. Klosa and Gouws 2015: 170-171).

In the following section, the practices of some German and African online dictionaries will be examined. The results are discussed in the light of the number of dictionary-external hyperlinks they contain, their presentation and placement, and whether and how the User Guide explains which external sources were chosen as link targets.

## 4.      Dictionary analysis

### 4.1      German dictionaries in *OWID*

*OWID*, the *Online-Wortschatz-Informationssystem Deutsch* (Online German Lexical Information System) at the Leibniz Institute for the German Language in Mannheim is "a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the Institute for the German Language" (Engelberg, Klosa and Müller-Spitzer 2009: 16; see also Engelberg, Klosa-Kückelhaus and Müller-Spitzer 2020). The focus is on specialized dictionaries that cover areas of the lexicon that have been partially neglected by lexicography and lexicology. All dictionaries contained in this portal were examined for this study; only four of the *OWID* dictionaries include dictionary-external hyperlink.

(1)      *elexiko — Online-Wörterbuch zur deutschen Gegenwartssprache: elexiko* is an online information system for contemporary German that documents, explains and comments on vocabulary based on current language data in individual modules. It includes detailed articles on individual words, more than 50 word group articles on meaning relations, topic fields and word fields as well as more than 250,000 articles providing only automatically generated information on the headwords (orthographic information, corpus citations, frequency data).

(2)      *Neologismenwörterbuch:* This dictionary presents over 2,500 new words, new phrases and new meanings of established words that have been added to the general part of the vocabulary of the standard German language between 1991 and today. The new vocabulary of the last three decades can

be searched via various access routes (by subject group, via advanced search).

(3) *Paronymwörterbuch:* This dictionary documents easily confused expressions in their current public usage. The paronyms (which have strong similarities in spelling, pronunciation and meaning) are presented together in contrasting dictionary articles. Their similarities and differences can be seen at a glance, and users can decide which sections or comparative views they want to see.

(4) *Schlüsselwörter der Wendezeit 1989/90:* This dictionary is the online version of the reference book by Dieter Herberg, Doris Steffens and Elke Tellenbach, *Schlüsselwörter der Wendezeit. Wörter-Buch zum öffentlichen Sprachgebrauch 1989/90* (Keywords of the time of the German reunification. Wordbook of words in public language 1989/90) (Berlin/New York: De Gruyter 2007). It provides information on the more than 1,000 words and phrases relevant to the period of German reunification, arranged according to 150 keywords and thematic groups.

With *OWID^{plus}*, a new experimental platform has been established at IDS to complement the dictionary portal *OWID*. Here the user can find a multitude of individual, also multilingual resources, which are implemented modularly as independent interactive applications. One resource in *OWID^{plus}* is the prototype of a dictionary of spoken German, which also contains dictionary-external hyperlinks:

(5) *LeGeDe — Lexikographie des gesprochenen Deutsch:* The *LeGeDe* resource is a lexicographic prototype in which specific features of the lexis of spoken German are presented. The dictionary articles are elaborated on the basis of authentic interactions from a corpus of spoken German. Numerous audio examples from the corpus illustrate the meanings and functions described.

More than half of the ten dictionaries included in *OWID* do not provide users with dictionary-external hyperlinks in their articles or from the outer features, either to other dictionaries within the portal or outside. The reason for this is that not all dictionaries included in *OWID* were developed directly for online publication, but some are retro-digitized online dictionaries that by their nature only have a dictionary-internal mediostructure. In one case, however, the user is referred to corpora: in the *Sprichwörterbuch* the user can find the lexicographer's exact corpus query, which can then be copied into the IDS corpus query system *COSMAS II* to find additional corpus citations in *DeReKo* — the German Reference Corpus of IDS. *OWID* itself also offers a corpus query service[6] that allows users to dynamically retrieve authentic corpus citations from the IDS's publicly available German corpora in a straightforward and fast manner. This service is specifically intended to complement the information provided by the dictionaries included in *OWID*.[7]

The next section will take a closer look at the different forms of presentation of such hyperlinks, how many of them can be found in the articles and the different types of lexicographic information containing hyperlinks.

## 4.2 Analysis of dictionary-external hyperlinks in *OWID* dictionaries

### 4.2.1 *elexiko*

Articles with automatically generated content systematically contain hyperlinks to an online German grammatical information system[8] and to an online database of collocations[9] (see examples in article *Briefempfänger*[10] in Figure 1). The link targets are shown by blue colour as well as an arrow indicating that a new website will be opened by clicking on the hyperlink. In text production situations, these dictionary-external hyperlinks allow the user to switch from the headword in *elexiko* to inflection tables or to search statistically significant collocators for the headword. An information button in the online article allows the user to read an explanatory text on these hyperlinks in the *elexiko* glossary.[11]



**Figure 1:**   Dictionary-external hyperlinks in *elexiko*

Fully lexicographically described articles in *elexiko* provide a number of hyperlinks to other online sources.[12] In the User's Guide, a short text about hyperlinks[13] explains the types of hyperlinks to be found (other language information sources, encyclopaedic sources). Users also learn where (in articles, in the glossary) and why (because the information in the linked online information systems is a useful supplement to the information in the *elexiko* articles, which focuses on meaning and usage) hyperlinks are included and that they are not set automatically, but only after careful examination.

External hyperlinks in fully described articles can be found regarding different types of lexicographic information, e.g. definition, encyclopaedic information, etymology and historical meanings, or language use and criticism. In articles on function words such as the pronoun *all*[14] ('all'), users find a note ("Hinweis") on further morphological, syntactic and semantic information containing a hyperlink to the online grammatical information system *Grammis* at IDS.

In some articles, encyclopaedic information is presented in addition to the definition. The article *Brief* ('letter') contains a hyperlink to the online encyclopaedia *Wissen.de*, where the user finds information on the origins of this form of communication or different types of letters. This example also illustrates the representation chosen in *elexiko* for dictionary-external hyperlinks in fully described words: they are introduced with a short descriptive paragraph and are dated (see encyclopaedic information under "Sachinformation" in the article *Brief*[15] in *elexiko* in Figure 2).



**Figure 2:**    Encyclopaedic information in *elexiko*

Since *elexiko* was planned as a dictionary of contemporary German and is based on an extensive corpus of texts from 1950 onwards, the articles do not contain any specially compiled information on etymology or the historical development of the

various meanings, but represent bibliographic references to etymological or historical dictionaries of German. Hyperlinks to online reference works (e.g. *Deutsches Wörterbuch von Jacob und Wilhelm Grimm*) can also be found here.

Special types of information in *elexiko* are so-called language-reflective notes. For example, words that have been chosen as words of the year include a hyperlink to websites where information can be found about why they were chosen by the public, by journalists and by language critics.

These examples show that most of the dictionary-external hyperlinks given in *elexiko* are of interest to dictionary users not in situations of text production or reception, but in situations where the dictionary is supposed to tell stories about words, i.e. when the dictionary has a cognitive function. And while in some cases the hyperlinks help to fulfil the very purpose of the dictionary (i.e. to provide information about the German language and its usage), in other cases they go beyond the language purpose. Overall, the selection of places where hyperlinks could or should be inserted, both in the articles with automatically compiled information and in the articles with the full information programme, has been carefully planned for this dictionary, and the practical implementation is systematic and, because of the information given in the outer texts, transparent to the users as well.

### 4.2.2   *LeGeDe*

In the few (prototypical) articles in *LeGeDe*, three types of dictionary-external hyperlinks can be found (see article *gucken* in *LeGeDe*[16] in Figure 3): hyperlinks to a corpus (of spoken German: "FOLK"), to a lexical corpus exploration tool for spoken data ("Lexical Explorer"), and to other online dictionaries (in *OWID*, but also beyond). The hyperlinks are presented in two groups and each group has a heading (e.g. "Informationen in anderen lexikografischen Ressourcen" ['information in other lexicographical sources']). Hyperlinks are underlined and the mouse indicator changes when hovering over the link target. While in hyperlinks to the corpus and the lexical exploration tool, both the word in question (here *gucken*) as well as the name of the external source ("in FOLK", "in Lexical Explorer") are given, the hyperlinks to other online dictionaries do not repeat the headword.

A third group of hyperlinks under the heading "Forschungsliteratur" ('research literature') enables users to open pop-up windows with bibliographical references. However, these are not dictionary-external hyperlinks.

All linked external sources allow the user to further explore spoken data on the headword, supporting the cognitive function of the dictionary. The user manual explains where to find the hyperlinks and which dictionaries and other sources have been selected.
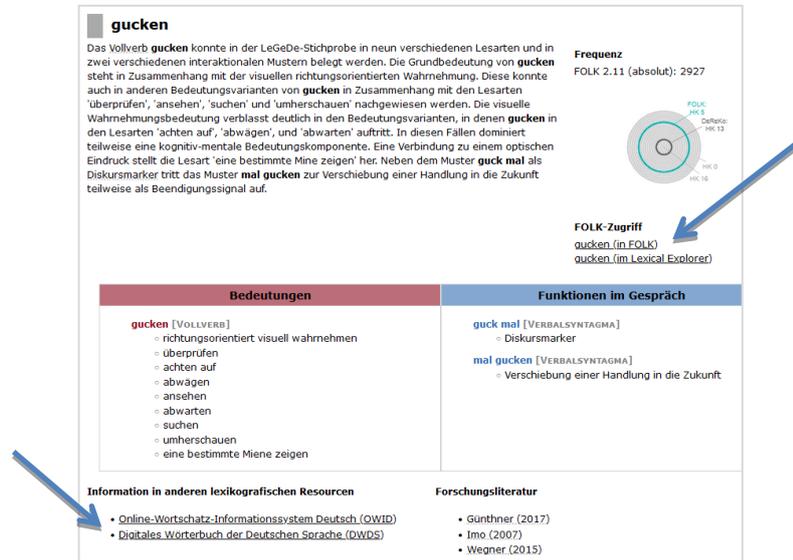
**Figure 3:**   Dictionary-external hyperlinks in *LeGeDe*

### 4.2.3   *Neologismenwörterbuch*

This dictionary on German neologisms offers users dictionary-external hyper-
links in the context of encyclopedic information or information on language
criticism. Here hyperlinks to other online dictionaries can be found which are
provided with a short explanatory text and show the headword in question as
a link target (see hyperlink to article *Debatte* in *Deutsches Fremdwörterbuch* in the
article *MeToo-Debatte* in *Neologismenwörterbuch*[17] in Figure 4). In these cases, dic-
tionary users are not presented with the name of the linked encyclopaedia, dic-
tionary, or other website, but the phrase "hier klicken" ('click here') serves as an
indicator of the hyperlink (in addition to the blue colour typical of hyperlinks
as well as the arrow symbol).



**Figure 4:**   Hyperlink to another dictionary in *Neologismenwörterbuch*

In the case of new meanings, a hyperlink is provided to a general online dictionary of German where the older meanings can be found (see hyperlink to a dictionary-external source to find information on the older meaning ("Ältere Bedeutung") for the new verb *wischen* ('to swipe') in *Neologismenwörterbuch*[18] in Figure 5).



**Figure 5:**   Hyperlink to a dictionary-external source in *Neologismenwörterbuch*

In line with this presentation, the hyperlink under the heading "Enzyklopädisches" ('Encyclopaedic information') also consists only of the text fragment "hier klicken" (see hyperlinks to encyclopaedic information and *Google Image Search* ("Google-Bilder") in article *Mikroplastik* in *Neologismenwörterbuch*[19] in Figure 6). Thus, the dictionary users cannot see the name of the link target. However, the user manual gives some information about the sources — in the case of encyclopaedic information, either the German *Wikipedia* or other websites with first-hand information.[20]

Rather specific hyperlinks are those to *Google Image Search* which — as far as possible — are offered as part of the encyclopedic information (see Figure 6) in the *Neologismenwörterbuch*. This hyperlink triggers a search in *Google Images* and serves to illustrate the concrete pictorial representation of what the headword denotes through a series of images. It should be noted that in these cases the link target does not present fixed content, but the content may change each time the hyperlink is clicked, since the link target is a search engine.



**Figure 6:**   Hyperlinks to encyclopaedic information in *Neologismenwörterbuch*

Many of the dictionary-external hyperlinks in the *Neologismenwörterbuch* contribute to fulfilling the dictionary's primary purpose (namely, to provide information about new words, phrases and meanings in German), but some go beyond that, taking users to websites with detailed encyclopaedic information. In text reception situations, hyperlinks to *Google Images* help users understand the meaning of neologisms, and other linked sources can help satisfy some users' curiosity about the origin of new words, etc.

### 4.2.4  *Paronymwörterbuch*

The *Paronymwörterbuch* contains dictionary-external hyperlinks to information under the heading "Wissenswertes" ('worth knowing'). Here, among other, information on use in terminology or in spoken language is given. In both cases, the user will find hyperlinks, either to a corpus of contemporary spoken German[21] or to an online dictionary of loan words (*Deutsches Fremdwörterbuch*) for information on the terminological usage of certain headwords.[22] These dictionary-external hyperlinks offer more language-related information to users in situations of text production, because they help users to differentiate between the easily confusable word pairs or triplets. They also support the cognitive function of the dictionary by providing users with the opportunity to learn more about written versus spoken language.

Unfortunately, the outer features of *Paronymwörterbuch* do not provide any information about the selection of sources for dictionary-external hyperlinks, and interestingly, the links are in fact not programmed as links, but only the URL is given in the online articles (see Figure 7).



**Figure 7:**   Dictionary-external references in *Paronymwörterbuch*

### 4.2.5  *Schlüsselwörter 1989/90*

In this dictionary, another form of external hyperlink to a dictionary is found: here, a web page showing all articles in groups contains a bundle of hyperlinks to *Neologismenwörterbuch* in *OWID*.[23] In an explanatory sentence, the user is informed that some articles contain hyperlinks to new lexical units that are described in the neologism dictionary (for an extract of the list, see Figure 8).



**Figure 8:**    Hyperlinks in *Schlüsselwörter 1989/90*

Link sources ("von" ['from']) and link targets ("zu" ['to']) are presented, and hyperlinks are identified only by the specific blue colour. They are clearly aimed at users in situations where they want to find (more) information about words, possibly in situations of text reception, but less so in contexts of text production. All hyperlinks here support the genuinely language information-oriented purpose of the dictionary.

### 4.2.6  Summary

Some *OWID* dictionaries contain a high variety of dictionary-external hyperlinks in various presentation forms. The sources that users may consult when following the links are also varied and include other dictionaries, online encyclopaedias, as well as links to the *Google Image Search*. But in those cases where print dictionaries were digitized, the *OWID* dictionaries do not offer dictionary-external links.

### 4.3      African online dictionaries

For this study, the following African online dictionaries were analysed:[24]

— *Aanlyn WAT* (Online version of the *Woordeboek van die Afrikaanse Taal*)
— *isiZulu.net: Bilingual Zulu–English Dictionary*
— *Mobile and Online Dictionary of South African English* (DSAE)

— *Multilingual N|uuki Dictionary*
— *Online Dictionary Chichewa/Chinyanja–English and English–Chichewa/Chinyanja*
— *Online Kiswahili (Swahili)–English Dictionary*
— *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete* (Online Explanatory Northern Sotho Dictionary)
— *Tšwela pelee-Eiwanika ly'Olusoga* (Online Lusoga Dictionary)

The *Aanlyn WAT* dictionary, the DSAE online dictionary, the *Chichewa/Chinaynja–English and English–Chichewa/Chinyanja online dictionary*, and the Lusoga online dictionary do not contain dictionary-external hyperlinks. All these dictionaries are retro-digitized dictionaries. The newly developed *Multilingual N|uuki Dictionary* does not contain dictionary-external hyperlinks either. As shown in Figure 9 (entry *!uu* ('woman's traditional apron') in the *Multilingual N|uuki Dictionary),* the only interactive element in this dictionary is the hyperlink button to play a sound file with the pronunciation of the headword.



**Figure 9:**   Entry *!uu* in the *Multilingual N|uuki Dictionary*

In the *Online Kiswahili (Swahili)–English Dictionary*, there are hyperlinks that take the user to "Microsoft's Swahili Terminology" database and "Microsoft's Language Portal".[25] These hyperlinks are integrated with the information licensed from "Microsoft's Swahili Terminology" that is displayed to users on the screen. In addition, there is a hyperlink to the website of "tlTerm" with information about this software application for creating terminology lists. All hyperlinks have the function of bibliographic references and do not provide the dictionary user with access to further language-related content. Thus, they are not counted as dictionary-external hyperlinks in this study.

In the online Explanatory Dictionary of Northern Sotho, users will find similar hyperlinks (e.g. to TshwaneDJe Software and Consulting[26]), but there is also a dictionary-external hyperlink to a website on African languages with

information about Northern Sotho[27] in the brief introductory text on the website (see the dictionary-external hyperlink to a website on African languages in article *maele* ('proverb, idiom') in *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete* in Figure 10). This could be counted as a dictionary-related hyperlink as the target source gives information on language that some users may find useful in a specific situation of information need. Both this dictionary and the *Online Kiswahili (Swahili)–English Dictionary* are based on printed dictionaries.



**Figure 10:**    Dictionary-external hyperlink in *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete*

The online only dictionary *isiZulu.net* offers a wide range of hyperlinks. In the left column on the screen users find a link to the dictionary forum page where

users can read information on updates, new entries and errata, i.e. meta-infor-
mation on the dictionary, under the heading "Dictionary forums". This may be
useful in cases of language-related information need and the hyperlink is thus
counted as a dictionary-external hyperlink. Links under the heading "Credits"
in the left column on the screen lead to different companies such as TLex or
Leo.org. All of these are not counted as dictionary-external hyperlinks, because
they have the function of bibliographic references and do not provide the dic-
tionary user with access to further language-related content.

**Learning isiZulu**
Cynthia's Podcast
isiZulu at Wikipedia
Wikipedia in Zulu
Oxford Bilingual School
Dictionary
IsiZulu & English
English-isiZulu / isiZulu-
English Dictionary
[C.M. Doke]
A Linguistic Analysis of
Zulu [G. Poulos]
Speak Zulu With Us
[African Voices]
**Readers**
African Storybook
Takealot
**Media**
Ilanga
Isolezwe
Ukhozi FM
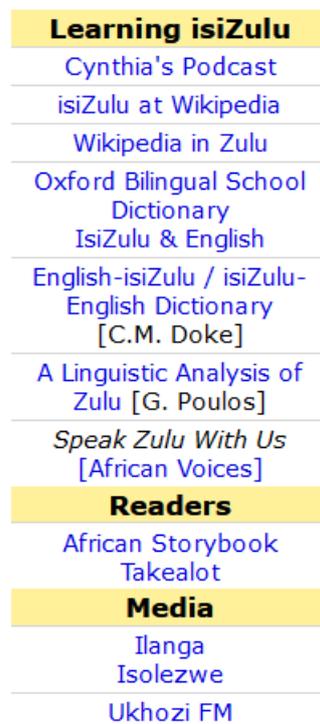
**Figure 11:**    Dictionary-external hyperlinks in *isiZulu.net*

In the right column on the screen, there are three groups of hyperlinks
(cf. Figure 11): "Learning isiZulu", "Readers", and "Media". The links under "Media"
lead to press webpages in isiZulu, one link under "Readers" leads to a website
with isiZulu storybooks, the other to a web shop where such storybooks can be
bought. All of these links are not counted as dictionary-external hyperlinks as
they offer a choice of texts in isiZulu that dictionary users may be interested to
read, but would not consult to gain additional information on a specific word.
Several links can be found under the heading "Learning isiZulu": two lead to web-
sites where print isiZulu–English dictionaries can be bought. One hyperlink
leads to the isiZulu *Wikipedia*, another to the entry *isiZulu* in the English *Wikipedia*.

The first is not counted as a dictionary-external hyperlink, because it is not useful in situations of text production or reception. The second is a dictionary-related hyperlink as the target source gives information on language that some users may find useful in a specific situation of information need. Similarly, three other links to a podcast in isiZulu, a website "Learn to speak Zulu" and the website "African Voices", where every day phrases in IsiZulu can be found, are dictionary-external hyperlinks.

The upper menu bar in *isiZulu.net* offers access to outer texts such as "Grammar", "Pronunciation" or "About". The menu "Usage" contains information on how to use the dictionary. Here, a hyperlink to the entry *Long and short scales* in the English *Wikipedia* is included, which is also categorized as a dictionary-external hyperlink. Finally, in the entries in *isiZulu.net*, there are several types of hyperlinks: all English or Zulu equivalents given are linked to the corresponding entries (i.e. dictionary-internal hyperlinks). There are also dictionary-internal links to conjugation tables. However, the abbreviations used are not linked to the abbreviation index that can be found in the menu "Usage". Overall, *isiZulu.net* offers the widest range of hyperlinks in the African online dictionaries in this study and all of them are presented in similar form (i.e. in blue font and by change of the mouse cursor when hovering over the hyperlink*)*.

Overall, the African online dictionaries in this study either do not offer dictionary-external links at all (mostly in cases where print dictionaries have been retro-digitized) or only (with the exception of *isiZulu.net*) a limited number of such hyperlinks, mostly to other websites with information on the specific language the dictionary covers.

## 5.      Conclusion

The German and African retro-digitized online dictionaries in this analysis do not provide dictionary-external hyperlinks, except for one case where a link target with language-related content can be found in the outer features. Nevertheless, lexicographers could and potentially should plan how to enrich the mediostructure of their retro-digitized dictionary through the introduction of dictionary-internal as well as dictionary-external hyperlinks that can be compiled automatically or by editorial means (see Klosa and Tiberius 2016: 77) to produce online dictionaries that offer more to users than their print counterparts.

In the genuine online dictionaries in this study, the number and extent of dictionary-external hyperlinks is usually higher. Here, hyperlinks leading from outer features (e.g. glossaries) to other websites with language-related content (e.g. grammatical information system, corpus) can be found. They also contain hyperlinks from articles to other online reference works (dictionaries and encyclopaedias) as well as to other informative websites and, in one case, to *Google Image Search*. These hyperlinks are often given when encyclopaedic information is added to the definition or when certain uses of the word need to be explained.

While in the first case, these hyperlinks may specifically be useful for native speakers, in the second case, they are addressed more towards (advanced) learners. Overall, the hyperlinks found in this study mainly support users in situations of both text production and text reception, and quite often they guide users to more language-related content, thus satisfying some users' needs for more information (i.e. cognitive function of dictionaries).

The example of the dictionaries in *OWID* shows that even within this dictionary portal not all dictionaries link to each other where this would theoretically be possible. When analysing the practice of these dictionaries, it also becomes clear that there seems to be no general rule for how dictionary-external hyperlinks are presented to dictionary users — regardless of existing recommendations on the design of online dictionaries or even of general website design rules. There are also great differences in whether and how the dictionaries in question explain their mediostructure to users in their User Guide. So even in cases of genuine online dictionaries, a careful plan for using dictionary-external hyperlinks as part of the outer features of the dictionary is not always discernible. Lexicographers should also pay attention to how useful and therefore necessary dictionary-external hyperlinks actually are. Linking to encyclopaedic information and/or images can support users in situations of text-production as well as text-reception, which are genuine dictionary functions; linking to websites with information on the language in questions satisfies information needs, which is often less important.

Based on the data in this study, it is still true that "lexicographers of online dictionaries need to realise that their venues for data go beyond the articles of the dictionary and even beyond the dictionary proper" (Klosa and Gouws 2015: 170), which is why it is so important that "online dictionaries need a well-devised data distribution structure" (ibid.). Only then will online dictionaries achieve results that support the genuine purpose and function of the dictionary envisaged and hence satisfy their users.

## Acknowledgement

## Endnotes

1.  The restriction on German online dictionaries in the *OWID* portal was adopted mainly because these dictionaries offer a wide range of dictionary types and should offer many hyperlinks as users can expect them to be hyperlinked within the portal.

2.    The basis for the choice of African dictionaries in this study is the information on (online) dictionaries by national lexicography units and Afrilex members provided on the website of the African Association for Lexicography: https://www.afrilex.co.za/useful-links. [Accessed: 24 October 2023]. One other dictionary mentioned in Prinsloo 2010 (*isiZulu.net*) was also included in this study because it is the only one offering a wide range of hyperlinks.

3.    There are also references to bibliographic information in printed sources, printed dictionaries and printed encyclopaedias. Of course, these references are not hyperlinks, but they are a different kind of outer features.

4.    See section 2. regarding this type of hyperlinks.

5.    On the advisability of following online design standards in online lexicography see Klosa-Kückelhaus and Michaelis 2022.

6.    See https://www.owid.de/service/cosmas. [Accessed: 25 October 2023].

7.    See information on https://www.owid.de/service/cosmas/info. [Accessed: 25 October, 2023].

8.    See *Verbformen.de*.

9.    See *CCDB*.

10.   https://www.owid.de/artikel/174720. [Accessed: 25 October 2023].

11.   https://www.owid.de/wb/elexiko/glossar/GrammatikKookkurrenzprofil.html. [Accessed: 25 October 2023].

12.   Articles in *elexiko* also contain references to print dictionaries, for example to etymological dictionaries, or to CD-ROM dictionaries (dating from a time when, for example, the Duden online dictionary was not yet published) and encyclopaedias (dating from a time when no freely usable encyclopaedic sources were available and *Wikipedia.de* was not yet as well developed as it is in 2022).

13.   https://www.owid.de/wb/elexiko/hilfe/beNachschlagen.html. [Accessed: 25 October 2023].

14.   https://www.owid.de/artikel/114743/Quantifikativpronomen. [Accessed: 25 October 2023].

15.   https://www.owid.de/artikel/174497/schriftliche%20Mitteilung. [Accessed: 25 October, 2023].

16.   https://www.owid.de/legede/article-main.jsp?id=Hut_gucken&level=hut. [Accessed: 25 October 2023].

17.   https://www.owid.de/artikel/407573. [Accessed: 25 October 2023].

18.   https://www.owid.de/artikel/403630 [Accessed: 25 October 2023].

19.   For example, in article *effektiv – effizient – effektvoll* ['effective – efficient – impactful'], see https://www.owid.de/artikel/407574. [Accessed: 25 October 2023].

20.   See https://www.owid.de/extras/neo/html-info/benutzerhinweise.html under "Enzyklopädisches". [Accessed: 25 October 2023].

21.   For example, in article *ethnisch – ethisch* ['ethnic – ethical'], see https://www.owid.de/parowb/artikel/effektiv_effizient_effektvoll/info/m%C3%BCndl_gebrauch. [Accessed: 25 October 2023].

22.   See https://www.owid.de/parowb/artikel/ethnisch_ethisch/info/fachsprache. [Accessed: 25 October 2023].

23.   https://www.owid.de/docs/swwz/map.jsp.

24.   See Prinsloo 2010 for general information on African online dictionaries and a list of URLs for African internet dictionaries (some of which were broken when accessed on 25 October 2023).

25.   Both hyperlinks are broken as on 10 October 2022.

26.   See https://tshwanedje.com/tshwanelex/.

27.   See https://africanlanguages.com/northern_sotho/. [Accessed: 10 October 2022].

## References

### A.     Dictionaries and other online sources

*Aanlyn WAT* — Online version of the *Woordeboek van die Afrikaanse Taal.*
     https://www.woordeboek.co.za/?timestamp=1569492060000.
*CCDB* — *Kookkurrenzdatenbank und korpuslinguistische Denk- und Experimentierplattform zur Explora-*
     *tion im Sprachgebrauch manifester Strukturen.*
     http://corpora.ids-mannheim.de/ccdb/.
*COSMAS II*. https://cosmas2.ids-mannheim.de/cosmas2-web/.
*Deutsches Fremdwörterbuch*. https://www.owid.de/wb/dfwb/start.html.
*Deutsches Wörterbuch von Jacob und Wilhelm Grimm*. https://woerterbuchnetz.de/?sigle=DWB#0.
*DeReKo* — *Deutsches Referenzkorpus*.
     https://www.ids-mannheim.de/digspra/kl/projekte/korpora/.
*Dictionary of South African English* (DSAE) *online*. https://www.dsae.co.za/.
*elexiko* — *Online-Wörterbuch zur deutschen Gegenwartssprache*.
     https://www.owid.de/docs/elex/start.jsp.
*Google Image Search*. https://www.google.de/imghp?hl=de.
*Grammis* — *Grammatisches Informationssystem*. https://grammis.ids-mannheim.de.
*isiZulu.net: Bilingual Zulu–English Dictionary*. https://isizulu.net/.
*LeGeDe* — *Lexik des gesprochenen Deutsch. Wörterbuchprototyp*. https://www.owid.de/legede.
*Multilingual N|uuki Dictionary*. https://dictionary.sadilar.org/.
*Neologismenwörterbuch*. https://www.owid.de/docs/neo/start.jsp.
*Online Dictionary Chichewa/Chinyanja–English and English–Chichewa/Chinyanja*.
     https://translate.chichewadictionary.org/.
*Online Kiswahili (Swahili)–English Dictionary*.
     https://africanlanguages.com/swahili/?timestamp=1569492060000.
*OWID* — *Online-Wortschatz-Informationssystem Deutsch*. https://www.owid.de.
*OWID^{plus}*. https://www.owid.de/plus.
*Paronymwörterbuch*. https://www.owid.de/parowb/.
*Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete* (Online Explanatory Northern Sotho Dictionary).
     https://africanlanguages.com/psl/?timestamp=1569492060000.
*Schlüsselwörter 1989/90* — *Schlüsselwörter der Wendezeit*.
     https://www.owid.de/wb/swwz/start.html.
*Sprichwörterbuch*. https://www.owid.de/wb/sprw/start.html.
*Tšwela pelee-Eiwanika ly'Olusoga* (Online Lusoga Dictionary*)*.
     https://www.menhapublishers.com/dictionary/?timestamp=1569492060000.
*Verbformen.de* — *Flexion aller deutscher Verben, Substantive und Adjektive*.
     https://www.verbformen.de.
*Wikipedia* — *die freie Enzyklopädie*. https://www.wikipedia.de.
*Wissen.de*. https://www.wissen.de.

## B.     Other literature

**Apel, U.** 2014. Linking a Dictionary to Other Open Data — Better Access to More Specific Information of the Users. Abel, A., Ch. Vettori and N. Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, Bolzano/Bozen, Italy, 15–19 July 2013*: 495-503. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.

**Blumenthal, A., L. Lemnitzer and A. Storrer.** 1988. Was ist eigentlich ein Verweis? Konzeptionelle Datenmodellierung als Voraussetzung computerunterstützter Verweisbehandlung. Harras, G. (Ed.). 1988. *Das Wörterbuch. Artikel und Verweisstrukturen. Jahrbuch 1987 des Instituts für deutsche Sprache*: 351-373. Düsseldorf: Schwann.

**Engelberg, St., A. Klosa and C. Müller-Spitzer.** 2009. Challenges to Internet Lexicography: The Internet Dictionary Portal at the Institute for German Language. *Kernerman Dictionary News* 2009: 10-19.

**Engelberg, St., A. Klosa-Kückelhaus and C. Müller-Spitzer.** 2020. Internet Lexicography at the Leibniz-Institute for the German Language. *K Lexical News* 28: 54-77.

**Engelberg, St., C. Müller-Spitzer and T. Schmidt.** 2016. Vernetzungs- und Zugriffsstrukturen. Klosa, A. and C. Müller-Spitzer (Eds.). 2016. *Internetlexikografie. Ein Kompendium*: 153-195. Berlin/Boston: De Gruyter.

**Heid, U., D.J. Prinsloo and T.J.D. Bothma.** 2012. Dictionary and Corpus Data in a Common Portal: State of the Art and Requirements for the Future. *Lexicographica* 28: 269-291.

**Klosa, A. and R.H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.

**Klosa, A. and C. Tiberius.** 2016. Der lexikografische Prozess. Klosa, A. and C. Müller-Spitzer (Eds.). 2016. *Internetlexikografie. Ein Kompendium*: 65-110. Berlin/Boston: De Gruyter.

**Klosa-Kückelhaus, A. and F. Michaelis.** 2022. The Design of Internet Dictionaries. Jackson, H. (Ed.). 2022. *The Bloomsbury Handbook of Lexicography*: 405-421. London: Bloomsbury Academic.

**Mann, M.** 2010. Internet-Wörterbücher am Ende der „Nullerjahre": Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien unter besonderer Berücksichtigung der Fachlexikographie. *Lexicographica* 26: 19-45.

**Müller-Spitzer, C.** 2007. Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis. Tübingen: Narr.

**Müller-Spitzer, C.** 2013. Textual Structures in Electronic Dictionaries Compared with Printed Dictionaries. A Short General Survey. Gouws, R., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 367-381. Berlin/Boston: De Gruyter.

**Prinsloo, D.J.** 2010. Internet Dictionaries for African Languages. *Lexicographica* 26: 183-194.

**Storrer, A.** 2013. Representing (computational) Dictionaries in Hypertextual Form. Gouws, R., U. Heid, W. Schweickard and H. E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 1244-1253. Berlin/Boston: De Gruyter.

**Wiegand, H.E. and M. Smit.** 2013. Mediostructures in Printed Dictionaries. Gouws, R., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 214-253. Berlin/Boston: De Gruyter.

# How to Solve Problems in Dictionary Use: The Georgian Case

Tinatin Margalitadze, *Centre for Lexicography and Language Technologies, Ilia State University, Georgia (tinatin.margalitadze@iliauni.edu.ge)*

and

George Meladze, *Centre for Lexicography and Language Technologies, Ilia State University, Georgia (giorgi.meladze.4@iliauni.edu.ge)*

**Abstract:** The present article was inspired by the plenary lecture given by Rufus Gouws at the XX International Congress of EURALEX in Mannheim: "Dictionaries: Bridges, Dykes, Sluice Gates" (Gouws 2022). His emphasis on the lack of a *dictionary culture* and the responsibilities that not only lexicography has towards society, but that society also has towards lexicography is an important perspective that is relevant for many societies, including Georgian society. While Gouws uses this term as an umbrella term to refer to both the user-friendliness of dictionaries, on the one hand, and the ability and willingness of society to use dictionaries and their knowledge of dictionary-use skills, on the other, in the present article we concentrate more on the latter aspect of the term *dictionary culture.*

Numerous studies conducted in the field of dictionary use since the 1970s have revealed that many users did not know how to use dictionaries, were not familiar with the structure of entries, could not identify what part of speech this or that word was, had no idea of homonymy, polysemy and related issues; this led to errors in comprehension, translation and other tasks related to foreign language acquisition. These studies, conducted in different countries, also clearly showed the need to teach dictionary use in schools and universities. Unfortunately, despite efforts from lexicographers and educators to improve this situation, and despite some positive steps taken in this direction which will be discussed in the paper, there is still no satisfactory teaching of dictionary use in educational institutions that would help create or sustain dictionary culture.

The aim of this article is to discuss problems in dictionary use in the Georgian context and share our experience of introducing the teaching of lexicography and developing dictionary-use skills at Georgian universities.

**Keywords:** DICTIONARY CULTURE, TEACHING DICTIONARY USE, DICTIONARY USE SKILLS, EXERCISES, DICTIONARY DIDACTICS, ENGLISH–GEORGIAN DICTIONARY

**Opsomming: Hoe om probleme in woordeboekgebruik op te los: Die Georgiese geval.** Hierdie artikel is geïnspireer deur die plenêre referaat wat deur Rufus Gouws by die XX Internasionale EURALEX Kongres in Mannheim gelewer is: "Dictionaries: Bridges, Dykes, Sluice Gates" ("Woordeboeke: Brue, dyke, sluisdeure") (Gouws 2022). Sy beklemtoning van die gebrek aan *woordeboekkultuur* en die verantwoordelikhede, nie slegs wat die leksikografie teenoor die gemeenskap het nie, maar wat die gemeenskap ook teenoor die leksikografie het, is 'n belangrike perspek-

tief wat relevant is vir baie gemeenskappe, insluitende die Georgiese gemeenskap. Terwyl Gouws hierdie term as sambreelterm gebruik om na beide die gebruikersvriendelikheid van woordeboeke, aan die een kant, en die vermoë en bereidwilligheid van die gemeenskap om woordeboeke en hul kennis van woordeboekgebruiksvaardighede te gebruik, aan die ander kant, te verwys, fokus ons in hierdie artikel meer op laasgenoemde aspek van die term *woordeboekkultuur*.

Verskeie studies wat sedert die 1970s in die woordeboekgebruiksveld uitgevoer is, het aangetoon dat baie gebruikers nie geweet het hoe om woordeboeke te gebruik nie, dat hulle nie bekend was met die struktuur van inskrywings nie, dat hulle nie die woordsoort van 'n woord kon identifiseer nie, dat hulle geen idee van homonomie, polisemie en verwante kwessies gehad het nie; dit het gelei tot foute in begrip, vertaling en ander take wat verband hou met vreemdetaalverwerwing. Hierdie studies wat in verskillende lande uitgevoer is, het ook die behoefte aan die onderrig van woordeboekgebruik in skole en universiteite duidelik aangetoon. Ongelukkig, ten spyte van pogings van leksikograwe en opvoeders om hierdie situasie te verbeter, en ten spyte van sommige positiewe stappe wat in hierdie rigting geneem is, wat in hierdie artikel bespreek sal word, is daar steeds nie bevredigende onderrig in woordeboekgebruik in opvoedkundige instansies wat daartoe sal bydra dat 'n woordeboekkultuur geskep of volhoubaar gemaak word nie.

Die doel van hierdie artikel is om probleme in woordeboekgebruik in die Georgiese konteks te bespreek en om ons ervaring in die implementering van die onderrig van leksikografie en die ontwikkeling van woordeboekgebruiksvaardighede aan Georgiese universiteite te deel.

**Sleutelwoorde:** WOORDEBOEKKULTUUR, DIE ONDERRIG VAN WOORDEBOEKGEBRUIK, WOORDEBOEKGEBRUIKSVAARDIGHEDE, OEFENINGE, WOORDEBOEKDIDAKTIEK, ENGELS–GEORGIESE WOORDEBOEK

## 1.      Introduction

In 2010 the online version of the *Comprehensive English–Georgian Dictionary* was posted on the internet (Margalitadze et al. 2010). It was based on the fascicles of the *English–Georgian Dictionary*, the publication of which began in 1995. We started a series of presentations of the Online Dictionary to different audiences just to find out that English–Georgian dictionaries were not used in teaching English, and even more bilingual dictionaries were banned from foreign language classes. This was the direct result of new methods for teaching foreign languages, which greatly diminished the role of translation, reduced the practice of using the native language in the process of teaching foreign languages, and prohibited the use of bilingual dictionaries, while the main emphasis was shifted to the use of explanatory, i.e. monolingual, dictionaries. This practice spread to schools and higher education institutions in Georgia. Nied Curcio, in her comprehensive survey of studies in the field of dictionary use, also considers the new methods of teaching foreign languages and banning bilingual dictionaries from foreign language classes as one of the factors responsible for the decline of interest in dictionaries (Nied Curcio 2022). A sceptical approach to these methods of teaching foreign languages was expressed by Arleta Adamska-Sałaciak in her article "Bilingual Lexicography: Translation Dictionaries" (2014).

"It is genuinely puzzling how methods which explicitly condemn the use of the native language in the classroom, effectively banning bilingual dictionaries, could ever have been considered beneficial in the teaching and learning of foreign languages", she wrote.

In 2011 we developed a training course for teachers of foreign languages: "Importance of Bilingual Dictionaries in Teaching Foreign Languages". During one year many training sessions were conducted with teachers of foreign languages both at schools and universities. The universities involved in the training sessions were Ivane Javakhishvili Tbilisi State University, Batumi Shota Rustaveli State University, and Kutaisi Akaki Tsereteli State University. Training sessions for teachers were held at Teachers' House in Tbilisi, attended by teachers from Tbilisi, as well as other towns in Georgia. The training course incorporated the theoretical part, which included such topics as the essence of word meaning, anisomorphism of languages, lack of equivalence between Georgian and English or other foreign languages, etc. It was stressed that while teaching foreign languages it was important to bring lexical, grammatical and syntactic information from a foreign language into correspondence with the respective lexical, grammatical and syntactic information from one's native language, and this is better and more efficiently achieved by means of bilingual dictionaries than by means of explanatory ones. The practical part demonstrated how the problem of equivalence was solved in the *Comprehensive English–Georgian Online Dictionary* and also gave teachers some practical advice on how to develop dictionary-use skills in students. We stopped the training sessions in 2012, as it was obvious that the official instructions that teachers received from the Ministry of Education of Georgia and from resource centres banned the use of the native language and bilingual dictionaries in foreign language classes. It was necessary to work with the official institutions responsible for working out the educational policies.

We started publication of articles on the current linguistic situation in the Georgian language, and problems in it caused mainly by the influence of English which, among other reasons, was the direct result of the methodology of teaching foreign languages (Margalitadze and Meladze 2016; Margalitadze 2018; Margalitadze 2020).

In 2022 the National Center for Educational Quality Enhancement in Georgia revised the benchmark for language acquisition programmes and introduced the use of bilingual dictionaries as an important method of teaching foreign languages. Nied Curcio underlines the same developments in many countries in her article: "we can find recommendations on dictionary use in educational guidelines and curricula in various European countries, i.e. Germany and Italy. The use of dictionaries is once again officially allowed; the practice, which had long been common, has thus been legitimized" (Nied Curcio 2022: 72).

Despite legitimizing dictionary use, the reality has not changed. Many teachers still do not use dictionaries in foreign language classes. There is no methodological support for those teachers who would like to introduce diction-

aries in their lessons and it is up to them to decide how to develop dictionary-use skills in students. There are no guidelines either for teachers or for students. Besides, one could also question how competent teachers are in developing study materials, exercises, etc. aiming at developing dictionary-use skills in learners. These are pressing questions in Georgia, but they are also the doubts and concerns in other countries as well (Nied Curcio 2022: 72-73).

As a result of the current state of affairs, lexicographers in many countries took it into their own hands to develop instructions, guidelines, exercises and other study materials, as well as practical modules in lexicography in order to assist the change of paradigm and help maintain a dictionary culture in their societies (see section 2.1).

This is the goal that we, as Georgian lexicographers, are pursuing in our country and which will be described in section 3 of the present article.

## 2.      Dictionary use

### 2.1      Research on dictionary use

In his article entitled "Dictionaries: Bridges, Dykes, Sluice Gates" Gouws notes the lack of dictionary culture in many societies and makes an important point by stressing that not only does lexicography bear a responsibility towards society, but society must in turn also accept its responsibility towards sustaining lexicography (Gouws 2022). Gouws uses the term *dictionary culture* as an umbrella term that includes both the user-friendliness of dictionaries, which is the responsibility of lexicographers, and the interest of society in lexicography, the ability and willingness to use dictionaries, knowledge of dictionary-use skills, etc. The decline of interest in dictionaries, followed by the gradual loss of dictionary-use skills, was revealed by numerous studies in the field of research on dictionary use conducted in many countries.

The scientific study of dictionary use began back in the 1970s. This included both quantitative and qualitative research on dictionaries and dictionary use (Quirk 1973; Tomaszczyk 1979; Mitchell 1983; Laufer and Hadar 1997; Atkins 1998; Béjoint 2010, among others); the total number of research studies conducted exceeds 200.

The studies in the field of dictionary use were important, as based on the feedback received from users, many dictionaries were improved from the point of view of both macrostructure and microstructure. However, the research also revealed that many users did not know how to use a dictionary effectively and how to extract the maximum information contained in it. Even more, surveys showed that there was no instruction on dictionary use in schools or universities. P. Márkus et al. write in their recently published paper, "Research on dictionary use has significantly changed the lexicographic practice and made dictionaries more user-friendly. Unfortunately, the same cannot be said for the field of dictionary didactics concerned with the education of users. No extensive teaching

of dictionary use is provided in schools and universities; this is especially the case in the Hungarian context" (P. Márkus et al. 2023).

In 1983, Evelyn Mitchell studied the look-up behaviours of 94 school students only to find out that the students did not read the contexts included in the dictionary entries, could not identify what part of speech this or that word was, and had no conception of the meaning or significance of homonymy. As a result, they selected the wrong homonyms and their translations were inevitably erroneous (Mitchell 1983).

Yukio Tono from the Tokyo University of Foreign Studies found out that his Japanese students were reluctant to look up words in dictionaries while reading and, if they still did, they read only the very first definitions, at the same time avoiding reading long definitions or/and explanations (Tono 2001).

Many studies also strongly indicated that without proper prior training and lexicographic education, students tend to use various types of dictionaries erroneously, inefficiently and ineffectively. The comparison of how trained/ experienced and untrained users look up words in dictionaries showed that these two categories of dictionary users have very different approaches to dictionary use. Rundell (1999) in particular pointed out that inexperienced users tend to look up only the very first sense of a dictionary entry. Such a superficial approach to dictionary use results in an inadequate command of foreign languages. Language competence in such cases becomes shallow and superficial. Another negative consequence of the neglect of dictionaries is the inability of such students to communicate their thoughts correctly, poor vocabulary, and a failure to associate foreign words and concepts with their proper counterparts from their native language or languages (Béjoint 2010).

Thus, along with the concern for the improvement in the techniques and skills required to dictionaries, it is also necessary to educate and teach students and the younger generation how to use dictionaries properly. Educators are well aware of this, and as a result, many countries have incorporated instruction on the use of dictionaries into their curricula.

In schools in Greece, for example, students are taught what a dictionary is, how to use dictionaries effectively, and how to extract maximum information from them (cf. Gavriilidou and Konstantinidou 2021). For this purpose, textbooks have been written that include special "lexicographical exercises". Elementary school students who complete these exercises develop some very useful skills in dictionary use. They are taught separately what the polysemy of lexical units means and how such polysemy is reflected in dictionary entries (Gavriilidou and Konstantinidou 2021).

In Lithuania, the 20-volume Academic Dictionary of the Lithuanian Language, which is rightfully regarded as a "major landmark of Lithuanian linguistics", is included in the curricula of Lithuania's secondary schools (Kavaliauskas 2012).

Open online lecture courses, initiated and masterminded by Michael Rundell, deserve special mention. The courses are well-structured, based on a pre-

designed plan, and provide the attendees with all the necessary information concerning dictionary use and dictionary making (Creese et al. 2018).

In 2023 P. Márkus published a workbook that contains specially developed exercises for students. Exercises train students in the use of dictionaries, both monolingual and bilingual. Exercises in training dictionary use in students cover many different issues (P. Márkus 2023).

Marello and her colleagues from Turin University created a website of online exercises, also aiming at the development of dictionary-use skills and so improving language competence (Marchisio et al. 2019).

## 2.2      Studies on dictionary use in Georgia

The information about the abovementioned studies, acquired from various sources such as international scientific fora, proceedings of international scientific conferences and symposia, etc., inspired us to conduct similar research in our country within the framework of MA and PhD programmes in lexicography at Ivane Javakhishvili Tbilisi State University (TSU), in order to find out how the dichotomy of trained and untrained dictionary users would play itself out in Georgia.

The first person to conduct such a study was Dr Antonina Kapanadze, then a doctoral student of the Doctoral Programme in Lexicography. She conducted two-stage research using special questionnaires aimed at the identification of dictionary-use skills in experimental subjects. The first stage involved the general public from the Georgian capital Tbilisi. The second stage focused on the students of TSU. For the second part of her study, Kapanadze relied on the experiment conducted by Laufer and Hadar (1997) to find out which types of dictionaries — monolingual, bilingual or bilingualized — were most effective both in language comprehension and language production.

The next study was the research conducted in 2019 by Khuskivadze, a PhD student in lexicography at TSU. The research took place in Tbilisi Public School #24, where she is employed as an English teacher. The research was focused on two separate groups of test subjects, group A and group B. The members of one group had received preliminary training in dictionary use with the help of exercises, specially compiled for this study, while the members of another group did not receive such training. In the second stage, students, including groups A and B, were given specially prepared contexts that contained familiar words but in rare senses, unknown to students. The students had to find out the contextual meanings of those words with the help of dictionaries. As material for this part of the experiment, Khuskivadze used the contexts, worked out by Donna Farina of New Jersey University together with her Slovenian colleagues for their study at Ljubljana University (Farina et al. 2019). She had written consent from the authors for the application of this material. The test series showed that those students/subjects who had received preliminary training showed better results in dictionary use (Khuskivadze 2023).

In 2020, another research project was conducted by Tskikolia, an MA student at TSU. The test subjects of the qualitative research conducted by Tskikolia were MA and BA students from TSU. Due to the restrictions imposed in response to the COVID-19 pandemic, the research was conducted through the online platform, Zoom. The aim of the research was to determine the extent to which the students had the necessary skills with regard to effective dictionary use.

One more study was implemented by our PhD student, Vardidze, who developed detailed plans of lessons for teaching polysemy at schools; they were based on her study of polysemous models of nouns in Modern English (cf. Vardidze 2020).

As expected, the findings of the studies conducted in Georgia generally correspond with those of the abovementioned dictionary research carried out by foreign lexicographers. They strongly indicate that education and training in dictionary use are absolutely necessary and that without such education, university students, and the students of languages in particular, end up with poor dictionary-use skills, with poor language proficiency and scanty vocabulary, both with respect to foreign and their native languages.

## 3.     Introduction of teaching lexicography at Georgian universities

### 3.1     Lexicography as a theoretical subject

"Introduction to Lexicography", a theoretical course of lectures for BA students, was introduced at Tbilisi State University in 2010. It was the first step toward familiarizing students with this discipline, and its complex, multidisciplinary nature.

The main topics of the theoretical course are as follows:

— the disciplinary status of lexicography;
— the functions of lexicography;
— macrostructure and microstructure of a dictionary and their components;
— the meaning of the word and methods of its study;
— documenting a language and comprehensive monolingual dictionaries;
— bilingual dictionaries and the problem of equivalence;
— learner's lexicography, the history of the genre and its characteristic features, the big five, bilingual learner's dictionaries;
— terminological dictionaries, terms and concepts, structural and semantic characteristics of modern terminology, problems of Georgian terminology;
— digital lexicography;
— history of Georgian monolingual lexicography;
— history of Georgian bilingual lexicography.

Alongside theoretical issues, students learn about academic dictionaries created by Georgian and foreign lexicographers, monolingual as well as bilingual and

terminological. In 2022 Margalitadze's manual "Introduction to Lexicography" was published by Ilia State University Press (Margalitadze 2022). The manual is not restricted to cater to the needs only of students of English departments, but is more general, incorporating examples from other languages. Still, the emphasis is on English as the Georgian language, at the current stage of its development, is mostly influenced by the English language. At the end of each chapter of the manual, there are practical tasks that students perform independently with the help of existing Georgian dictionaries. Students also prepare presentations dedicated to different lexicographic issues, famous Georgian and foreign lexicographers, and various types of dictionaries: slang, dialectal, synonyms, etymological, idioms, etc.

At present the theoretical subject "Introduction to Lexicography" is taught at four universities in Georgia. Students find the theoretical course both interesting and useful. At the end of the course they usually stress the usefulness of the knowledge acquired during this course for their future careers. Our experience of teaching this subject revealed that it was not sufficient to teach theory; it was also necessary to develop a series of practical exercises for students aimed at developing dictionary-use skills. We identified the most problematic issues, mostly connected to the meaning of words, understanding the complexity of the interaction between English and Georgian lexis, the problem of equivalence, polysemy, homonymy, conversion, phrasal verbs, idioms and terminology. These findings prompted us to develop exercises for the treatment of the abovementioned issues. Exercises are being developed together with the PhD students of the "Digital Lexicography" programme at Ilia State University who are involved in teaching lexicography to BA students.

### 3.2    Development of dictionary-use skills: Exercises

In this section, we describe some types of exercises we developed for engaging with homonymy, conversion and polysemy.

Exercises on conversion are quite simple at the initial stage. Example 1 shows such a case, when students are asked to circle the correct part of speech and they have to select the correct answer from two options.

(1) **SIDE**
   1. The side of the car was damaged.
      a. noun   b. adjective
   2. The side mirror was broken.
      a. noun   b. adjective

In the next type of exercises, students have to select the correct answer from a wider range of options, as is shown in example 2. Example 3 demonstrates the exercises in which students are given sentences without any suggestions about possible parts of speech. These sentences contain words known to them, but

there are also exercises that contain unknown words and students have to find out the part of speech as well as the lexical meaning of a word. As mentioned in previous sections, one of the problems identified in many studies on dictionary use is the fact that learners cannot identify what part of speech this or that word belongs to. The aim of exercises of types 1 and 2 is to help students understand and realize that a word may have the same meaning but belong to different parts of speech. While working on these exercises students are asked to complete them without the help of dictionaries. After completion of the task, they work in pairs, check words in dictionaries, and discuss examples. As for the exercises of type 3, in this case, students work with the help of dictionaries and they determine both, the part of speech of a word and its lexical meaning.

(2) **BACK**
    1. Nick was not careful and fell on his back.
      a. verb   b. noun   c. adverb   d. adjective
    2. When we were children we liked to sit in back rows at the cinema.
      a. verb   b. noun   c. adverb   d. adjective
    3. I could not notice him, because he stood back in the crowd.
      a. verb   b. noun   c. adverb   d. adjective
    4. We seldom back his plan.
      a. verb   b. noun   c. adverb   d. adjective

(3)  He was three days back in his work.
    I have a back view of things.
    We walked toward the garden at the back of the house.

Exercises on homonymy contain pairs of homonyms or converted forms and students have to identify which pair of words represents homonyms and which pair is an example of conversion, e.g. example (4).

(4)  **bargain**  *n*  something you buy cheaply or for less than its usual price.
    **bargain**  *n*  a small farm-holding

    **barrow**  *n*  a hill, hillock.
    **barrow**  *n*  a castrated boar.

    **beal**  *v*  a pustule or boil.
    **beal**  *n*  the mouth of a river.

    **beat**  *n*  a stroke or blow in beating.
    **beat**  *v*  to strike with repeated blows.

Type 4 exercises concentrate on the differences between converted forms and homonyms. They help students realize that there are words with the same meaning, but they belong to different parts of speech, while there are also words that are spelled and pronounced in the same way but convey totally different

meanings. Students write out example sentences from bilingual and monolingual dictionaries, compare and discuss them in pairs or groups.

The above-discussed types of exercises (1, 2, 3 and 4) develop one of the most important skills in learners, namely the ability to look for the right word in a dictionary, i.e. the right part of speech and the right homonym.

Students particularly enjoy and find useful exercises on polysemy. We usually start with simple exercises, like example (5) in which students are given several polysemous meanings of a word from the *Comprehensive English–Georgian Online Dictionary,* followed by several sentences containing this word in one of the senses and they have to circle the correct sense of a word.

(5)  **TO BREAK**

1. გატეხა (გატეხს);
2. დამტვრევა (დაამტვრევს; დაიმტვრევა); დამსხვრევა;
3. გარღვევა (გაარღვევს; გაირღვევა); გაწყდომა; გაგლეჯა;
4. შესუსტება (შეასუსტებს);
5. გაღარიბება (გააღარიბებს), გაკოტრება;
6. ცნობება (აცნობებს), უწყება, შეტყობინება;
7. გატეხა (გატეხს), დარღვევა (*სიტყვისა, დაპირებისა, კანონისა და ა.შ.*).

   a.  Tom **broke** the news of her husband's death to Mary.
       **1**    **2**    **3**    **4**    **5**    **6**    **7**
   b.  The money lenders soon **broke** him.
       **1**    **2**    **3**    **4**    **5**    **6**    **7**

The next exercises on polysemy are more difficult. Students are given sentences, in which familiar words are used in senses not known to them (example 6). In order to find the contextual meaning of *sharp* in the second sentence, for example, students have to read the whole entry of this highly polysemous word. In the *Comprehensive English–Georgian Online Dictionary,* this meaning of *sharp* '(of clothes or their wearer) smart and stylish' is described under sense 10. Such exercises make students understand that the contextual meaning of a word is not always the first sense in an entry, a common misconception among users.

(6)  1. Governments around the world have raised **interest** rates in a desperate attempt to keep post-pandemic inflation under control.
    2.  The gentleman attracted everyone's attention in his **sharp** blue Dolce & Gabbana suit.
    3.  The influx of refugees **taxed** the scarce water resources of this small village.

In exercises of type 7, students have to insert omitted words in sentences and the words to be inserted are either used in different senses of a word or are converted forms or homonyms (example 7). For example, *tip* in the first sentence of example 7 is used in the sense 'to give (someone) a small sum of money

as a reward for performing a service, such as serving in a restaurant'. However, in the 4th sentence, students have to insert a homonym of *tip*, meaning 'overbalance so as to fall or turn over'. Likewise, *book* in the second sentence is a verb and means 'to engage (a performer or guest) for an event', while in the third sentence, it is a noun, a converted form, meaning 'a main division of the Bible'. Students use different online dictionaries for these exercises. After accomplishing the task, they discuss their findings about different senses of polysemous words or converted/homonymous pairs and their meanings.

(7) Insert the omitted words:  **TIP  BOOK  PUT**

1. We generally _____ about 5 GEL per drink.
2. Your band is _____ for the school party.
3. Even though he swore on the _____ he was telling the truth, he was still sent to jail.
4. The little girl _____ the vase full of water all over the new carpet.
5. The agreement was never _____ into writing.
6. They feared being _____ to death.

The aim of the exercises is to develop a habit in students to consult a dictionary, to check the meanings of words in different sources, both monolingual and bilingual, and to appraise the efficiency of dictionaries. Discussions after completion of exercises also help students better understand difficulties stemming from the anisomorphism of native and foreign languages, polysemy, homonymy, or other linguistic issues that are quite problematic to be dealt with by learners without the help of relevant reference sources.

### 3.3    Course of online lessons

In 2022 we started recording and publishing on the internet a series of online lessons entitled "Why Consult a Dictionary". The video lessons are recorded on the Zoom platform. Lessons are short, 8–10 minutes long, and each lesson is dedicated to the analysis of one particular translation mistake. The lessons analyse mistakes with the help of relevant dictionary entries and demonstrate how these mistakes could have been avoided if dictionaries had been consulted.

Below we describe one type of video lesson, dedicated to the analysis of cases when an English word is polysemous but its equivalent in Georgian has only one meaning. Often, the polysemous meaning of an English word is ascribed to the Georgian equivalent.

For example, *classified* has two meanings in English: 1. arranged in classes or categories, and 2. (of information or documents) designated as officially secret and accessible only for authorized people. The Georgian equivalent კლასიფიცირებული '*klasipicirebuli*' has only the first meaning, but კლასიფიცირებული '*klasipicirebuli*' is often erroneously used in the second sense

of *classified* in Georgian and so instead of "NATO's classified information", we get the Georgian translation, which literally means "NATO's information arranged in classes".

*Roadmap* has two senses in English: 1. a map showing roads, especially for automobile travel, and 2. a detailed plan to guide progress toward a goal. The Georgian equivalent of this word საგზაო რუკა '*sagzao ruka*' has only the first meaning of its English counterpart. Despite this, საგზაო რუკა '*sagzao ruka*' is used to render the second meaning of English *roadmap.* As a result, the English sentence "Roadmap for accession to the European Union" is translated literally and the Georgian translation is "a map showing roads for accession to the European Union".

*A bald eagle*, a beautiful bird, in one of the TV programmes, was translated in Georgian literally, მელოტი არწივი '*meloṭi arċivi*', as an eagle with a bald head, i.e. in the primary sense of the English adjective *bald* "without hair on the head". *Bald* is a highly polysemous word. One of its senses is "having white feathers or skin (of birds and animals)". This bird has white feathers on the head and *bald* in this case means not "without hair", but "having white feathers".

Each video lesson contains screenshots of relevant dictionary entries, both monolingual and bilingual, with instructions on how to navigate in them and how to avoid mistakes by consulting dictionaries.

These, as well as similar mistakes, are collected and used not only in video lessons but also in exercises on polysemy, as discussed above.

## 3.4    Establishment of an international conference in lexicography

For the wider popularization of lexicography in Georgia, a biannual international conference was launched in 2022 organised by Ilia State University and Batumi Shota Rustaveli State University with the theme "Lexicography in the XXI Century". The conference will be held alternatively at Ilia State University and Batumi State University. This will draw attention to lexicography not only in the capital city, but in Western Georgia as well. The aim of the conferences is to create a forum in Georgia for the discussion of vital issues of modern lexicography, including issues of dictionary didactics.

The rapid development of artificial intelligence and computer and mobile technologies has created new prospects for lexicography. Online dictionaries, mobile dictionary applications, machine translation programs, vast text corpora, and other advances in modern digital humanities pose new challenges to lexicographers. Lexicography and modern philology in general have become truly multidisciplinary fields. In an interdisciplinary activity such as modern lexicography scholars need to be innovative, to look for new directions, new methods and new approaches. Yet with the advent of digital technologies in philology, excessive preoccupation with new technologies and relying upon them too exclusively may be sometimes counterproductive. Here too, a creative combination

of various methods and approaches is of great importance.

All of this points to the need for intensive communication among scholars working in this field, frequent contacts and exchanges of experience. In terms of the circumstances described above, the current situation of Georgia differs from that of developed European countries. There is still much to be done in Georgian lexicography. Therefore, we consider it important to establish such a forum in Georgia, which will contribute to the proper development of this important field both in our country and in countries similar to Georgia. The conference will contribute to familiarizing scholars working in this field with European as well as general, international knowledge and expertise; it will promote the popularization of lexicography and will encourage talented young people to work in this field. At the conference, special attention will be devoted to the participation of young researchers as there will be sessions dedicated to MA and PhD students. We hope that the biannual conferences will play an important role in the popularization of teaching dictionary-use skills in schools and universities.

## 4.      Conclusion

With the development of new directions in theoretical lexicography and research on dictionary use, the quality, structure and user-friendliness of existing dictionaries have improved considerably. Unfortunately, we have also witnessed the decline of interest in dictionaries. Many studies carried out in the field of dictionary use have revealed that students do not have elementary skills in dictionary use; they even cannot identify what part of speech this or that word is. Some students do not even know that online resources are also dictionaries and regard this term referring to something obsolete and outdated. One of the reasons for the decline of interest in dictionaries lies in the new methods of teaching foreign languages, which banned the use of bilingual dictionaries in foreign language classes.

The latest developments indicate that the educational policy has changed and dictionary use is now officially legitimized in the official educational guidelines of many countries. This important change has not affected dictionary use at all. There is still no training in dictionary use in schools and universities; there are no guidelines for teachers on how to develop dictionary-use skills in students.

Many lexicographers in different countries took the initiative into their own hands and undertook the development of study materials, workbooks with exercises both for monolingual and bilingual dictionaries, and online platforms for learners.

Lexicography in Georgia largely followed the same trend: decline of interest in dictionaries, adopting new methods of teaching foreign languages, banning bilingual dictionaries from foreign language classes, implementing changes to

the educational policy recently and the introduction of dictionary use in official guidelines; yet this did not show any palpable results for the discipline and in our efforts to maintain dictionary culture in our country.

Our studies have revealed that students enjoy working with dictionaries, once they are taught how to do this effectively; they like targeted lexical exercises and at the end of the semester they show better results in dictionary use, which is tested with specially devised exercises. Students like discussing theoretical issues of lexicography; they discover how much fun it is to explore meaning, to find links between the senses of a polysemous word, and so on.

At this stage, we think that lexicographers from different countries should combine their experience and efforts to develop unified programmes and guidelines for teachers, and produce workbooks with exercises for students aiming at the development of dictionary-use skills. At the same time, it is important to hold conferences and round-table discussions dedicated to the analysis of existing problems in our methods of education and devise concrete steps for their improvement. We are pursuing this policy in our country and it is encouraging to see that lexicographers in other countries also come to the same conclusions. Gouws is right (Gouws 2022) to stress the responsibility of society towards sustaining our discipline, but it is probably also our duty to explain to our societies the importance of this discipline for the health of our languages, education and culture. We need to gain the support of the media in this endeavour, and above all to convince and win the support of teachers who are the ones to introduce the teaching of dictionary use in their classes, both at schools and at universities.

## References

**Adamska-Sałaciak, A.** 2014. Bilingual Lexicography: Translation Dictionaries. Hanks, P. and G.-M. de Schryver (Eds.). 2014. *International Handbook of Modern Lexis and Lexicography:* 1-11. Berlin/Heidelberg: Springer.
DOI: 10.1007/978-3-642-45369-4_6-1 #

**Atkins, S.** 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators.* Tübingen: Max Niemeyer.

**Béjoint, H.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.

**Creese, S., B. McGillivray, H. Nesi, M. Rundell and K. Sule.** 2018. Everything You Always Wanted to Know about Dictionaries (But Were Afraid to Ask): A Massive Open Online Course. Čibej, J., V. Gorjanc, I. Kosem and S. Krek (Eds.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global* Contexts, 17–21 July 2018, Ljubljana, Slovenia: 59-66. Ljubljana: Ljubljana University Press, Faculty of Arts.

**Farina, D.M., 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.

**Gavriilidou, Z. and E. Konstantinidou.** 2021. The Design of an Explicit and Integrated Intervention Program for Pupils Aged 10–12 with the Aim to Promote Dictionary Culture and Strategies. Gavriilidou, Z., L. Mitits and S. Kiosses (Eds.). 2021. *Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, Alexandroupolis, Greece, 7–9 September 2021. Volume 2*: 735-745. Komotini, Greece: Democritus University of Thrace.

**Gouws, R.** 2022. Dictionaries: Bridges, Dykes, Sluice Gates. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs and P. Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress, 12–16 July 2022, Mannheim, Germany*: 36-52. Mannheim: IDS-Verlag.

**Kapanadze, A.** 2017. Examining the Effectiveness of Monolingual, Bilingual and Bilingualized Dictionaries for Georgian Learners. PhD dissertation, Ivane Javakhishvili Tbilisi State University.

**Kavaliauskas, V.** 2012. Academic Dictionary of the Lithuanian Language — The Most Outstanding Monument of the Lithuanian Language. *Proceedings of the II International Symposium in Lexicography in Batumi, Georgia, 18–20 May 2012.* Tbilisi: Meridian.

**Khuskivadze, E.** 2023. Development of Dictionary Use Skills in Learners of English. Margalitadze, T. (Ed.). 2023. *International Conference "Lexicography in the XXI Century", 10–12 November, 2023. Book of Abstracts:* 148-149. Tbilisi: Ilia State University.

**Laufer, B. and L. Hadar.** 1997. Assessing the Effectiveness of Monolingual, Bilingual, and "Bilingualised" Dictionaries in the Comprehension and Production of New Words. *The Modern Language Journal* 81(2): 189-196.

**Marchisio, M. et al.** 2019. Adapting STEM Automated Assessment System to Enhance Language Skills. *The 15th International Scientific Conference eLearning and Software for Education, Bucharest, 11–12 April 2019. Volume 2:* 403-410. Bucharest: "CAROL I" National Defence University Publishing House.

**Margalitadze, T.** 2018. Once Again Why Lexicography is Science. *Lexikos* 28: 245-261.
Available at: https://doi.org/10.5788/28-1-1464

**Margalitadze, T.** 2020. Language and Ecology of Culture. *Lexicographica* 36: 225-240.
Available at: https://doi.org/10.1515/lex-2020-0012

**Margalitadze, T.** 2022. *Introduction to Lexicography.* Tbilisi: Ilia State University Publishing House.

**Margalitadze, T. and G. Meladze.** 2016. Importance of the Issue of Partial Equivalence for Bilingual Lexicography and Language Teaching. Margalitadze, Tinatin and George Meladze (Eds.). 2016. *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity. Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia, 6–10 September, 2016*: 787-797. Tbilisi: Ivane Javakhishvili Tbilisi State University.

**Margalitadze, T. et al.** 2010. *Comprehensive English–Georgian Online Dictionary.* Tbilisi: Lexicographic Centre.
Available at: www.dict.ge

**Mitchell, E.** 1983. *Search-do Reading: Difficulties in Using a Dictionary*. Formative Assessment of Reading, Working Paper 21. Aberdeen: Aberdeen College of Education.

**Nied Curcio, M.** 2022. Dictionaries, Foreign Language Learners and Teachers. New Challenges in the Digital Era. Klosa-Kückelhaus, A., S. Engelberg, Ch. Möhrs and P. Storjohann (Eds.). 2022. *Dictionaries and Society: Proceedings of the XX EURALEX International Congress, 12–16 July 2022, Mannheim, Germany:* 71-84. Mannheim: IDS-Verlag.

**P. Márkus, K.** 2023. *Teaching Dictionary Skills.* Budapest: L'Harmattan.

**P. Márkus, K. et al.** 2023. Dictionary Skills in Teaching English and German as a Foreign Language in Hungary: A Questionnaire Study. *International Journal of Lexicography* 36(2): 173-194.

**Quirk, R.** 1973. The Social Impact of Dictionaries in the U.K. McDavid, R. and A. Ducket (Eds.). 1973. *Lexicography in English*: 76-88. New York: Annals of the New York Academy of Sciences.

**Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-53.

**Tomaszczyk, J.** 1979. Dictionaries: Users and Uses. *Glottodidactica* 12: 103-119.

**Tono, Y.** 2001. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension.* Lexicographica. Series Maior 106. Berlin/Boston: Max Niemeyer.

**Vardidze, T.** 2020. The Impact of Teaching English Polysemy through Similarity-Based Approach. *Journal of Education in Black Sea Region* 5(2): 112-125.

# Gender Stereotypes in Dictionaries: The Challenge of Reconciling Usage-based Lexicography with the Role of Dictionaries as Social Agents

Carolin Müller-Spitzer, *Institut für Deutsche Sprache, Mannheim, Germany (mueller-spitzer@ids-mannheim.de)*

**Abstract:** In many countries of the world, perspectives on gender equality and racism have changed in recent decades.[1] One result has been more attention being devoted to traces of androcentric and racist language in society. This also affects dictionaries. In lexicography there are discussions about whether or to what extent social asymmetries are inscribed in dictionaries and if this is still acceptable. The issue of the nature of description plays an important role in this discussion. If sexist usages are often found in language use, i.e. in the corpus data on which the dictionary is based, does the dictionary also have to show them? How is this, in turn, compatible with the normative power of dictionaries? Do dictionaries contribute to the perpetuation of gender stereotypes by showcasing them under the banner of descriptive principles? And what roles do lexicographers play in this process? The article deals with these questions on the basis of individual lexicographical examples and current discussions in the lexicographic and public community.

**Keywords:** GENDER AND LANGUAGE, GENDER STEREOTYPES, COLLOCATIONS, CORPUS-BASED LEXICOGRAPHY, DICTIONARIES AS SOCIAL AGENTS

**Opsomming: Genderstereotipes in woordeboeke: Die uitdaging om die gebruiksgebaseerde leksikografie met die rol van woordeboeke as sosiale werktuie te versoen.** In baie lande van die wêreld het perspektiewe op gendergelykheid en rassisme in onlangse dekades verander.[1] Een uitvloeisel hiervan is dat meer aandag geskenk word aan spore van androsentriese en rassistiese taal in die gemeenskap. Dit beïnvloed ook woordeboeke. In die leksikografie is daar besprekings oor of en tot watter mate sosiale asimmetrieë in woordeboeke beskryf word en of dit nog aanvaarbaar is. Die kwessie rondom die aard van beskrywing speel 'n belangrike rol in hierdie bespreking. Moet die woordeboek ook seksistiese gebruike opneem indien dit dikwels in taalgebruik, bv. in die korpusdata waarop die woordeboek gebaseer is, gevind word? Hoe is dit op sy beurt weer versoenbaar met die normatiewe krag van woordeboeke? Dra woordeboeke by tot die verewiging van gendersterotipes deur hulle onder die vaandel van beskrywende beginsels op te neem? En watter rolle moet leksikograwe in hierdie proses speel? In dié artikel word op grond van individuele leksikografiese voorbeelde en huidige besprekings in die leksikografiese en openbare gemeenskap aan hierdie vraagstukke aandag geskenk.

**Sleutelwoorde:** GENDER EN TAAL, GENDERSTEREOTIPES, KOLLOKASIES, KORPUS-GEBASEERDE LEKSIKOGRAFIE, WOORDEBOEKE AS SOSIALE WERKTUIE

> *Culture does not make people. People make culture.*
> Chimamanda Ngozi Adichie

## 1.    Introduction

In a recent article, Rufus Gouws classifies dictionaries not only as utility instruments commenting on linguistic issues, but also as "agents": "As an agent in social impact, it reflects the culture of the speech community but often also the attitudes and assessment of the lexicographer and the broader speech community towards their world and the prevailing ideologies and different ways of thinking" (Gouws 2022: 40). This position is not new. Wiegand, for example, had already pointed out that so-called cultural dictionaries offer possibilities of identification through their descriptions: "It is also true that, especially in the field of so-called cultural words, in the field of political-social lexis, as well as in the case of moral-ethical terms, general monolingual dictionaries always make offers of identification." (Wiegand 1995: 212).[2] Bergenholtz also stated that it is not only politically loaded vocabulary (such as *democracy* or *freedom*) that requires a political positioning of the lexicographer: "Lexicography without value assessments and without social commitment is an illusion and, furthermore, not possible. … traces of political and cultural-political attitudes are found not only in meaning paraphrases, but also in the external and internal selection process" (Bergenholtz 2001: 500).[3] Bergenholtz and Gouws even go so far as to say that "every single lexicographical decision has a language-political relevance and therefore, in the end, a political dimension" (Bergenholtz and Gouws 2006: 14).

This provides both opportunities and risks. One of the opportunities is that dictionaries may contribute to social transformation, as Gouws outlined for the South African dictionary landscape after the end of apartheid. He emphasised that the dictionaries to be compiled promote communication between the newly acknowledged national languages (cf. Gouws 2001: 522; for a different context cf. also Wiegand 1995). On the other hand, there is a risk that dictionaries may contribute to exclusion, for example, by perpetuating racist attributions or gender stereotypes. In this sense, lexicographers have a high degree of responsibility: "This puts an enormous responsibility on the shoulders of lexicographers to include a representative selection of items from a given language and to treat these items in an objective, responsible and unbiased way" (Gouws 2022: 40).

In addition to racism, the issue of gender constructions has received much more attention in recent years. Therefore, greater attention has been devoted to traces of an androcentric society in language. This also affects dictionaries. There are lexicographical discussions about whether or to what extent social asymmetries should be inscribed in dictionaries and if this is still acceptable, e.g. if

listing a lexeme such as *bitch* as a synonym for *woman* is still a justifiable lexicographic choice. The issue of descriptivity plays an important role in this discussion. One of the questions arising from descriptive considerations is whether dictionaries have to portray everything the corpus basis suggests, even if it goes against moral considerations.

The construction of large digital corpora, the potential they offer, and the partial automation of lexicographic processes were central topics in lexicography and dictionary research in the last decades. Corpus-based lexicography is a great achievement when the aim is to base dictionary descriptions on actual language use. However, the question of how different sources differ and how to deal with more problematic aspects of language use has been somewhat forgotten in the light of these promising possibilities. For example, if sexist usage is common in language use, i.e. in the corpus data on which the dictionary is based, does the dictionary also need to show it? How is this, in turn, compatible with the function of dictionaries as instruments of inclusion and as social agents? Within this context, this paper aims to highlight the debate about gender stereotypes in dictionaries (section 2), show how the lexicographic database can introduce biases into dictionaries with a case study on collocations for *man* and *woman* in German (section 3), and what can be learned from this, using the example of primary school dictionaries in South Africa (section 4).

## 2.      Gender stereotypes in dictionaries[4]

Personal designations such as *man* and *woman* can be understood as 'cultural words' in a broader sense (cf. Nied Curcio 2020: 186): cultural changes in society are also reflected in discussions about appropriate language, e.g. for women and men or for non-binary people. For example, in the wake of the #MeToo movement, there are debates about whether and how everyday sexism is reflected in language. These debates also have an impact on dictionaries. One example is the petition "Have you ever googled 'woman'?" which Maria Beatrice Giovanardi started in 2019. There she primarily complained about the description of women in various dictionaries, including lexicographic works by Oxford University Press, e.g. that *filly*, *biddy* or *bitch* are listed as synonyms for *woman*:

> The first search involved googling 'woman synonyms' and boom — an explosion of rampant sexism. I thought to myself, 'What would my young niece think of herself if she read this?' […] Should data about how language is used control how women are defined? Or should we take a step back and, as humans, promote gender equality through the definitions of women that we choose to accept? […] We talked about how the dictionary is the most basic foundation of language and how it influences conversations. Isn't it dangerous for women to maintain these definitions — of women as irritants, sex objects and subordinates to men? (Giovanardi 2020)

Oxford University Press responded to these questions saying that their dictionaries tried to reflect the language, not to "dictate" language use (cf. Flood 2019).

First, however, some brief remarks on gender roles in dictionaries need to be made.

Dictionaries in general are often a reflection of their time, i.e. the way that they describe the meaning of certain lexical units must always be seen in their respective historical contexts. They are linguistic resources that reflect and reproduce gender stereotypes (Nübling 2009: 594; cf. also Bergenholtz and Gouws 2006: 15). Consider the following example phrases taken from the entries on *man*, *woman*, *girl* and *boy* in the up-to-date online version of the *Cambridge Dictionary*,[5] reproducing stereotypical gender concepts (on stereotypes in general cf. McGarty, Yzerbyt and Spears 2002):

— "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."
— "She's a really nice woman."
— "Who was that beautiful girl I saw you with last night?"
— "Both girls compete for their father's attention."

Such entries concerning gender can be found in many dictionaries and have been discussed for a long time. Very pointedly and amusingly, Luise Pusch has shown how gender stereotypes that were prevalent at that time can be read in example phrases of the German *Duden Bedeutungswörterbuch* from 1970:[6] The man, i.e. "he", "shows an acrobatic mastery of his body", "his soul is able to encompass the universe" and "he had a great impact". "She," on the other hand, "is always neatly dressed," "took the baby out daily," "awaits his return with great anxiety," and "she looked up to him as to a god."[7] Pusch summarizes: "In the preface, the editors write that the 'basic vocabulary of German in its basic meanings' is to be presented. They succeed in much more: they convey a deep, unforgettable insight into the soul of German, into its basic treasure of feelings and thoughts" (Pusch 1984: 144; for more details on various German dictionaries cf. Nübling 2009). This illustrates that dictionaries are often a mirror of their time and thus also one of the important "platforms for productions of gender" (Nübling 2009: 594 [my translation]). Similarly, in their analysis of a contemporary Chinese dictionary, Hu, Xu and Hao (2019) point out that women "are often constructed in peripheral and domestic roles, as daughter, mother or grandmother" (Hu, Xu and Hao 2019: 28). In contrast, men "are described as strong in physical strength, versatile in skills and noble in their actions. In other words, men are represented as valuable, active social members" (Hu, Xu and Hao 2019: 28) These stereotypical representations of men and women are therefore not only to be found in dictionaries of the past, but also in current lexicographical works (cf. e.g. Fuertes-Olivera and Tarp 2022).

Dictionary editors' conscious positioning is particularly relevant because dictionaries can be understood as normative instances, even if they are primarily intended to be descriptive (Ripfel 1989; Sinclair 1992; Barnickel 1999; Hidalgo Tenorio 2000; Kotthoff and Nübling 2018). Therefore, lexicographers

have a special responsibility not to contribute to perpetuating gender stereo-types by showcasing them under the banner of descriptive principles. After Pusch's essay cited above, attempts were made by the Duden editors to improve the dictionary in many ways, e.g. by avoiding unnecessarily stereo-typical example phrases and by systematically including female occupational designations (Kunkel 2004; cf. also Westveer et al. 2018; Nübling 2009: 595)

The representation of gender in dictionaries is a matter of both language use and lexicographic-moral responsibility. Regarding language use, however, it is necessary to look closely at the data basis that is considered to represent this use. As in the field of machine learning (ML) and artificial intelligence (AI), lexicographers should discuss whether the selection of a particular database intro-duces additional, probably unintended, biases into their analyses (and then into the dictionaries). To illustrate the problem of gender bias in dictionaries, I will approach the basic problem from two very different perspectives. First, a summary of a comparative corpus study of German will show what influence the corpus base can have on the representation of gender roles in a dictionary. Secondly, I would like to illustrate more anecdotally how gender stereotypes are (unconsciously) reproduced on thematic pages in two concrete dictionaries for children from the South African dictionary landscape, and how these repre-sentations can be made more inclusive through greater awareness of gender stereotypes (and the desire to avoid them).

## 3.     The database as a source of bias

Biases that are introduced by the database are a well-known problem in com-puter-linguistic research. For example, Caliskan et al. summarise the results of a study on semantics derived automatically from language corpora like this: "Our results indicate that text corpora contain recoverable and accurate imprints of our historic biases, whether morally neutral as toward insects or flowers, problematic as toward race or gender, or even simply veridical, reflecting the status quo distribution of gender with respect to careers or first names." (Caliskan, Bryson and Narayanan 2017: 184). This is in line with a growing number of studies conducted in the field of ML or AI, stating that these technological sys-tems "often influence, reflect, and reinforce gender stereotypes" (Singh et al. 2020: 1281; cf. also Chen et al. 2021; Criado Perez 2020).

To investigate whether this problem also applies to corpus-based diction-aries, i.e. whether different corpus samples lead to different meaning paraphrases, we examined collocations for *Mann* (man) and *Frau* (woman) in different text sources (cf. Müller-Spitzer and Rüdiger 2022). The analyses presented in the fol-lowing are based on three corpora for German, compiled from different source materials. The first is a corpus of 'Fiction Books' which is based on various works of fiction (20th and 21st century). The second, 'Magazines', is a corpus con-taining various periodicals on food, lifestyle, psychology, nature, education, etc. The third corpus consists of 'Newspapers' (for details cf. Nied Curcio 2022: 134).

For the comparison, co-occurrences were filtered to include only tokens that were annotated at least once by the TreeTagger[8] with the *adjective* POS tag. Common co-occurrences were filtered by the 100 most significant entries (based on Poisson distribution; cf. Heyer, Quasthoff and Wittig 2006: 134). In the following figures (Figures 1–3), we see all adjectives co-occurring with *Mann* and *Frau* in these three corpora. Adjectives are especially interesting to examine gender stereotypes as they are used, among other things, to describe entities. Thus, they would be included in a collocation set answering a question such as "What is a *woman* or a *man* like?" Our three corpora show different tendencies in this regard.
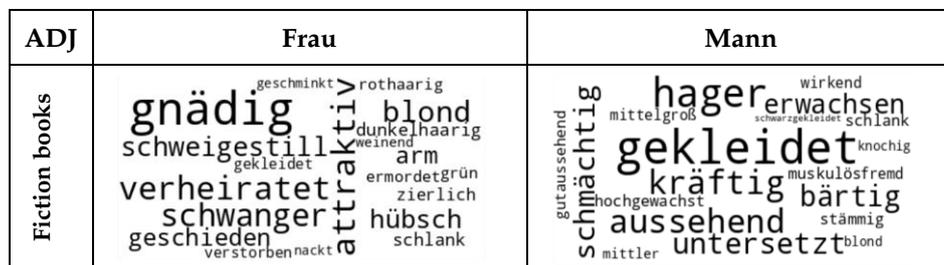


**Figure 1:**    Adjective co-occurrences for *Frau* and *Mann* in fiction books (the font size indicates significance based on Poisson distribution)

In works of fiction women are mainly described regarding their appearance, e.g. *blond*, *pretty* or *attractive* (*blond, hübsch, attraktiv;* n=10/20), but also in terms of their marital status (*married* or *divorced – verheiratet, geschieden*), or *pregnant* (*schwanger*). The adjective *schweigestill* seems to be a tagging error (it is not a German adjective) and *gnädig* points to the quasi-lexicalized, outdated address comparable to *milady* or *madam/ma'am* (*gnädige Frau*). For *man*, in contrast, adjectives describing appearance clearly dominate, e.g. *gaunt, stout, stocky, bearded* or *lanky* (*hager, kräftig, untersetzt, bärtig, schmächtig*).



**Figure 2:**    Adjective co-occurrences for Frau and *Mann* in magazines (the font size indicates significance based on Poisson distribution)

In magazine texts only three collocators refer to appearance (*attraktiv, kurvig, zier-lich – attractive, curvy, petite*), while *pregnant* and *childless* (*schwanger, kinderlos*) refer to the role of (not) being a mother. Adjectives referring to marital status are not part of the most significant collocates. Interestingly, women are rather characterized as *self-confident, employed, independent, strong* or *emancipated* (*selbst-bewusst, berufstätig, unabhängig, stark, emanzipiert*). Therefore, the magazine corpus is the only dataset in which women are more characterized by other adjectives than those referring to appearance, social roles or violence. Men, on the other hand, are described as *attractive, married, bearded, naked, gay* or *\*-looking (attrak-tiv, verheiratet, bärtig, nackt, schwul, aussehend*). Surprisingly, a considerable num-ber of terms relate to appearance, social role or sexual orientation.

| ADJ | Frau | Mann |
|---|---|---|
| newspapers |  |  |

**Figure 3:**   Adjective co-occurrences for *Frau* and *Mann* in newspapers (the font size indicates significance based on Poisson distribution)

In the newspaper texts, women are described in terms of marital roles (*allein-stehend, verheiratet, geschieden – single, married, divorced*) and as *pregnant* (*schwanger, hochschwanger*). We also find collocators referring to work (*berufstätig, erwerb-stätig*) and to emancipation (*emanzipiert, gleichberechtigt*). What is striking, how-ever, is the diverse and dominant vocabulary referring to physical violence: *raped* (*vergewaltigt*), probably both in its adjectival and its participial/passive use, is the most significant collocate after *pregnant* and *employed*. *Sexual* (*sexuell*) and *affected* (*betroffen*) could point to contexts such as *affected by sexual violence*. Other collocators indicating violence are *misshandelt* and *missbraucht* (mis-treated, misused). Also for men, violent acts or potentially dangerous physical/ mental states are a predominant topic in the newspaper corpus: *armed, masked, alcoholised, drunk, previously convicted (bewaffnet, maskiert, alkoholisiert, angetrunken, vorbestraft)* are significant collocators of *Mann*.

The examples show how strongly 'language use' depends on the empirical basis. The linguistic-thematic embedding of the words *Frau* and *Mann* in the various text groups show several overlaps (e.g. appearance and, for women, marital status), but also significant differences (fiction: dominance of appear-ance; magazines: topics of emancipation arise; newspapers: violent acts domi-nate for both lexemes). It is therefore too simplistic to state that meaning para-phrases and examples dictionaries should be based on 'language use', as this

can mean very different things in different contexts. At least in academic lexicography, scrutiny is necessary to determine which source is responsible for which descriptions in the dictionary. This finding also highlights the importance of good metadata for linguistic corpora. For example, when corpora used in lexicography are crawled from the web and lack metadata such as text type, source or authorship, it is even harder to adequately reflect or classify the different contexts in which texts were created.

When lexicographers aim at designing dictionaries as instruments of inclusion, it is certainly a sensible approach to avoid over-emphasising gender in entries concerning men and women, while still basing the descriptions on actual language use. For example, linguistic contexts that point to violence and are taken from newspaper corpora must be classified as such — it is not 'general language use', but that of a specific genre or jargon. It is another question whether such examples should be part of a general language dictionary at all, which will not be discussed in this paper. Contrastingly, the various sets of collocations from magazines show that men and women can be portrayed in totally different ways (e.g. as self-confident and emancipated). These collocations are certainly preferable when balanced portrayal is an aim in dictionary-making. Kaplan also argues for a carefully curated selection: "Unlike those offered by the regular English dictionaries, bias-free and inclusive paraphrases of meaning reflect real usage and the real consequences of bias and exclusion. Such paraphrases of meaning and illustrations are necessary in order for dictionaries and lexicographical resources as a whole to provide users with information that is trustworthy, culturally aware, and socially responsible" (Kaplan 2020: 218).

## 4.    Gender stereotypes in the Longman school dictionaries

As far as I know, the South African dictionary landscape is unique in that, after apartheid, nine indigenous languages were recognised as national languages alongside the colonial languages English and Afrikaans, and dictionaries were developed specifically for these 'new' national languages, initiated by PanSALB (Pan South African Language Board). PanSALB's goal is to "accelerate the production of dictionaries" in order to create "the conditions for the development of and the equal use of all official languages".[9] Another special feature is that leading academics are involved in the development of very different types of dictionaries (cf. e.g. Gouws 1993), for example, the Longman dictionaries for all official languages of South Africa, which are published with Rufus H. Gouws and Danie Prinsloo as series consultants. Such dictionaries or textbooks can play an important role in (un)learning gender stereotypes: "Textbooks are used by teachers as a core means of teaching in 70-95% of classroom time. Gender-sensitive books can encourage children to discuss gender stereotypes and help promote equitable behaviour" (Benavot 2016).

The last two decades have seen many societal changes regarding gender constructions. Not only are women's rights being discussed more often and with

more positive outcomes — the binary gender order itself is gradually being deconstructed. Therefore, it is worth looking at teaching material and reference works with a new critical eye to see if there is room for improvement. The dictionaries I examine are the 2010 *Longman Foundation Phase Bilingual Sepedi–English Dictionary* (Mabule 2010) and the 2010 *Longman Foundation Phase Bilingual IsiXhosa–English Dictionary* (Dlali 2010). These dictionaries contain word lists and a picture dictionary with themes that are frequently used in the Foundation Phase. They include scenes from children's everyday lives and the corresponding English and Sepedi/IsiXhosa terms. The second part of the paper therefore focuses on visual representations of gender stereotypes, as opposed to the verbal expressions examined in the preceding section.
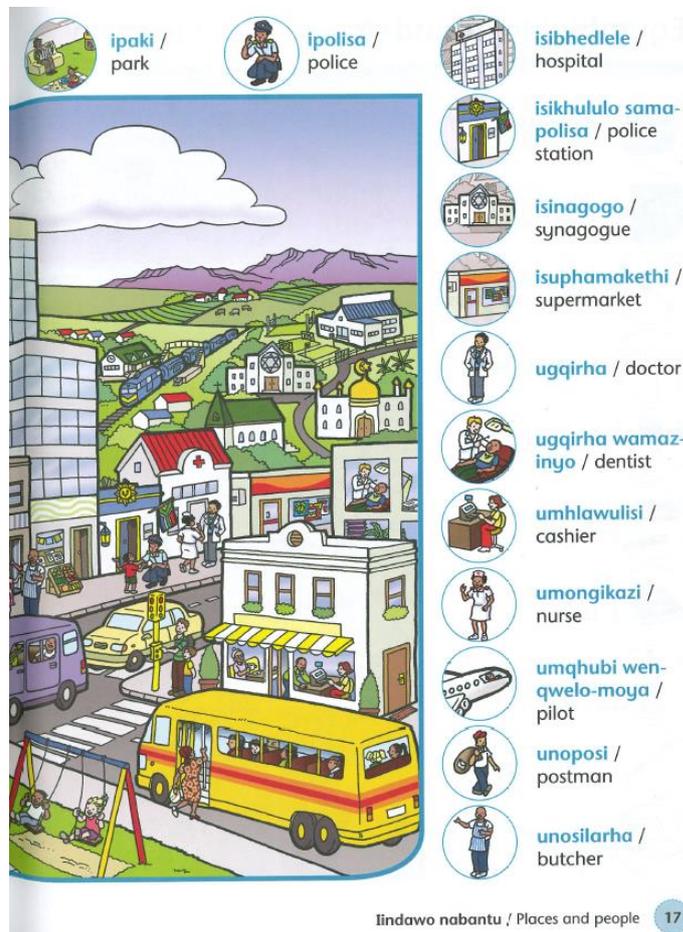


**Figure 4:** Theme page "Places and people" in 2010 *Longman Foundation Phase Bilingual IsiXhosa–English Dictionary* (Dlali 2010: 17)

On the theme page of "Places and people", we see various people in different professions (cf. Figure 4). Professional roles are often stereotypically associated with one gender. It is therefore worth taking a critical look at them: We see a male *police officer*, a male *doctor*, a male *dentist*, a female *cashier*, a female *nurse*, a male *postman* and a male *butcher*. In total, we have a 5:2 ratio of men to women. Besides that, the higher status occupations are all portrayed by men (*doctor, dentist*). Jobs performed traditionally by women are respectively illustrated with female personas (*cashier, nurse*) This is a problem that is also known from AI systems: "Especially, in widely used language translation systems ML/AI models assign pronouns to professions confirming the gender stereotypes; for example, the models automatically assign he/him pronouns to professions such as doctors and pilots, whereas it assigns she/her pronouns to nurses and flight attendants" (Shrestha and Das 2022: 1). A study of Malaysian, Indonesian, Pakistani and Bangladeshi school textbooks describes the same asymmetry: "The overall findings show that professions attached to female characters are traditional, and lower in prestige and income" (Islam and Asadullah 2018: 14). On the other hand, it is also important that these theme pages must relate to the children's world, otherwise they might regard the dictionary as a 'bad' reference book. This is the well-known problem of usage and authority that Sinclair has already described: "This brings up the question of usage and authority. These concepts must support each other or no-one will respect either of them. If their close relationship breaks down, and authority is not backed up by usage, then no-one will respect it. … Similarly, no-one will respect usage if it is merely an unedited record of what people say and write. … Any successful record of a language such as a dictionary is itself a contribution to authority" (Sinclair 1992). Nevertheless, more inclusive portrayals can be implemented without going against either authority or usage. While certain occupations might still be gender-biased, others certainly have started to see more mixed groups (e.g. doctors are predominantly female in younger age groups (Tiwari et al. 2021: 7). In this way, asymmetries can be reduced and gender role perceptions gradually altered.

Another area where stereotypes are encoded is the gendered assignment of verbs or actions. On the topic page "Things that we do" (Figure 5), we see children involved in various actions: girls *read*ing, *danc*ing, *writ*ing, *push*ing (someone on the swing) and *swim*ming; boys *throw*ing, *sit*ting, *giv*ing, *catch*ing, *kick*ing, *play*ing, *laugh*ing and *wav*ing. As these are all gender-unrelated activities, a balanced portrayal of girls and boys would be desirable (boys are assigned 8 activities, girls only 5). Moreover, some children could easily be depicted without any clear gender attribution.

Among the adults it is mainly the helping activities that are attributed to women: *Carry*ing baskets and *wash*ing dishes are done by women, while the man is *stand*ing in the playground with his hands in his pockets. From today's point of view, the illustrations should be less gender-stereotypical in order to give children a freer idea of role assignments (without being unrealistic). Mohd Faeiz

Ikram bin Mohd Jasmani et al. (2011) argue along these lines that overly stereo-typical role allocation in textbooks could have very negative impacts; accord-ingly, I suggest that more conscious and diversified representations could have a positive impact. Mohd Faeiz Ikram bin Mohd Jasmani et al. (2011: 71) point out:

> portrayals of stereotypical ideals and ideologies about men and women in the English language textbooks can have lasting consequences. Children and adoles-cents usually develop their ideas about the world at an early stage and this will last well into their adult lives. The textbooks, by depicting gender bias ideolo-gies, seem to suggest that women have a limited or restricted role to play in a male-dominant society and women, by accepting such a perspective, perceive this view as normal.



**Figure 5:**    Theme page "Things that we do" in the 2010 *Longman Foundation Phase Bilingual Sepedi–English Dictionary* (Mabule 2010: 20-21)

## 5.    Concluding remarks

Lexicography is a discipline that (if it wants to produce user-friendly diction-aries) stands at the centre of society. This means that societal values are reflected in dictionaries and that lexicographers have to be aware of changes and adapt their work accordingly, e.g. when it comes to paraphrases of meaning. This is something that Rufus Gouws pointed out early and repeatedly: "In the treat-

ment of any linguistic expression lexicographers should not only look at the linguistic features of that form, but they should also look beyond that at the cultural values it conveys. Taking due cognizance of the culture also implies being aware of specific culture-bound lexical items and suggesting new ways to present and treat them in an optimal way" (Gouws 2020: 4).

In the context of descriptive, corpus-based lexicography, this aspect has been somewhat neglected, as exemplified by OUP's statement above saying that their dictionaries try to reflect and not to dictate language use (Flood 2019). And the results of the empirical analyses presented here can be seen as indicative of gender bias in dictionaries, although this would require much more extensive empirical investigation. In my view, it is not enough to simply observe language use and record it in the dictionary. Dictionaries are made by humans, and if dictionaries are to be different from search engines or other ML or AI tools, they should show a carefully curated selection of language use (cf. Nied Curcio 2020: 200; for the field of ML and AI cf. Caliskan, Bryson and Narayanan 2017; Bolukbasi et al. 2016). As Bergenholtz emphasized, such processes of selection from language use or comments on it have little to do with language policy. Rather, it is a conscious classification that makes dictionaries more helpful (cf. Bergenholtz 2001: 517; Bergenholtz and Gouws 2006: 39-40) and, hopefully, less discriminating.

Critical analyses of already published dictionaries, as shown briefly in section 4, are not ends in themselves. They enable us to learn from the past and to avoid making the same 'mistakes' again in the future. Rufus Gouws once expressed this very well regarding digital lexicography, so the last word here is his:

> What should be learned from the past, and this applies to both printed and electronic dictionaries, is to conscientiously avoid similar traps and mistakes, especially in cases where what are now seen as mistakes were then regarded as the proper way of doing things. ... In these new endeavours, we as lexicographers are still bound to make mistakes in the future, but we have to restrict ourselves to making only new mistakes. (Gouws 2011: 18)

## Endnotes

1.    I thank Samira Ochs for insightful comments on this paper and both reviewers for helpful remarks.
2.    My translation of original: "Weiterhin gilt: Besonders im Bereich der sog. kulturellen Wörter, im Bereich der politisch-sozialen Lexik sowie bei den moralisch-ethischen Bezeichnungen machen die allgemeinen einsprachigen Wörterbücher stets auch Identifikationsangebote."
3.    My translation of original: "Lexikographie ohne Werturteile und ohne gesellschaftliches Engagement ist ein Unding und darüber hinaus gar nicht möglich. […] Dies gilt jedoch für alle Wörterbücher, nicht nur bei Bedeutungsangaben, sondern auch bei der äußeren und inneren Selektion finden sich Spuren politischer und kulturpolitischer Haltungen."
4.    Parts of the following study (section 2 and 3) are already published in (Müller-Spitzer and Rüdiger 2022; Müller-Spitzer and Lobin 2022).
5.    https://dictionary.cambridge.org/de/.

6.  *Duden Bedeutungswörterbuch*, Mannheim 1970.

7.  Own translation, original: "Der Mann, also 'er', 'zeigt eine akrobatische Beherrschung seines Körpers', 'seine Seele vermag das All zu umfassen' und 'große Wirkung ging von ihm aus'. 'Sie' dagegen 'ist immer adrett gekleidet', 'hat das Baby täglich ausgefahren', 'erwartet mit großer Angst seine Rückkehr' und 'sie sah zu ihm auf wie zu einem Gott'."

8.  Helmut Schmid (1995): Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop, Dublin, Ireland*: 47-50; https://www.cis.uni-muenchen. de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf.

9.  https://pansalb.org/.

# References

**Barnickel, Klaus-Dieter.** 1999. Political Correctness in Learners' Dictionaries. Herbst, Thomas and Kerstin Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?):* 161-174. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110947021.161

**Benavot, Aaron.** 2016 (updated 2022). Gender Bias is Rife in Textbooks. *World Education Blog,* 8 March 2016 (updated version: 18 April 2022 by Aaron Benavot and and Catherine Jere). https://world-education-blog.org/2016/03/08/gender-bias-is-rife-in-textbooks/ (24 January, 2023)

**Bergenholtz, Henning.** 2001. Proskription, oder: So kann man dem Wörterbuchbenutzer bei Textpro- duktionsschwierigkeiten am ehesten helfen. Lehr, Andrea, Matthias Kammerer, Klaus-Peter Konerding, Angelika Storrer, Caja Thimm and Werner Wolski (Eds.). 2001. *Sprache im Alltag: Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*: 499-519. Berlin/New York: De Gruyter. https://doi.org/10.1515/9783110880380

**Bergenholtz, H. and Rufus H. Gouws.** 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13-45. https://doi.org/10.4314/lex.v16i1.51486

**Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama and Adam Kalai.** 2016. *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. arXiv: 1607.06520. https://doi.org/10.48550/arXiv.1607.06520

**Caliskan, Aylin, Joanna J. Bryson and Arvind Narayanan.** 2017. Semantics Derived Automatically from Language Corpora Contain Human-like Biases. *Science* 356(6334): 183-186. https://doi.org/10.1126/science.aal4230

**Chen, Yan, Christopher Mahoney, Isabella Grasso, Esma Wali, Abigail Matthews, Thomas Middle- ton, Mariama Njie and Jeanna Matthews.** 2021. Gender Bias and Under-representation in Natural Language Processing across Human Languages. *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, May 19–21, 2021, Virtual Event USA:* 24-34. New York: ACM (Association for Computing Machinery). https://doi.org/10.1145/3461702.3462530

**Criado Perez, Caroline.** 2020. *Invisible Women. Exposing Data Bias in a World Designed for Men.* Lon- don: Vintage.

**Dlali, M.** 2010. *Longman Foundation Phase Bilingual IsiXhosa/English Dictionary*. Cape Town: Maskew Miller Longman. https://www.graffitiboeke.co.za/ (7 February, 2023)

**Flood, Alison.** 2019. Thousands Demand Oxford Dictionaries "Eliminate Sexist Definitions". *The Guardian*, 17 Sept. 2019.
https://www.theguardian.com/books/2019/sep/17/thousands-demand-oxford-dictionaries-eliminate-sexist-definitions (22 March, 2022)

**Fuertes-Olivera, Pedro A. and Sven Tarp.** 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132.
https://doi.org/10.5788/32-2-1699

**Giovanardi, Maria Beatrice.** 2020. Open Letter calling on @OxUniPress to change their entry for the word "woman" #SexistDictionary. *Change.org*.
https://www.change.org/p/change-oxford-dictionary-s-sexist-definition-of-woman/u/25841171 (17 March, 2022)

**Gouws, Rufus H.** 1993. Afrikaans Learner's Dictionaries for a Multilingual South Africa. *Lexikos* 3: 29-48.
https://doi.org/10.5788/3-1-1099

**Gouws, Rufus H.** 2001. Der Einfluß der neueren Wörterbuchforschung auf einen neuen lexikographischen Gesamtprozeß und den lexiko graphischen Herstellungsprozeß. Lehr, Andrea, Matthias Kammerer, Klaus-Peter Konerding, Angelika Storrer, Caja Thimm and Werner Wolski (Eds.). 2001. *Sprache im Alltag: Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*: 521-531. Berlin/New York: De Gruyter.
https://doi.org/10.1515/9783110880380

**Gouws, Rufus H.** 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography. The Internet, Digital Initiatives and Lexicography*: 17-29. London/New York: Continuum.

**Gouws, Rufus H.** 2020. Metalexicography, Dictionaries and Culture. *Lexicographica* 36(2020): 3-9.
https://doi.org/10.1515/lex-2020-0001

**Gouws, Rufus 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

**Heyer, Gerhard, Uwe Quasthoff and Thomas Wittig.** 2006. *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. IT Lernen. Herdecke: W3L.

**Hidalgo Tenorio, Encarnación.** 2000. Gender, Sex and Stereotyping in the *Collins COBUILD English Language Dictionary*. *Australian Journal of Linguistics* 20(2): 211-230.
https://doi.org/10.1080/07268600020006076

**Hu, Huilian, Hai Xu and Junjie Hao.** 2019. An SFL Approach to Gender Ideology in the Sentence Examples in the Contemporary Chinese Dictionary. *Lingua* 220: 17-30.
https://doi.org/10.1016/j.lingua.2018.12.004

**Islam, Kazi Md Mukitul and M. Niaz Asadullah.** 2018. Gender Stereotypes and Education: A Comparative Content Analysis of Malaysian, Indonesian, Pakistani and Bangladeshi School Textbooks. *PLOS ONE* (Public Library of Science) 13(1): e0190807.
https://doi.org/10.1371/journal.pone.0190807

**Kaplan, Steven M.** 2020. Boys Will Be Boys: An Example of Biased and Exclusive Usage. *Lexicographica* 36: 205-223.
https://doi.org/10.1515/lex-2020-0011

**Kotthoff, Helga and Damaris Nübling.** 2018. *Genderlinguistik: Eine Einführung in Sprache, Gespräch und Geschlecht*. Narr Studienbücher. Tübingen: Narr Francke Attempto.

**Kunkel, Kathrin.** 2004. Die Frauen und der Duden — der Duden und die Frauen. Eichhoff-Cyrus, Karin M. (Ed.). 2004. *Adam, Eva und die Sprache : Beiträge zur Geschlechterforschung*: 308-315. Mannheim: Dudenverlag.

**Mabule, M.** 2010. *Longman Foundation Phase Bilingual Sepedi/English Dictionary*. Cape Town: Maskew Miller Longman. (7 February, 2023)

**McGarty, Craig, Vincent Y. Yzerbyt and Russell Spears.** 2002. Social, Cultural and Cognitive Factors in Stereotype Formation. McGarty, Craig, Vincent Y. Yzerbyt and Russell Spears (Eds.). 2002. *Stereotypes as Explanations: The Formation of Meaningful Beliefs about Social Groups*: 1-15. Cambridge: Cambridge University Press.
https://doi.org/10.1017/CBO9780511489877.002

**Mohd Faeiz Ikram bin Mohd Jasmani, Mohamad Subakir Mohd Yasin, Bahiyah Abdul Hamid, Yuen Chee Keong, Zarina Othman and Azhar Jaludin.** 2011. Verbs and Gender: The Hidden Agenda of a Multicultural Society. *3L: Language, Linguistics and Literature: The Southeast Asian Journal of English Language Studies* 17 (Special issue): 61-73.

**Müller-Spitzer, Carolin and Henning Lobin.** 2022. Leben, lieben, leiden: Geschlechtersterotype in Wörterbüchern, Einfluss der Korpusgrundlage und Abbild der sprachlichen ‚Wirklichkeit'. Diewald, Gabriele and Damaris Nübling. 2022. *Genus — Sexus — Gender*: 33-64. Berlin/Boston: De Gruyter.
https://doi.org/10.1515/9783110746396-002

**Müller-Spitzer, Carolin and Jan Oliver 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, Annette, Stefan Engelberg, Christine Möhrs and Petra Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congres*: 129-141. Mannheim: IDS Verlag.
https://euralex.org/publications/the-influence-of-the-corpus-on-the-representation-of-gender-stereotypes-in-the-dictionary-a-case-study-of-corpus-based-dictionaries-of-german/ (21 December, 2022)

**Nied Curcio, Martina.** 2020. Kulturell geprägte Wörter zwischen sprachlicher Äquivalenz und kultureller Kompetenz. Am Beispiel deutsch–italienischer Wörterbücher. *Lexicographica* 36(2020): 181-204.
https://doi.org/10.1515/lex-2020-0010

**Nied Curcio, Martina.** 2022. Dictionaries, Foreign Language Learners and Teachers. New Challenges in the Digital Era. Klosa-Kückelhaus, Annette, Stefan Engelberg, Christine Möhrs and Petra Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congres*: 71-84. Mannheim: IDS Verlag.
https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202022/EURALEX2022_Pr_p71-84_Nied-Curcio.pdf (21 December, 2022)

**Nübling, Damaris.** 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

**Ripfel, M.** 1989. Die normative Wirkung deskriptiver Wörterbücher. Hausmann, Franz Josef, Oskar Reichmann, H.E. Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Ein Internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dic-*

*tionnaires. Encyclopédie internationale de lexicographie*. Volume 5.1: 189-207. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin/New York: De Gruyter.

**Shrestha, Sunny and Sanchari Das.** 2022. Exploring Gender Biases in ML and AI Academic Research through Systematic Literature Review. *Frontiers in Artificial Intelligence* 5.
https://www.frontiersin.org/articles/10.3389/frai.2022.976838 (24 January, 2023)

**Sinclair, John.** 1992. Introduction. *Collins Cobuild English Language Dictionary*, xv-xxi. London: HarperCollins.

**Singh, Vivek K., Mary Chayko, Raj Inamdar and Diana Floegel.** 2020. Female Librarians and Male Computer Programmers? Gender Bias in Occupational Images on Digital Media Platforms. *Journal of the Association for Information Science and Technology* 71(11): 1281-1294.
https://doi.org/10.1002/asi.24335

**Tiwari, Ritika, Angelique Wildschut-February, Lungiswa Nkonki, René English, Innocent Karangwa and Usuf Chikte.** 2021. Reflecting on the Current Scenario and Forecasting the Future Demand for Medical Doctors in South Africa up to 2030: Towards Equal Representation of Women. *Human Resources for Health* 19(1).
https://doi.org/10.1186/s12960-021-00567-2

**Westveer, Thom, Petra Sleeman and Enoch O Aboh.** 2018. Discriminating Dictionaries? Feminine Forms of Profession Nouns in Dictionaries of French and German. *International Journal of Lexicography* 31(4): 371-393.
https://doi.org/10.1093/ijl/ecy013

**Wiegand, Herbert Ernst.** 1995. Der kulturelle Beitrag der Lexikographie zur Umgestaltung Osteuropas. *Lexicographica* 11: 210-218.

# Lexicography, Culture and Mediation. Or Why a Good Lexicographer Must Also Be a Good Cultural Mediator

Martina Nied Curcio, *Università degli Studi Roma Tre, Italy (martina.nied@uniroma3.it)*

**Abstract:** Culture-bound items are often omitted in lexicographic resources, and when they are present, they are rarely described in an appropriate way, especially in bilingual dictionaries: the listed equivalents often do not reflect the meaning of the cultural word precisely, usage examples may be missing and above all — especially for a non-native speaker — important information on the cultural level is omitted (Nied Curcio 2020). Communication problems and misunderstandings which are triggered not by divergences on the linguistic level, but on the cultural level, can arise between speakers of different languages and cultures. Specific cultural knowledge and an advanced level of intercultural competence are required.

Mediation of texts and concepts and using mediation strategies can provide significant input developing the lexicographer's intercultural competence, which is essential for an adequate lexicographic description of cultural aspects. Mediators often resort to their plurilinguistic and pluricultural repertoire and use mediation strategies. The same skills can also be harnessed by lexicographers for adequate and successful descriptions of culture-bound items in lexicographic resources. In this article the focus is on the concept of mediation; the *Companion Volume* of the *Common European Framework of Reference for Languages* (2018) shows parallels in the requirements of skills between mediators and lexicographers, and advocates the use of mediation strategies by lexicographers to ensure that cultural information is adequately represented in the dictionary.

**Keywords:** BILINGUAL DICTIONARIES, CULTURAL ITEMS, TRANSLATION, INTERCULTURAL COMPETENCE, MEDIATION, MEDIATION STRATEGIES, PLURILINGUAL COMPETENCE, PLURICULTURAL COMPETENCE

**Opsomming: Die leksikografie, kultuur en bemiddeling. Of waarom 'n goeie leksikograaf ook 'n goeie kultuurbemiddelaar moet wees.** Kultuurgebonde items word dikwels weggelaat uit leksikografiese bronne, en wanneer hulle aanwesig is, word hulle selde, veral in tweetalige woordeboeke, op 'n toepaslike wyse verklaar: die gelyste ekwivalente reflekteer dikwels nie die betekenis van die kultuurwoord akkuraat nie, gebruiksvoorbeelde mag ontbreek, en bowenal — veral vir 'n niemoedertaalspreker — word belangrike inligting op kultuurvlak weggelaat (Nied Curcio 2020). Kommunikasieprobleme en misverstande wat, nie deur afwykings op linguistiese vlak nie, maar op kultuurvlak, veroorsaak word, kan tussen sprekers van verskillende tale en kulture ontstaan. Spesifieke kultuurkennis en 'n gevorderde vlak van interkulturele vaardigheid word vereis.

Bemiddeling van tekste en konsepte en die gebruik van bemiddelingstrategieë kan belangrike insette lewer om die leksikograaf se interkulturele vaardigheid, wat noodsaaklik is vir 'n voldoende leksikografiese beskrywing van kultuuraspekte, te ontwikkel. Bemiddelaars maak dikwels gebruik van hul veeltalige en multikulturele repertoire en gebruik bemiddelingstrategieë. Leksikograwe kan ook dieselfde vaardighede vir die voldoende en suksesvolle beskrywing van kultuurgebonde items in leksikografiese hulpbronne inspan. In hierdie artikel word die fokus op die mediasiekonsep geplaas; die *Companion Volume* van die *Common European Framework of Reference for Languages* (2018) toon ooreenkomste in die vaardigheidsvereistes tussen bemiddelaars en leksikograwe, en steun die gebruik van bemiddelingstrategieë deur leksikograwe om te verseker dat kultuurinligting in die woordeboek voldoende verteenwoordig word.

**Sleutelwoorde:** TWEETALIGE WOORDEBOEKE, KULTUURITEMS, VERTALING, INTERKULTURELE VAARDIGHEID, BEMIDDELING, BEMIDDELINGSTRATEGIEË, VEELTALIGE VAARDIGHEID, MULTIKULTURELE VAARDIGHEID

## 1.    Cultural items in translation and in lexicography

Because of the globally interconnected and networked world, we find constantly ourselves engaged in cross-cultural communication (cf. Gudykunst and Ting-Toomey 1988), whether in everyday life, when travelling, learning a foreign language or in professional interpreting and translating. Comparable situations in which communication is misunderstood and even disrupted as a result of cultural divergences are numerous, precisely because communication partners often concentrate on the linguistic level and disregard the cultural level. However, every language contains a large number of culturally determined words. These relate to everyday life (plants and animals, food, clothing, housing, work, leisure), politics, religion, geography, nature, ethnography, justice and administration, as well as social rules and values and the image that people or a nation have of themselves and others (cf. Newmark 1988). In the broadest sense languages are carriers and symbols of identity of a national/ethnic entity or culture assigned to a country, a region, a continent.[1]

In translation studies it is acknowledged that there are various terms that refer to a concept that is clear to a native speaker, but often remain unclear in their cultural connotations to a second-language speaker.

> In translation studies, these references have been variously referred to as *cultural references* (or *referents*), *cultural elements*, *culture-specific items*, *realia* or *culturemes*. All these terms intend somehow to point to a concept that can be more or less intuitively grasped, at least by translators and translation scholars (Marco 2019: 21).

Cultural items are repeatedly the subject of theoretical and methodological discussions, since there is often a lexical gap in the other language as a result of cultural incongruence[2] (Stolze 1999: 224).

> [I]f there is a lexical gap, i.e. if words or phrases are not known or when lexical equivalents do not exist in the target culture and language, such culture-specific items cause problems in translation (Persson 2015: 1).

Translators are challenged to address this lack of appropriate words in one language; they have to fill the lexical gap and — if possible — to find the appropriate lexical expression (while also devoting attention to the register and possible linguistic connotations of the variety). Alternatively, if no equivalent of the culture-bound item exists, they are challenged to describe the cultural item in an adequate and comprehensible way. It is their task to ensure the most precise reception of the culturally determined word or expression by the users.

Not only in translation studies, but also in lexicography, the question of translatability or strategies for the transfer of cultural items has been dealt with extensively, because the lexicographer has to face the same challenges when dealing with cultural items in a dictionary. Based on the theoretical studies of Molina Martínez (2006) and Luque Nadal (2009) and combining the different nuances of this concept, Sanmarco Bande (2017) provides a very useful definition of cultural items for meta-lexicography, putting the focus on contrasts, i.e. language and culture comparison,[3] thus working towards overcoming the problem. Sanmarco Bande (2017: 303) states that the most effective method of dealing with culturally specific words is not only offering the closest equivalent to the source language term, but also addressing the entire microstructure with a combination of different strategies: clarifications in brackets, numerous linguistic labels, clear and well-contextualised examples, semantic, cultural and pragmatic remarks, all reinforced by targeted typography. It is necessary to continue reinforcing the translation with other parts of the microstructure, because without the support of these tools, it is very likely that the user will misinterpret the information provided within the translation.[4]

One could assume that lexicographers have by now solved the translation problem of cultural items in bilingual dictionaries and that these dictionaries offer suitable equivalents for the users. However, this is — as already mentioned — a fallacy with regard to the cultural level. While much attention has been devoted to cultural items that are typical of a particular culture or nation and that have neither a linguistic nor a cultural equivalent in another language (German: "Kulturspezifika"),[5] less attention has been devoted to an adequate translation or lexicographic description of cultural items that are typical of several (at least two) cultures and for this reason have linguistic equivalents in the other language (German: "Kulturwort")[6] (for a categorization of culture-bound items for lexicography cf. Nied Curcio 2020). The equivalents are very often not completely equivalent; without intercultural competence, the cultural divergence is not recognisable at first glance. Words such as *breakfast* or *coffee* always have an equivalent in bilingual dictionaries, i.e. *breakfast*: ontbyt (Afrikaans), Frühstück (German), *petit déjeuner* (French), (*prima*) *colazione* (Italian), *café da manhã* (Portuguese) *kahvaltı* (Turkish), *reggeli* (Hungarian) or *coffee*: koffie (Afrikaans), *Kaffee* (German), *café* (French), *caffè* (Italian), *café* (Portuguese) *kahve* (Turkish), *kávé* (Hungarian). Some of these "Kulturwörter" can be international and quite easy to understand, but at the cultural level they are often *false friends* in terms of content, usage, habits and traditions. As can be seen in Figs. 1–4, the choice of ingre-

dients that are part of a (typical) breakfast can vary across different cultures. (Figs. 1–4)

| | | | |
|---|---|---|---|
| **Fig. 1:** *Frühstück* (German) | **Fig. 2:** *(prima) colazione* (Italian) | **Fig. 3:** *kahvaltı* (Turkish) | **Fig. 4:** *reggeli* (Hungarian) |

Let's take a closer look at *coffee*. Despite the linguistic equivalence between German *Kaffee* and Italian *caffè*, cultural differences are apparent, which remain hidden if only the linguistic level is taken into account. It is generally known that we are dealing with two different types of coffee with different preparation processes. At first glance, this comparison at the word level seems very trivial, but what is less known, however, and can therefore also lead to misunderstandings, is that the cultural concept of *drinking coffee* differs in Germany compared to Italy. The question "Shall we have a coffee?" can lead to major misunderstandings. Some Germans become irritated when they are invited to have coffee in Italy because the Germans usually link "drinking coffee" to the cognitive concept of the German *Kaffeeklatsch*, which includes sitting at a table, having a chat and probably eating cake (*Kaffee und Kuchen*). By contrast, they experience that drinking a coffee in Italy is all about standing at the counter with an espresso, not sitting comfortably at a table, not chatting for a long time and most probably not even having a piece of cake. Everything is over in a few minutes. If you are offered an espresso as a guest at home in Italy, the coffee is usually served perfectly, i.e. the amount of sugar requested by the guest is added and stirred. A German may feel uncomfortable or even offended in his autonomy. It might seem to them that they are not capable of doing this themselves. Misunderstandings like these occur frequently in everyday life, and are usually resolved through tolerance and understanding. If we move on an international political level, cultural differences can also lead to more serious conflict if the parties are not familiar with the other culture and are not aware of differences in specific cultural contexts.

## 2.    Mediation and Lexicography

### 2.1    Cultural items in an Italian–German dictionary — a case of mediation

In bilingual dictionaries — even online — cultural items are usually considered as equivalences and their cultural level and possible differences between cultures/ culture nations are not taken into account (Fig. 5 and Fig. 6).
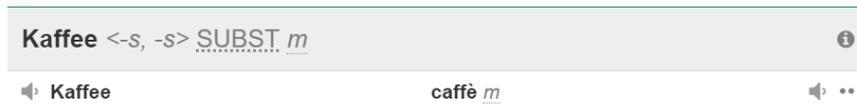
**Fig. 5:**   Excerpt from "Kaffee" in the bilingual version German–Italian of PONS[7]



**Fig. 6:**   Excerpt from "Kaffee" in the bilingual version German–Italian of LEO[8]

An exception with regard to the German–Italian language pair are the diction-
aries Giacoma Kolb 2019 and Verdiani 2010 in which cultural information has
been added in specially designed boxes at the end of the dictionary entry
(Table 1 and Table 2).[9]

These two examples show how important cultural information has been
added. The entry points out that *Frühstück* in German-speaking countries — in
contrast to the customs and traditions in Italy — can go on until lunchtime on
certain days and that both savoury and sweet food can be served. Additional
explanations in the case of *Kaffee* refer not only to the fact that the typical coffee
is a filtered coffee, but that the Italian espresso has become equally widespread;
therefore naming the two types should be done carefully to distinguish clearly
between the two. The entry also explicitly refers to the habit of having cake
with coffee, usually in the afternoon and especially during the holiday season.

| Original text | Translation in English |
|---|---|
| ***Frühstück*** ● Il *Frühstück*, "prima colazione", è un pasto importante e sostanzioso in tutti i paesi di lingua tedesca. La domenica e i giorni di festa può addirittura protrarsi per tutta la mattinata sconfinando nell'ora del pranzo. In genere l'offerta è vastissima: caffè, tè, latte, succhi di frutta, varie qualità di pane e panini (→ **Brot**), burro, marmellata, miele, formaggio, prosciutto, salumi, uova alla coque (*weiches Ei*) o strapazzate (*Rührei*), yogurt con o senza müsli o altri cereali. | ***Frühstück*** ● *breakfast*, "prima colazione", is an important and substantial meal in all German-speaking countries. On Sundays and feast days, it can even extend throughout the morning into the lunch hour. Generally, there is a wide range on offer: coffee, tea, milk, fruit juices, various types of bread and rolls (→ **Brot**), butter, jam, honey, cheese, ham, cold meats, soft-boiled eggs (*weiches Ei*) or scrambled eggs (*Rührei*), yoghurt with or without müsli or other cereals. (Translation from Italian to English by the author) |

**Table 1:**    The cultural box for "Frühstück" in the bilingual dictionary German–Italian of Giacoma Kolb 2019 (online)

| Original text | Translation in English |
|---|---|
| ***Kaffee*** ● Il caffè è una bevanda fondamentale per le aree di lingua tedesca. Viene preparato versando lentamente acqua bollente sopra il caffè macinato, contenuto in un filtro (per questo si parla anche di *Filterkaffee*). Tale procedura è eseguita di norma da una macchina per il caffè (*Kaffeemaschine*), ma può anche essere effettuata a mano. Sempre più frequente è anche l'espresso di tipo italiano, che viene chiamato *Espresso* proprio per distinguerlo dal caffè lungo tradizionale.Molto diffusa in Germania è la tradizione del *Kaffee und Kuchen*, che consiste nel consumare caffè e qualche dolcetto a metà pomeriggio in compagnia, soprattutto nei giorni festivi (→ **Kaffeeklatsch**, → **Kaffeekränzchen**). | ***Kaffee*** ● coffee, is an important drink in German-speaking areas. It is prepared by slowly pouring boiling water over ground coffee in a filter (hence the term *Filterkaffee*). This procedure is usually performed by a coffee machine (*Kaffeemaschine*), but can also be done by hand. Also increasingly common [in Germany] is Italian-style coffee, which is called *espresso* precisely to distinguish it from traditional *Filterkaffee*. Very widespread in Germany is the tradition of *Kaffee und Kuchen*, which consists of drinking coffee and having a piece of cake in the middle of the afternoon in company, especially on holidays (→ **Kaffeeklatsch**, → **Kaffeekränzchen**) (Translation from Italian to English by the author) |

**Table 2:**    Excerpt from "Kaffee" in the bilingual dictionary German–Italian of Giacoma Kolb 2019 (online)

## 2.2    Mediation activities and strategies — tools for a successful lexicographer

In Giacoma Kolb (2019) and Verdiani (2010), these space-consuming information boxes are present in the print version, too. This means that an adequate representation of cultural items does not primarily depend on the space available in the dictionary. In contrast, both PONS and LEO, where the space argument does not play a role, lack qualitatively reliable cultural information. There we even find the tendency to link all word entries to one linguistic equivalent in the other language, so that a bi-directional search is possible, which makes the same word appear as a lemma in back and forward translations. (cf. Figs. 5, 6) This falsely reinforces the assumption of a 1:1 correspondence, which in reality does not exist. So it is not a question of whether a dictionary is printed or made available online, but whether the lexicographers are aware of the specific cultural meanings and of the intercultural divergences, as well as on whether they explicitly choose to represent such differences in the dictionary (Nied Curcio 2020: 200).

The prerequisite for this approach is that lexicographers — even in a time of digitalisation, in which translation machines and computer-assisted translation tools surpass dictionaries — have very good knowledge of both their own culture and the culture of the target user. Furthermore, lexicographers — especially in bilingual or plurilingual dictionaries — should act as *mediators* in communicating the adequate description of cultural items. Only on this condition can culturally specific items have a chance of being included in a dictionary at all, and being adequately presented.

What does it mean to act as a mediator? Mediators are intermediaries between interlocutors who cannot understand each other directly. They act as social agents who build bridges and help construct or mediate meaning, generally from one language to another (cross-linguistic mediation) and sometimes within a language, between different varieties (e.g. standard language — dialect) or modalities (e.g. from spoken language to sign language). A very good mediator (C2 level)

> [c]an mediate effectively and naturally, taking on different roles according to the needs of the people and situation involved, identifying nuances and undercurrents and guiding a sensitive or delicate discussion. Can explain in clear, fluent, well-structured language the way facts and arguments are presented, conveying evaluative aspects and most nuances precisely, and pointing out sociocultural implications (e.g. use of register, understatement, irony and sarcasm) (Council of Europe 2018: 105).

In the CEFR 2018 *mediation* is differentiated into *mediation activities* and *mediation strategies*. Mediation activities are further subdivided into *mediating a text*, *mediating concepts* and *mediating communication*.
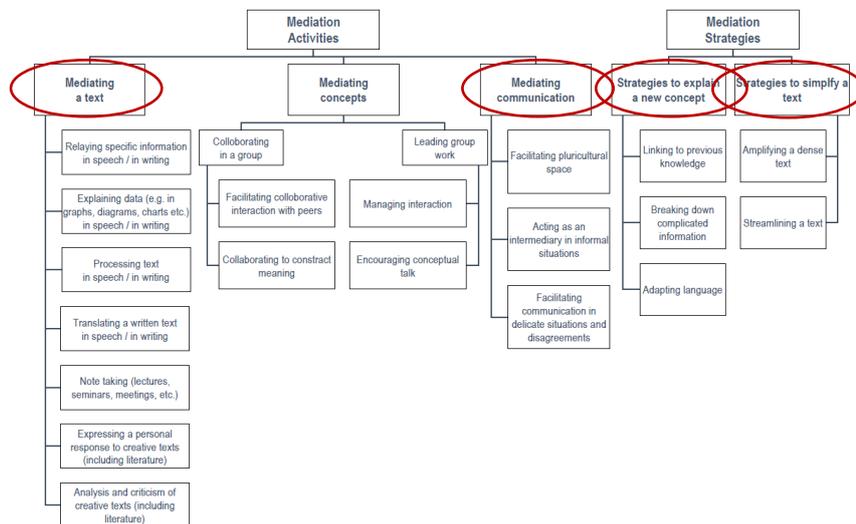
**Fig. 7:**    Mediation activities and mediation strategies (CEFR 2018: 104)

Of particular concern to the lexicographer as a mediator are *mediating a text*, *mediating communication,* and effectively using *mediation strategies* while describing the meaning of cultural items.

> *Mediating a text* involves passing on to another person the content of a text to which they do not have access, often because of linguistic, cultural, semantic, or technical barriers. [...] *Mediating communication:* The aim of mediating communication is to facilitate understanding and to shape successful communication between users/learners who may have individual, sociocultural, sociolinguistic or intellectual differences in standpoint (CEFR 2018: 106).

When *mediating a text*, especially *relaying a text in speech or in writing*, selection is fundamental. The mediator has to decide on the specific content of the source text which has to be conveyed to somebody else. The mediator — and this is also valid when preparing a lexicographic article — must first select from the vast volume of information available and then transfer it to the target person in a more condensed but still truthful/accurate form.

By *Mediating communication,* the authors of the CEFR intend mediators to facilitate the understanding of other people's needs and previous knowledge and they expect the capacity to switch from their own perspective to that of the other, keeping both perspectives in mind and facilitating the communication. Mediators, therefore, need to have a well-developed emotional intelligence and empathy for the viewpoints and emotional states of other participants. They have to show interest in understanding cultural norms, demonstrate sensitivity and respect for

different sociolinguistic and sociocultural points of view, and anticipate mis-understandings which could arise because of the differences (CEFR 2018: 122).

Especially with regard to cross-linguistic mediation, being a good media-tor includes social and cultural competence, as well as plurilingual competence. (CEFR 2018: 106) Plurilingual competence involves the ability to call flexibly upon an inter-related, uneven, plurilinguistic repertoire to:

— switch from one language or dialect (or variety) to another;
— express oneself in one language (or dialect, or variety) and understand a per-son speaking another;
— call upon the knowledge of a number of languages (or dialects, or varieties) to make sense of a text;
— recognise words from a common international store in a new guise;
— mediate between individuals with no common language (or dialect, or vari-ety), even with only a slight knowledge oneself;
— bring the whole of one's linguistic equipment into play, experimenting with alternative forms of expression;
— exploit paralinguistics (mime, gesture, facial expression, etc.) (CEFR 2018: 28).

The mediator's task — as well as the lexicographer's, especially in describing cul-tural items — is to facilitate pluricultural space:

> Rather than simply building on his/her pluricultural repertoire to gain accept-ance and to enhance his own mission or message […], he/she is engaged as a cultural mediator: creating a neutral, trusted, shared 'space' in order to enhance the communication between others. He/she aims to expand and deepen inter-cultural understanding between participants in order to avoid and/or overcome any potential communication difficulties arising from contrasting cultural view-points. Naturally, the mediator him/herself needs a continually developing aware-ness of sociocultural and sociolinguistic differences affecting cross-cultural com-munication (CEFR 2018: 122).

To be able to realise this goal, the mediator should resort to various communi-cation strategies in a mediation situation. To explain new concepts and mean-ings s/he should *link to previous knowledge* ("[…] introduce complex concepts (e.g. scientific notions) by providing extended definitions and explanations which draw upon assumed previous knowledge") (ibid.: 128)", *break down complicated information* ("[…] facilitate understanding of a complex issue by explaining the relationship of parts to the whole and encourage different ways of approaching it") (ibid.), and *adapt their language* ("[…] adapt the language of a very wide range of texts in order to present the main content in a register and degree of sophistication and detail appropriate to the audience concerned") (ibid.). For the *simplification of a text* it is necessary to *amplify a dense text* ("[…] elucidate the information given in texts on complex academic or professional topics by elabo-rating and exemplifying"), and to *streamline a text* ("[…] redraft a complex

source text, improving coherence, cohesion and the flow of an argument, whilst removing sections unnecessary for its purpose") (ibid.: 129).

By acquiring a decent level of intercultural competence, using mediation and applying mediation strategies, the lexicographer will not only be able to select an important cultural item for the (bilingual/plurilingual) dictionary but also adequately present it. In this way the lexicographer will become a successful mediator between cultures.

The importance of cultural knowledge, intercultural awareness, and the consideration of mediation as an activity and strategy for adequately presenting lemmas in bilingual dictionaries was already noted by Robert Morrison for the compilation of the first English–Chinese dictionary in 1815. It would be important to bring these issues back into focus in lexicography and in the training of lexicographers.

> The way Morrison positioned himself as a bilingual lexicographer between two very different languages and their associated cultures, which had very little contact and interaction prior to his task of dictionary compilation, was significant. His dictionary entries, illustrative examples, and detailed explanations went beyond typical bilingual lexicography. His efforts were more an experience in intercultural mediation than merely giving lexical equivalents. He ensured the provision of detailed cultural and contextual explanations of the Chinese language, making the language and associated culture accessible to a Western readership wishing to learn Chinese and understand its culture through his text. Morrison sought not only to enhance the understanding of his English equivalents through providing additional illustrative phrases, he also relied on an additional strategy in trying to mediate his understanding of the culturally embedded nature of the meanings he applied to characters and words included in his dictionary. He provided additional, non-linguistic cultural information through explanatory notes, which quoted or described information drawn from original Chinese works, from the Chinese classics, popular literature, and contemporary official notices that reflected the philosophy, values, social systems, and everyday experiences of Chinese, as expressed through language. His inclusion of illustrative examples, commentaries, and explanations went far beyond the work of traditional bilingual lexicography and provided extensive insights into the world of Chinese culture through language, in a manner described as a culture-oriented approach […] to dictionary compilation. (Scrimgeour 2016: 444-445)

As we have seen, this historical reference indicates that the recognition of cultural knowledge in dictionary compilation dates back to at least 1815, emphasizing the enduring importance of these considerations in lexicography and lexicographer training. Lexicographers have to mediate communication. They have — like mediators — to bridge linguistic and cultural gaps, facilitating communication. They have to be able to convey nuanced meaning, with sensitivity and respect for different sociolinguistic and sociocultural points of view, with emotional intelligence and empathy, with the target user in mind, who do not have a direct access to the culture and the various culture-bound items described by the

lexicographer. With their plurilingual competence, cultural knowledge, and inter-cultural awareness lexicographers have to facilitate pluricultural space. To achieve the goal, lexicographers should resort to various communication strategies, just as mediators do, i.e. link to previous knowledge, adapt their language, explain in clear words the cultural information, all this to create understanding of specific cultural information and to guarantee a successful interpersonal and inter-group communication.

## Endnotes

1.  „[…] Identitätsträger eines nationalen/ethnischen Gebildes, einer nationalen/ethnischen Kultur — im weitesten Sinne — und werden einem Land, einer Region, einem Erdteil zugeordnet." (Markstein 1998: 288)

2.  „kulturellen Inkongruenz" (Stolze 1999: 224)

3.  "Basandosi sugli studi teorici di Molina Martínez (2006) e Luque Nadal (2009) e combinando le diverse sfumature che assume questo concetto, possiamo fornire la seguente definizione per la metalessicografia contrastiva: la parola culutrale è il termine o l'espressione che potrebbe far parte di un articolo lessicografico, il cui contenuto culturale, semantico o pragmatico appartiene a una lingua determinata e il cui trattamento dipende dal rapporto interculturale fra le lingue messe a contatto. La constraitivtà, cioè il rapporto esistente fra due sistemi linguistici messi a confronto, determina il modo di gestire i contenuti all'interno della traduzione o dell'articolo lessicografico." (Sanmarco Bande 2017: 292)

4.  "Il metodo più efficace nel trattamento delle parole culturali non solo deve offrire il traducente più vicino al termine della lingua di partenza, ma concerne tutta la microstruttura, con la combinazione di diverse strategie: chiarimenti fra parentesi, numerose etichette linguistiche, esempi chiari e ben contestualizzati, osservazioni di tipo semantico, culturale e pragmatico, tutto ciò rafforzato da una tipografia mirata a questo scopo. È necessario continuare a rafforzare il traducente con altre parti della microstruttura, perché senza il sostegno di questi strumenti è molto probabile che l'utente interpreti sbagliatamente le informazioni fornite all'interno dell'articolo." (Sanmarco Bande 2017: 303)

5.  "Kulturspezifika sind kulturell geprägte Wörter, typisch für eine bestimmte Kultur, die weder ein sprachliches noch ein kulturelles Äquivalent in einer anderen Sprache haben. […] (Nied Curcio 2020: 186)

6.  "Als Kulturwörter werden kulturell geprägte Wörter bezeichnet, die typisch für mehrere (mindestens zwei) Kulturen sind, und aus diesem Grund dafür sprachliche Äquivalente in der anderen Sprache existieren. Die Entsprechung gilt jedoch nicht auf der kulturellen Ebene; die kulturelle Divergenz ist ohne interkulturelle Kompetenz auf den ersten Blick nicht erkennbar. Es kann sich dabei auch um einen Internationalismus handeln […] oder/und um sog. *Falsche Freunde* […]." (Nied Curcio 2020: 186-187)

7.  https://de.pons.com/%C3%BCbersetzung/deutsch-italienisch/Kaffee

8.  https://dict.leo.org/italienisch-deutsch/Kaffee

9.  For a more detailed description cf. Nied Curcio 2020.

## References

### A.    Dictionaries

**Giacoma Kolb 2019** = Giacoma, Luisa and Susanne Kolb. 2019. *Il nuovo dizionario di Tedesco. Dizionario Tedesco–Italiano / Italiano–Tedesco. Das Großwörterbuch Deutsch–Italienisch / Italienisch–Deutsch.* Fourth edition. Bologna/Stuttgart: Zanichelli PONS.

**LEO** = https://www.leo.org/englisch-deutsch (Accessed 15 January 2023.)

**PONS** = https://de.pons.com/ (Accessed 15 January 2023.)

**Verdiani 2010** = Verdiani, Silvia. 2010. *Tedesco Junior. Dizionario di apprendimento della lingua tedesca. Tedesco Italiano — Italiano Tedesco.* Torino: Loescher.

### B.    Other literature

**Council of Europe.** 2018. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors.* Strasbourg: Council of Europe.
https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989
(Accessed 21 December 2022.)

**Gudykunst, William B. and Stella Ting-Toomey.** 1988. Culture and Affective Communication. *American Behavioral Scientist* 31(3): 384-400.
https://doi.org/10.1177/000276488031003009

**Luque Nadal, Lucía.** 2009. Los culturemas: ¿unidades lingüísticas, ideológicas o culturales? *Language Design* 11: 93-120.
http://elies.rediris.es/Language_Design/LD11/LD11-05-Lucia.pdf (Accessed 15 January 2023.)

**Marco, Josep.** 2019. The Translation of Food-related Culture-specific Items in the Valencian Corpus of Translated Literature (COVALT) Corpus: A Study of Techniques and Factors. *Perspectives* 27(1): 20-41.
DOI:10.1080/0907676X.2018.1449228 (Accessed 21 December 2022.)

**Markstein, Elisabeth.** 1998. Realia. Snell-Hornby, Mary, Hans G. Hönig, Paul Kußmaul and Peter A. Schmitt (Eds.). 1998. *Handbuch Translation*: 288-291. Tübingen: Stauffenberg.

**Molina Martínez, Lucía.** 2006. *El otoño del pingüino. Análisis descriptivo de la traducción de los culturemas.* Castelló de la Plana: Publicacións de la Universitat Jaume I.

**Newmark, Peter.** 1988. *A Textbook of Translation.* London: Prentice-Hall.

**Nied Curcio, Martina.** 2020. Kulturell geprägte Wörter zwischen sprachlicher Äquivalenz und kultureller Kompetenz. Am Beispiel deutsch–italienischer Wörterbücher. *Lexicographica* 36: 181-204.

**Persson, Ulrika.** 2015: *Culture-specific Items. Translation Procedures for a Text about Australian and New Zealand Children's Literature.* MA Thesis. Växjö: Linnaeus University, Sweden.
http://www.diva-portal.org/smash/get/diva2:850704/FULLTEXT01.pdf (Accessed 21 December 2022.)

**Sanmarco Bande, María Teresa.** 2017. Le parole culturali della gastronomia nella lessicografia italospagnola odierna. Bajini, Irina, M.V. Calvi, G. Garzone and G. Sergio (Eds.). 2017. *Parole per mangiare. Discorsi e culture del cibo*: 289-306. Milan: LED.

**Scrimgeour, Andrew.** 2016. Between Lexicography and Intercultural Mediation: Linguistic and Cultural Challenges in Developing the First Chinese–English Dictionary. *Perspectives* 24(3): 444-457.
DOI: 10.1080/0907676X.2015.1069859 (Accessed 21 December 2022.)

**Stolze, Radegundis.** 1999. *Die Fachübersetzung. Eine Einführung.* Tübingen: Narr.

# Eppur si muove: Lexicography is Becoming Intelligent!

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

**Abstract:** The paper focuses on an ongoing R&D project which the author is conducting together with Spanish lexicographers as well as computer scientists from a high-tech company specialized in online dictionaries and language services. The objective is to develop an AI-powered Spanish writing assistant for both native and non-native writers and learners. After briefly discussing the current experiences with digital writing assistants, the paper will detail the concrete project, where the lexicographers' task is, on the one hand, to contribute to the training of the underlying language model, and on the other hand, to outline a model for good communication between the tool and its users. Based on a study of practice in existing writing assistants, the paper will then formulate a set of principles for user communication that will be implemented in the writing assistant. The article gives examples of how this implementation takes place, what new challenges it poses, and how the writing assistant will eventually work. Finally, it discusses the lexicographers' new tasks and outlines some perspectives for future work.

**Keywords:** WRITING ASSISTANTS, USER COMMUNICATION, LEXICOGRAPHICAL CONTEXTUALIZATION, INTEGRATED DICTIONARIES, GLOSSES, LEXICOGRAPHICAL DATA OF A NEW TYPE, INCIDENTAL LEARNING, INTENTIONAL LEARNING, LEMMA-CENTRED LEXICOGRAPHICAL DATABASES, PROBLEM-CENTRED LEXICOGRAPHICAL DATABASES

**Opsomming: Eppur si muove: Die leksikografie word intelligent!** In hierdie artikel word gefokus op 'n deurlopende N&O-projek wat die outeur saam met Spaanse leksikografie, asook rekenaarwetenskaplikes van 'n hoëtegnologiemaatskappy, wat spesialiseer in aanlynwoordeboeke en taaldienste, onderneem. Dit het die ontwikkeling van 'n KI-gedrewe Spaanse skryfhulpmiddel vir beide moeder- en niemoedertaalskrywers en -leerders ten doel. Ná 'n kort bespreking van die huidige ervaring met digitale skryfhulpmiddels word die konkrete projek, waarvan die leksikografie se taak, aan die een kant, is om 'n bydrae te lewer tot die afrigting van die onderliggende taalmodel, en aan die ander kant, om 'n model vir goeie kommunikasie tussen die werktuig en die gebruikers daarvan te beskryf, uitvoerig uiteengesit. Gebaseer op 'n studie van die ervaring in bestaande skryfhulpmiddels, sal daarna 'n stel beginsels vir gebruikerskommunikasie geformuleer word wat in die skryfhulpmiddel geïmplementeer sal word. In die artikel word voorbeelde

gegee van hoe die implementering plaasvind, watter nuwe uitdagings teengekom word, en hoe die skryfhulpmiddel uiteindelik sal funksioneer. Laastens word die leksikograaf se nuwe take bespreek en 'n paar vooruitsigte vir toekomstige werk word uiteengesit.

**Sleutelwoorde:** SKRYFHULPMIDDELS, GEBRUIKERSKOMMUNIKASIE, LEKSIKOGRAFIESE KONTEKSTUALISERING, GEÏNTEGREERDE WOORDEBOEKE, GLOSSE, 'N NUWE TIPE LEKSIKOGRAFIESE DATA, TOEVALLIGE LEER, DOELBEWUSTE LEER, LEMMAGESENTREERDE LEKSIKOGRAFIESE DATABASISSE, PROBLEEMGESENTREERDE LEKSIKOGRAFIESE DATABASISSE

## 1.    Introduction

This article is dedicated to my friend and colleague Rufus H. Gouws. His 43 years of highly productive academic work may seem a lot to many of us (and indeed it is), but it could also be argued that these 43 years represent only a short span of time from the perspective of more than 4,300 years of lexicographical history. Be that as it may, the past four decades have seen more dramatic and rapid changes than any previous period. And the transformation process shows no signs of slowing down, hence the famous quotation from Galileo included in the title.

Today, lexicography is becoming intelligent. Artificial Intelligence (AI) is penetrating the millennial discipline from various angles: data selection, preparation, storage, presentation and integration into different tools; cf. Tarp (2019). The disruption is almost total. AI-based tools are taking over more and more creative tasks, sending almost fully processed data to lexicographers, whose job is increasingly reduced to control, revision, and final editing. There are even examples of AI-generated data, which can rightly be considered lexicographical, now being presented directly to the target audience without a single lexicographer having been over it. Thus, the relationship between the human and the artificial lexicographer is constantly changing, often strained by mutual mistrust or ignorance of each other's capabilities. Time calls for both constructive thinking and *disruptive interdisciplinary collaboration* between lexicographers, on the one hand, and computer scientists who are the embodiment of the new technologies, on the other hand.

A quarter of a century ago, on the eve of the third millennium, Grefenstette (1998) asked whether there would be lexicographers in the year 3000. In the second decade of the new millennium, Rundell (2012) replied that "there will still be lexicographers, but they will no longer do the same job". A decade later, amid an unprecedented technological revolution, it can be added beyond doubt that not only their job but also the outcome of that job will be different. Until now, the vast majority of lexicographers have regarded their field to be that of dictionaries and similar reference works. In the current disruptive period, however, entirely new types of lexicographical products are emerging. As Tarp and

Gouws (2023) have suggested based on an analysis of the historical schism between *glossography* and *dictionography*, this requires a redefinition of the very subject matter of the discipline, which should not only include dictionaries, but also modern glosses and other types of lexicographical data integrated into writing assistants, reading assistants, machine translators, and other tools. It is therefore absolutely essential that today's lexicographers engage in the design of these digital tools. In this sense, their job should not only be to integrate existing lexicographical data types into them (linear innovation), but even more to develop entirely new lexicographical data types in close collaboration with computer scientists and programmers. This is what is meant by *disruptive inter-disciplinary collaboration*.

From the perspective of accessibility, Nomdedeu-Rull and Tarp (2024) propose a classification of digital dictionaries into three main types: stand-alone, embedded, and integrated. The *stand-alone dictionary* is the digital equivalent of the traditional printed dictionary that can be consulted independently of the context in which the user's information needs occur. Today, this type of digital dictionary is mostly published on Internet portals or in the form of applications that can be downloaded to laptops, tablets, and mobile phones. The *embedded dictionary* is a variant of the previous one with almost the same features. The only difference is that its search engine is located in another tool, typically a learning application, and only users who have access to it can consult the dictionary. Finally, the *integrated dictionary* is characterised by the fact that the user can consult it by clicking directly on a word or phrase that appears in a digital text. This type of digital dictionary, which is much easier to consult than the other two, is becoming increasingly common. However, although it represents an important innovation, it is also a clear expression of the persistence of *linear thinking* when it comes to putting cutting-edge technology at the service of lexicography. Nomdedeu-Rull and Tarp (2024) therefore add a fourth solution, the modern *gloss*, that is, a single lexicographical data stripped of all the traditional characteristics of a dictionary: book format, lemma, article, macrostructure, and microstructure. When designed to be visualized on demand in digital texts, the short glosses make allowance for a more fluent reading and writing process, which also creates the conditions for *incidental learning* as opposed to *intentional learning*; cf. Tarp (2022a). The latter requires immersion and contemplation, which can only be achieved by providing access to additional lexicographical data, including traditional dictionary articles. This kind of integrated combination of old, new and as yet undeveloped lexicographical data types — for example, in writing or reading assistants — may prove to be highly relevant for language learning in the digital age.

With this perspective, the paper will focus on an ongoing R&D project that the author is conducting together with computer scientists at Ordbogen A/S, a Danish high-tech company specialising in online dictionaries and language services. The objective is to create a completely new digital *Spanish writing assistant* that has many similarities to *Grammarly*, but also differs in some important

aspects, such as being both *monolingual* and *bilingual*, which requires additional functionalities and features. The writing assistant under construction is primarily driven by two interconnected, AI-powered language models called GECToR and BERT, each with its own specific function. These language models are also the main ones used to support *Grammarly*; cf. Omelianchuk et al. (2020).

Section 2 provides a brief overview of existing writing assistants and their functionalities. These observations lead directly to Section 3, where the research project and its aims are presented. Section 4 then discusses the results of a critical analysis of various *Grammarly*-like writing assistants and how they communicate with their users and explain the suggestions for improving language and style. Based on the analysis, a set of principles for good user communication in writing assistants will be formulated. The following section gives examples of how these principles are implemented and what new — and unexpected — challenges this poses. Section 6 summarises the tasks, some of them entirely new, that lexicographers have in developing user-centred writing assistants, and then defines their new role. Finally, the paper offers some perspectives and challenges for contemporary lexicographers.

## 2.    Writing assistants

Today, the writing of first- and second-language texts is increasingly done on laptops, tablets and mobile phones. At the same time, a deterioration in written language can be observed almost everywhere, especially among the younger generations. This undoubtedly calls for the development of new didactic methods that can motivate students in a new way, and it therefore seems logical to start precisely where people write, that is, on the devices mentioned. Instead of being passive writing tools, these devices — by incorporating writing assistants — can be turned into active tools that interact with users and their written language. From this perspective, it is not surprising that the last decade has seen the appearance of a number of digital writing assistants, like the ones presented by Verlinde and Peeters (2012), Granger and Paquot (2015), Tarp et al. (2017), Alonso-Ramos and García-Salido (2019), and Frankenberg-García et al. (2019), among others.

These writing assistants incorporate different technologies, with some of them, like *Grammarly* and *DeepL Write*, relying on AI-powered language models. Depending on their specific philosophy, they have different purposes, features, and functionalities. Some of them, like the *Interactive Language Toolbox*, have a pronounced didactic purpose. A few, such as *ColloCaid* and *HARenEs*, focus on a specific language problem, namely collocations. The probably best-known — such as *Grammarly*, *ProWritingAid*, and *Write Assistant* — provide more general language support for L1 and L2 writing. Most of these general writing assistants exist in only one language (usually English), whereas only a few, like *LanguageTool* and *DeepL Write*, have versions in several languages. Surprisingly,

almost all of them are monolingual, even though they claim to assist both native and non-native writers. As far as the different functionalities of the existing writing assistants are concerned, the following main types can be observed:

(1)   **Detection** of possible problems in the written text
(2)   **Correction** with suggestion of alternative solutions
(3)   **Prediction** of word terminations and next words
(4)   **Transformation** of syntax, style, sentences, etc.
(5)   **Translation** with provision of L2 equivalents
(6)   **Consultation** with look-ups in lexicographical databases

The author of this article is not aware of any writing assistant that has all the above functionalities. Most have three or four. *Write Assistant*, for instance, offers prediction, translation and consultation of lexicographical databases. *Grammarly*, *ProWritingAid*, *LanguageTool*, *DeepL Write*, and *Ginger* offer detection, correction and transformation (rephrasing). In the long run, the ideal writing assistant will probably be the one that combines the functionalities of *Write Assistant* and one of the other four, for instance, *Grammarly* (see Figure 1). Especially the possibility of consulting the meaning of suggested words and using one's own mother tongue seems to be highly relevant when writing in a second language.

| WRITE ASSISTANT | GRAMMARLY |
|---|---|
| | 1.  **DETECTION** (possible problems) |
| | 2.  **CORRECTION** (alternative suggestions) |
| 3.  **PREDICTION** (word completions, etc.) | |
| | 4.  **TRANSFORMATION** (syntax, style etc.) |
| 5.  **TRANSLATION** (L2 equivalents) | |
| 6.  **CONSULTATION** (look-ups in databases) | 6.  CONSULTATION (look-ups in databases) |

**Figure 1:**   Complementary functionalities in two writing assistant concepts

As we will see in the next section, lexicographical data are in one way or another involved in all six functionalities, but it is obviously most pronounced when it comes to consultation. In this respect, there are different levels of consultation in existing writing assistants. After a text has been typed or pasted into the tool, the problems detected are usually underlined in different colours depending on their severity and the actions users are expected to perform (see Figure 2).

The traditionel dictionary is the result of a long historical development. It started, both in China and Europe, in ancient times when the old scribes copied manuscript works from earlier periods and inserted glosses to explain obsolete and difficult words. These glosses were later compiled into glossaries which, over the centuries, developed into the modern dictionary. The process had both advantages and disadvantages and saw at least three important inovations.

**Figure 2:**    Possible problems detected by *ProWritingAid*

Users then have to click on the underlined words to find out what the problem might be and what to do about it. The window that pops up will usually offer a brief explanation of the problem as well as an alternative solution, such as a suggestion to change the spelling, inflection, wording, syntax or style. In Figure 3, *LanguageTool* suggests a correction of a spelling mistake appearing in the same text as in Figure 2. The suggested alternative words are undoubtedly *lexico-graphical data of a new type*, automatically generated by the underlying language model without any direct human intervention. And the same applies to the explanatory notes which are, of course, written by humans, but automatically visualized and inserted into specific contexts without their intervention and knowledge.
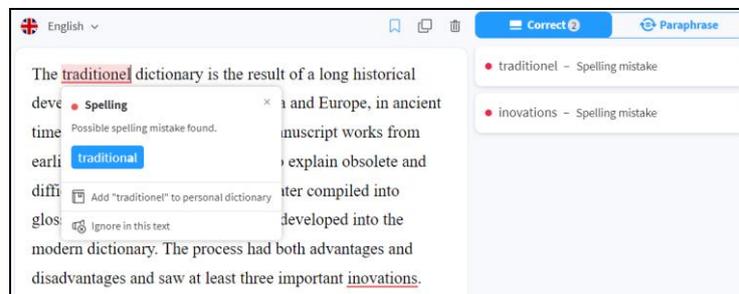
**Figure 3:**    Pop-up window with corrections in *LanguageTool*

In Figure 3, the detected problem is briefly explained with the note *Possible spelling mistake found*. When necessary, some writing assistants also allow their users to access additional explanations. *Grammarly* and *ProWritingAid*, for example, frequently offer this option in the form of a button that users can click on to get more information. This kind of user communication, and the challenges it poses, will be discussed in detail in Section 4.

If the *Paraphrase* button in Figure 3 is activated instead of the *Correct* button, and a sentence is simultaneously marked in the inserted text, *LanguageTool* immediately lists a number of alternative AI-generated sentences with a different wording that the writer can use to vary his or her language and express a
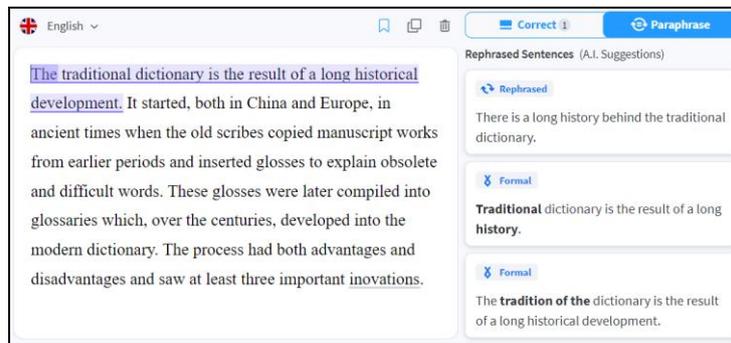
particular idea even more clearly (see Figure 4).



**Figure 4:**    AI-generated rephrasing of sentences in *LanguageTool*

The beta version of *DeepL Write* also offers alternatives to single words and rewording of whole sentences, often listing more than 20 alternative suggestions (see Figure 5). The quality of these suggestions seems quite high, but the design itself, with their presentation in scrolling windows, can make them appear rather unwieldy to users who may need some time to find their way around them. The main problem with the AI-generated rephrasing of whole sentences, however, is that it requires a rather high proficiency level in the language concerned to be able to judge which alternative suggestions are most appropriate in the concrete context, and to understand some of the new words included. As we will briefly discuss in Section 4, this is a serious challenge for *DeepL Write*, *LanguageTool*, and other similar writing assistants that offer rephrasing without further explanation, especially when targeting non-native users.
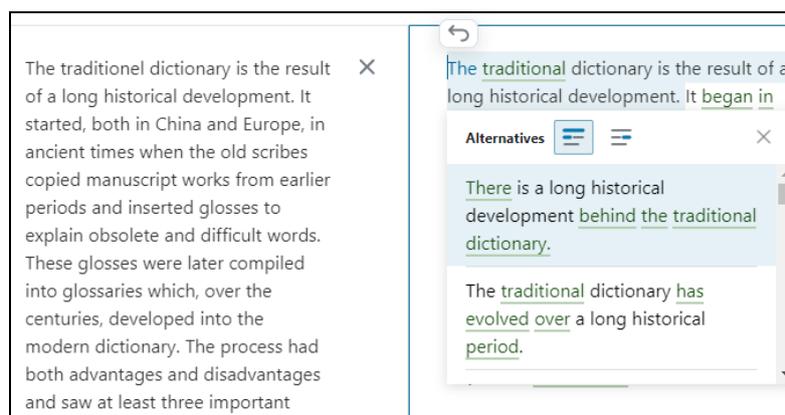


**Figure 5:**    AI-generated rephrasing of sentences in *DeepL Write*

Some writing assistants, like *Grammarly* and *ProWritingAid*, also allow their users to click on some words to get synonyms (see Figure 6). They do not, however, offer any definitions, thus leaving users who do not know the meaning of the suggested alternatives with no recourse. The consultation options are therefore only partial and, as such, particularly problematic for non-native users (this is why this functionality is written in subdued colour in Figure 1). In this way, the most inquisitive writers may feel compelled to consult one of the traditional stand-alone dictionaries (see above), thereby severely disrupting the writing flow. The problem lies neither in cutting-edge technology nor in lexicographical expertise, but in insufficient interdisciplinary collaboration between the two disciplines. This is a tradition inherited from the renowned English thesauri that list synonyms and antonyms, but usually without explaining their meaning, although this could be easily done in the digital environment.
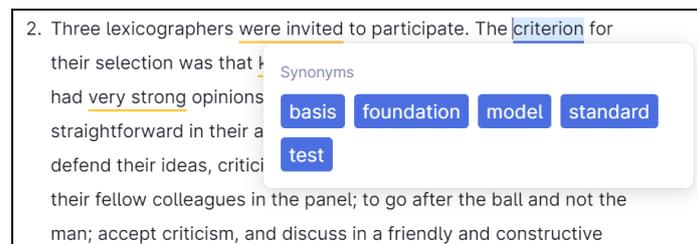


**Figure 6:**    Pop-up window with synonyms suggested by *Grammarly*

This brief review of some of the most interesting and prestigious writing assistants clearly indicates that we are facing huge and rapid changes in the way the written language is produced. In this sense, unprecedented opportunities are opening up to write quickly, correctly and in linguistically varied ways in the desired genre and style. But there are also great risks associated with the new technology if it is not used wisely. If future generations allow themselves to be seduced and become slaves of the new tools, without making a sustained effort to acquire the art of writing, they could end up losing track of what they themselves write.

After discussing the impact of writing assistants on second-language learning, Tarp (2020: 60-61) concludes that "it is above all necessary to *learn to learn by means of technology*" and, therefore, calls for "new didactic methods and convincing arguments in order to motivate language learners". With this in mind, more attention should be paid to the didactic dimension in writing assistants, which existing ones tend to downplay, especially with regard to second-language learning. In addition, a more user-friendly design is necessary to make the writing and correction process as fluid as possible, allowing for not only intentional learning, but also incidental learning. Finally, as we will discuss in Section 4, there is a need for improved user communication, with clear and understandable explanations of problems and suggestions.

## 3.    Description of the project

As mentioned, the overall objective of the current R&D project is to develop a Spanish writing assistant that in many aspects is similar to some of the general writing assistants described above. But there are also some significant differences, inasmuch as the new software aims at being both a writing tool and a learning tool, for both native and non-native users. In addition to a strictly monolingual Spanish version, it will therefore, as a research matter, include bilingual dimensions with four different languages. The first part of the project is carried out exclusively by specialists from Ordbogen A/S. The lexicographers then gradually join in, starting when the writing assistant is trained to detect specific linguistic problems in Spanish, and then continuing when it is taught to communicate with its audience in the respective languages.

The planned writing assistant is based on a set of language models with very different functions, of which the two most important, as already mentioned, are GECToR and BERT. In the following, we will focus on GECToR, which is the one that requires the most lexicographical expertise. This language model is freely available on the internet, but only in a version trained on an English corpus. The computer scientists' first task is therefore to clean it of all English and train it on a Spanish corpus. The training basically consists of special software that breaks the corpus down into its constituent sentences and then automatically introduces between one and five errors into each of them, such as misspelling or misconjugating a word, inserting or deleting a word, and swapping words. In this way, the language model learns to distinguish between right and wrong. Once this training is complete, the language model needs to be fed three different data types to perform optimally: synthetic, semi-synthetic and organic data. This is where lexicography comes in.[1]

*Synthetic data*, a term introduced by computer scientists, are lexicographical data retrieved from a monolingual Spanish database. They are made up of all the lemmata and inflectional forms contained in the database, as well as their respective grammatical categories, i.e. part of speech, gender, number, person, tense and mood, including the more than 60 conjugated forms of Spanish verbs. The aim is to enable the language model to recognise (or dispute) the words and morphological forms that users type and, at the same time, to provide it with a tool for communicating with the lexicographers who prepare the short notes and explanations, as we will see in Section 5. This phase, carried out by a computer scientist, has been completed at the time of writing.

The next phase is to feed the model with *semi-synthetic data*. This is where lexicographers take on a new role. Semi-synthetic data, another computer science term, are sentences framed like the fill-in exercises familiar to many language teachers. It concerns the filling in of words that learners typically confuse, misuse or misconjugate, involving elements such as prepositions, adverbs, pronouns, articles, congruence, tense and mood. However, the main difference with traditional fill-in exercises is that the semi-synthetic data must always be unambiguous, since binary artificial intelligence does not understand ambigu-

ity, which means that there must be only one correct solution. The sentences can either be taken from a corpus or constructed, as they are only for internal use as training material. Figure 7 shows some examples of this material, in this case concerning when — and when not — a verb in the infinitive should be preceded by the preposition *de*. The real challenge, however, is that there is no complete inventory of these types of problem and no one knows how many sentences are needed, so this is a task that is likely to continue to some extent even after the writing assistant is launched. For now, the plan is to prepare 30 correct and 30 incorrect sentences for each problem type, and their final number will probably run into tens of thousands.

| Wrong | Right |
|---|---|
| Sus productos son más fáciles vender este año. | Sus productos son más fáciles **de** vender este año. |
| Estos ordenadores son difíciles producir. | Estos ordenadores son difíciles **de** producir. |
| Este asunto es fácil aclarar. | Este asunto es fácil **de** aclarar. |
| Tengo la intención volver a vivir a Sudáfrica. | Tengo la intención **de** volver a vivir a Sudáfrica. |
| La casa es muy bonita y será fácil vender. | La casa es muy bonita y será fácil **de** vender. |
| Esta casa es muy bonita y será fácil **de** venderla. | Esta casa es muy bonita y será fácil venderla. |
| Resultó imposible **de** vender la casa. | Resultó imposible vender la casa. |
| Me es totalmente imposible **de** pagar ahora. | Me es totalmente imposible pagar ahora. |
| Es necesario **de** hacer más investigaciones. | Es necesario hacer más investigaciones. |
| Nos parece fácil **de** aclarar el asunto. | Nos parece fácil aclarar el asunto. |

**Figure 7:**    Semi-synthetic data used to train the Spanish writing assistant

*Organic data* are the ones that real users contribute when they start working with the writing assistant. They will therefore only be obtained when the latter has reached a quality that allows it to be tested and then launched. At the time of writing, the writing assistant is available only in its very first version. It can recognise a large number of language structures, but it is also full of bugs that make it useless for its intended users. Further training with semi-synthetic data is therefore required.

In order to fulfil its purpose, a digital writing assistant must not only detect typical errors, but also communicate alternative suggestions to the user in an understandable way, which is a key quality required for such a tool. In addition to providing training material, this is where lexicographers come in with their particular expertise, but in a new role. This is also the main objective of the research project. The idea is to formulate principles for good user communication in the described Spanish writing assistant. Thus, once the quality of the writing assistant has reached a satisfactory level, the next step is to investigate how it can best explain both the problems it detects and the alternative suggestions it automatically generates after appropriate training.

To this end, a critical analysis of existing English writing assistants has identified five different levels of user communication (see Section 4). However, not

all of these levels represent good communication or are relevant to the current project, which will therefore focus on the following two: (1) *Suggestions with a short explanation* and (2) *Suggestions with a supplementary explanation*. As we will see in Section 5, these suggestions must be written from the codes that the language model automatically generates for the respective *problem types*, since they cannot be elaborated for the practically *infinite number of concrete problems* that might arise.

The plan is to write the texts in Spanish. They will then be automatically translated into English, Danish, Italian and Chinese, followed by linguistic revision based on the experience of another recent research project at the Danish company (cf. Tarp 2022b). The reasons for choosing these four languages are that (1) English is the most interesting language in terms of the number of potential users; (2) Danish is relevant to the author's home university and its students; (3) Italian is close to Spanish and may pose specific problems, e.g. with false friends; and (4) Chinese is both linguistically and culturally so distant from Spanish that it may also pose specific challenges.

The next step will be to test the benefits that both native and non-native Spanish speakers derive from the explanatory notes, with a view to possibly modifying them. This will be done using qualitative methods, mainly through observation of individual users, followed by unstructured interviews and analysis of written texts. The initial intention was to answer two main research questions:

(1)   How should the balance between plain language and necessary terminology be struck so that users who cannot be expected to know the grammatical categories can fully benefit from the explanations?
(2)   Do the explanations, even when correctly translated, work for native speakers of all the languages involved, or should they be adapted to each of these languages?

However, a third research question arose during the initial drafting of the explanatory notes (see Section 5):

(3)   To what extent do the automatically generated codes allow lexicographers to write the explanations in the necessary detail so that they are not too superficial, but provide the information that users need to understand the suggested corrections and alternatives?

These three research questions will not be answered conclusively here, as they are beyond the scope of this article. It would require the writing assistant's basic parameters, including the explanations, to have reached a level that would allow it to be tested on real users. However, the answers to these questions are crucial to ensuring good user communication in the final product. They will therefore to some extent inform the discussion in the next two sections and illuminate the direction it takes.

## 4.      User communication in existing writing assistants

This section presents and discusses the results of a critical analysis of various writing assistants that are based on artificial intelligence and lexicographical databases. Most of these tools — such as *Grammarly*, *Ginger*, *LanguageTool*, and *ProWritingAid* — have a premium version, but only the free versions are used here, as the purpose is not to review them but to draw inspiration from them. This also means that writing assistants such as *WhiteSmoke* and *Writefull* are not included, as they are currently only available in premium versions. The analysis is inspired by Norman's (2013) ideas about good communication from machines to humans, and focuses on how these tools communicate with their users and explain the suggestions for improving language and style. In the terminology of this American engineer and cognitive scientist, writing assistants are "every-day things", so their design should allow their users to use them intuitively without further instruction. This implies, among other things, that they should clearly indicate to their users "what actions are possible, what is happening, and what is about to happen" (Norman 2013: 8).

It is worth noting that the writing assistants studied here largely meet this requirement, although, as we shall see, this does not necessarily mean that they are user-friendly in all respects. Their use of *affordances* and *signifiers* that allow users to navigate intuitively in the respective tools seems generally very professional, while there are some challenges related to the *feedback* provided, i.e. how the suggestions are explained or not explained. (See Tarp and Gouws (2020: 6) for a definition of these industrial designer terms.)

According to Norman's human-centred design philosophy, feedback must be given after every action in order to confirm the action and communicate its results. As such, it is an important form of communication that should be both immediate and informative. In the case of writing assistants, this means that they should not only suggest alternative solutions, but also explain why they are suggesting them.



**Figure 8:**    Wrong word in WhatsApp message after automatic correction

There are actually some widely used writing assistants that completely ignore Norman's usability advice. For example, if the autocorrect option is enabled on a smartphone, messages sent too quickly and without enough attention can easily go awry. Figure 8 shows a typical case of a WhatsApp message. A Spanish expatriate, back in her country after two years of Covid shutdown, expresses her joy at finally enjoying a *tinto de verano* (summer red wine). But when she accidentally hits the wrong key, the built-in writing assistant automatically corrects it to the nonsensical *tinto de veranda* (veranda red wine) without her consent. Such an automatic correction without the user's explicit consent is clearly the result of a bad model for the tool's communication with its users.
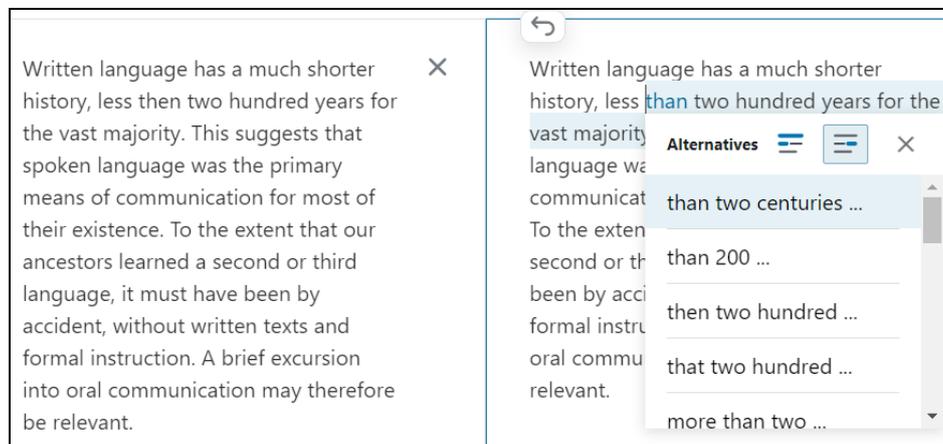


**Figure 9:**    Alternatives to confused word in *DeepL Write*[2]

*DeepL Write* is an example of a writing assistant that gives its users a choice between several alternatives, but which, at least in its beta version, does not support these suggestions with explanations to make the decision easier for the users (see Figure 9). This can cause serious problems. The design used makes it impossible for users to see whether the proposed alternative is a correction of a spelling error or a confused word, or a suggestion to use a better alternative, or simply to vary the language. Users will find that the word they originally typed often appears as number two or three in the list of alternatives. And even if the word is clearly misspelt or confused, they will sometimes also find it further down the list, as in Figure 9. This seems to be an example where artificial intelligence has been given too free a rein, and therefore needs to be further domesticated to become user-friendly. As a minimum, one could have expected *DeepL Write* to explain to its users what type of problem it had detected and provided alternatives for. This is at least what other writing assistants do, as we will see below.
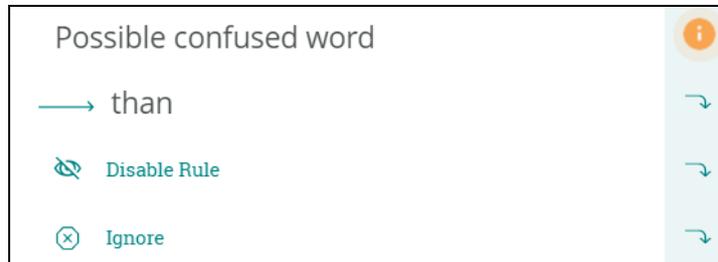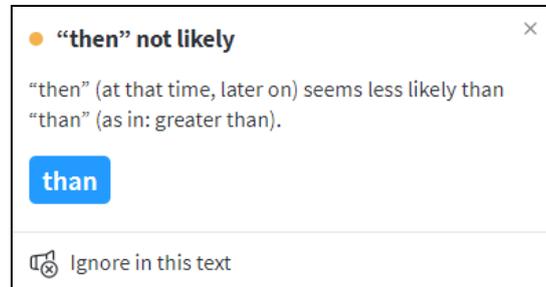
**Figure 10:**     Suggestion with short explanation and reference in *ProWritingAid*

The same text as in Figure 9 was pasted into *Grammarly*, *Ginger*, *LanguageTool* and *ProWritingAid* to see how they dealt with the above challenge. Unsurprisingly, they all offer *than* as an alternative. But beyond that, they describe the problem remarkably differently and with varying degrees of completeness. *ProWritingAid*, for example, describes it as a "possible confused word" (Figure 10), while *Grammarly* classifies it as a spelling problem and a word that "may be used incorrectly" (see Figure 11). In addition, both refer their users to a supplementary explanation (see Figures 13 and 14).
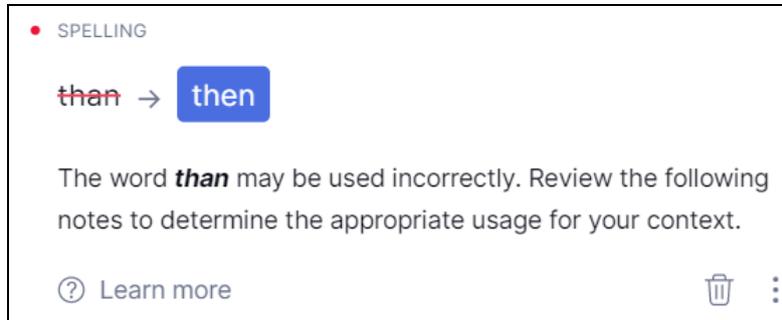


**Figure 11:**     Suggestion with short explanation and reference in *Grammarly*

As can be seen in Figure 12, *LanguageTool* describes *then* as "not likely" and "less likely than" *than* in the specific context. At the same time, it includes two parenthetical definitions of the words as "at that time, later on" and "as in: greater than", respectively. The short definitions themselves are easy to understand, but the condensed definition style known from traditional print dictionaries is likely to be an unnecessary challenge for many users. It simply takes longer time to read this type of text. As such, it hardly represents a model of good user communication.

**Figure 12:**    Suggestion window with short explanation in *LanguageTool*

*Ginger*, for its part, classifies *then* in the concrete context as a "misused word" and explains the incorrect usage as a "confusion between words" with similar orthography or pronunciation (see Figure 13). It also adds two short lexicographical definitions of the words, using grammatical terms such as *conjunction*, *sentence connector* and *comparative*, with which part of the target audience cannot be expected to be familiar. Besides, the whole default text seems too voluminous when it is just a matter of quickly explaining to users why they should replace *then* with *than*. It is a case of *relative data overload*, according to the classification proposed by Gouws and Tarp (2017: 408). A much better solution would therefore be to hide the definitions and allow users to visualize them when needed. As we will see, this is the option chosen by both *Grammarly* and *ProWritingAid*.
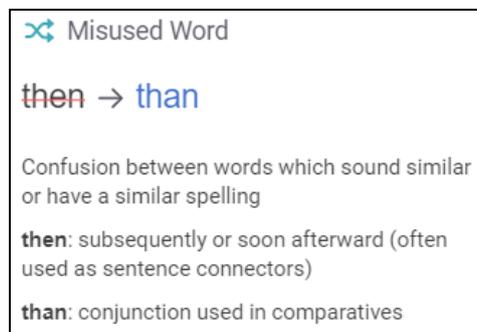


**Figure 13:**    Suggestion window with explanation in *Ginger*

Figure 14 shows the hidden text that *Grammarly* displays when users click on the *Learn more* button in Figure 11. It contains definitions of the two words and examples of their correct and incorrect use. This seems to be all a user needs to get a general idea of the meaning and usage of the two words. It is also worth noting the use of grammatical terms such as *conjunction*, *preposition* and *adverb*. Contrary to Figure 13, the use of such specialized terms makes sense here, as it

can be assumed that the users who click through are also the most motivated and knowledge-seeking. In general, the text in Figure 13 reflects an appropriate way of providing additional explanations in writing assistants. It represents good user communication.
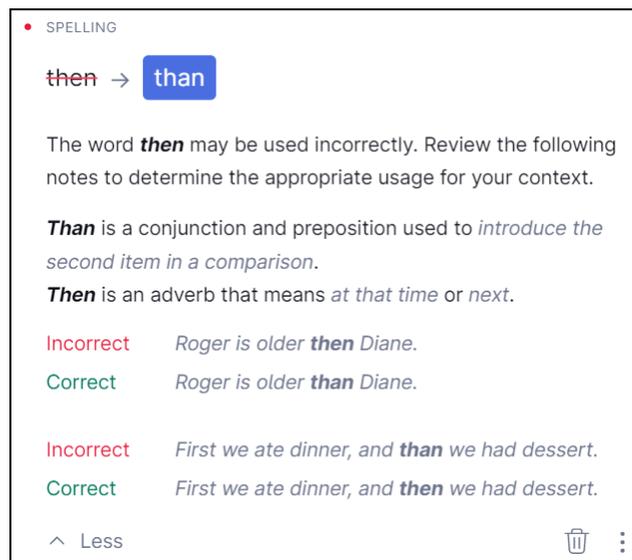


**Figure 14:**    Window with additional explanation in *Grammarly*

Unfortunately, the same cannot be said for *ProWritingAid*. When the users, after being informed that *than* is a "possible confused word", click on the icon for more information (Figure 10), a window pops up with the text shown in Figure 15. This text only deals with the general problem of confused words and does not mention the specific problem of *than* and *then* with a single word. Even if the most inquisitive users then click on *Read Full Article*, no more details about the two words appear, just a long list of other confused word pairs. The users themselves are likely to be left confused and less motivated to perform similar searches. This is clearly not an example of good user-centred communication from the writing assistant to its target audience, but rather the opposite. It is also a bit strange. A year and a half earlier, when users of *ProWritingAid* clicked on the icon for more information, they actually activated a window that explained the use of the two words in much the same way as *Grammarly*, and also included separate links to two definitions, a full article, and even a video. It is difficult to see the logic of replacing an acceptable text with one that is unspecific and too general. The merit of *Grammarly's* supplementary comment in Figure 14 is precisely that it provides a specific explanation for a specific problem, exactly what users need in a concrete writing situation.
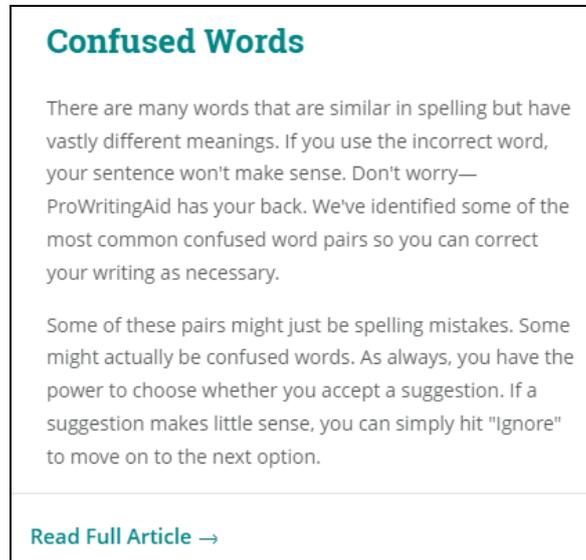
**Figure 15:**     Window with additional explanation in *ProWritingAid*

The screenshots and examples discussed above have been chosen because they are representative of the main conclusions that can be drawn from the analysis of the different writing tools. This does not mean, of course, that there are no other interesting lessons to be learned from them. The first important finding is that the communication in the existing English writing assistants can be divided into five categories according to the degree of elaboration:

(1)    auto-correction without user's acceptance (Figure 8)
(2)    suggestion without explanation (Figure 9)
(3)    suggestion with short explanation (Figures 10, 11, 12)
(4)    suggestion with supplementary explanation (Figures 14, 15)
(5)    extended explanation (after clicking through in Figure 15)

The first two categories clearly indicate poor communication, which invites mistakes and does little to support learning, and should therefore be avoided wherever possible. The fifth category consists of links to both internal and external pages. Such references may be appropriate when explaining a complex grammatical challenge, such as the use of the subjunctive in Spanish. However, it is important that references are used judiciously and do not become references for reference's sake, but are only offered when they are relevant to the use of the particular writing assistant and the associated learning process. The current research project therefore focuses on the third and fourth categories, which are directly related to the new understanding of incidental and intentional learning

in the digital context (cf. Tarp 2022a). These two categories or levels of communication do not in themselves constitute good communication. The latter also depends on a number of other factors, such as the overall design, the language used, the style and terminology, the size of the texts and the use of definitions.

The design of existing writing assistants is generally intuitive and also in line with the recommendation of Tarp and Gouws (2019, 2020) to assist users directly in the context where they have problems, without requiring them to have special reference skills. It is therefore not discussed in detail here. By contrast, language can be a challenge for many users. All the writing assistants described above use English as the sole language of explanation, at least for the time being. This is of course best for native English speakers, but for non-native speakers, including learners who are also part of the target group, it can be very challenging, especially if they are at a beginner or intermediate level. Therefore, if these writing assistants are serious about addressing non-native English speakers, they should consider explaining themselves in the native language of these users.

In terms of style, it is noticeable that most writing assistants take some care in formulating the explanations to support their alternative suggestions. This is reflected in the use of words and expressions such as *may be*, *possible*, *it seems*, *it appears*, etc. This approach is understandable, as the technology is not yet able to identify problems and suggest alternatives with 100% accuracy. For example, in Figure 2, *ProLanguageAid* recommends correcting both *manuscript* and *glosses*, even though these words are perfectly valid in the given context. Still, these reservations can be annoying and are probably not always necessary. So it is refreshing that *Ginger* dares to say things directly, as shown in Figures 13 and 16. Whatever the risk of making mistakes, this points to a more user-friendly product, for which further research and refinement is needed. Since no language model is 100% perfect, the key question in any case is whether a few errors can be accepted and, if so, where the error threshold should be.

Regarding the extent of the explanations and the use of terminology, it has already been argued that the default texts should not be too long and that specialized terms should be avoided as far as possible and preferably used only in the supplementary comments. The latter, as mentioned above, is an aspect of one of the research questions to be answered in the current project.

The *main problem* — and therefore the *main challenge* — that the analysis of the different writing assistants revealed was the lack of definitions for the suggested alternative words and expressions, as users, especially non-native English speakers, cannot always be expected to know the meaning of all of them. The risk is that users simply click on the suggested words and have no idea what they are actually typing. This could have unpleasant consequences for them and certainly does not help them to develop their own writing. It is therefore gratifying that *Ginger*, as an inspiring exception, offers such straightforward definitions, although it could be done more elegantly (see Figure 16).
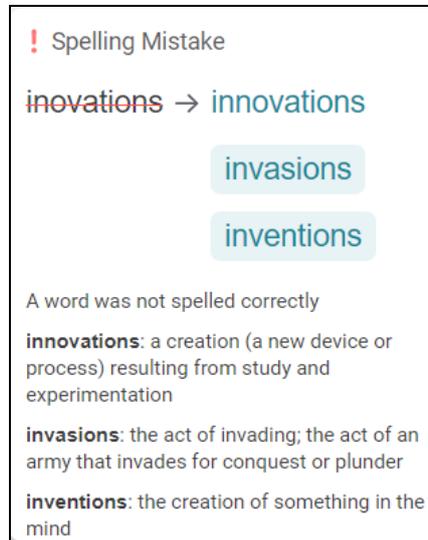
**Figure 16:**     Explanation and definition of suggested alternatives in *Ginger*

The challenge goes far beyond individual words and phrases that users may be unfamiliar with, as it also involves rephrasing entire sentences. This is a problem similar to the new AI-powered chatbots that students and others are relying on to write essays on topics they are not well versed in, and where they often have no deeper understanding of what the chatbot is writing. With good reason, Chomsky (2023) characterizes the use of these tools as "a way of avoiding learning". In the case of writing assistants, it will probably take some time to develop software that can explain the meaning of the spontaneously rephrased sentences generated by artificial intelligence. This is certainly a challenge that must be met if a further decline in written language is to be avoided. An interim solution could be to develop software that can identify the concrete meaning of individual words in a given context, based on the visionary ideas of Bothma and Gouws (2022). It might also be relevant to use constructive hints to motivate understanding. When writing in a second language, it could even be an option to provide automatic translation into the user's native language. In any case, the whole issue is an urgent topic for research.

## 5.     Some reflections on writing explanations

At the time of writing, the implementation of the above ideas has not yet begun. However, some testing has been done and is discussed here. So far, the GECToR language model has generated over 8,000 internal codes as a prerequisite for writing short explanations to support the alternative suggestions. This number is likely to increase to 30-40,000 as the model is trained further.

Most of the problems currently identified relate to replacing single words, where writing explanations would be relatively straightforward, although time-consuming. However, there are also a large number of cases involving the conjugation of Spanish verbs. In this respect, some of the generated codes are elaborate and easy to handle, while others are very short and uninformative. Below are two examples to illustrate the challenges.
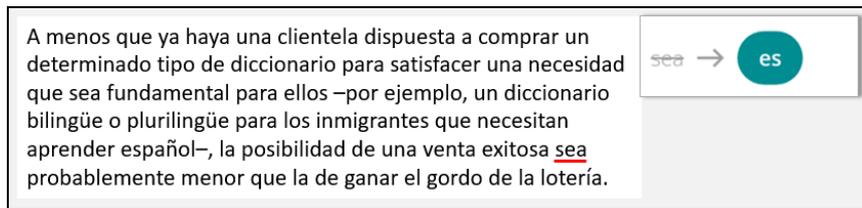


**Figure 17:**    Writing assistant with suggestion for correction

Figure 17 shows a Spanish text pasted into the new writing assistant. One of the verbs was then incorrectly changed from the indicative to the subjunctive and the software immediately reacted by underlining the error and suggesting the correct conjugation. The correction was linked to the following internal code, which the GECToR language model had automatically generated after training:

—    $TRANSFORM_INFLEXION_verb_verb-indicative-present-third-person

The first part in capital letters indicates what the main problem is, i.e. that the inflection needs to be changed. The second part in lower case is based on synthetic data from a lexicographical database and indicates a verb in the third person present indicative as an alternative. Using a specially designed interface, the lexicographer can now write an explanation that links directly to the internal code. This explanation, translated into English in Figure 18, consists of a macro in which the two words in bold are inserted. From now on, whenever there is a typing problem involving the same code, the writing assistant will display this text and insert the two concrete words involved in the particular case.
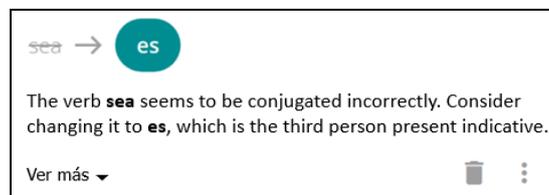


**Figure 18:**    Window with short explanation of the suggestion to change *sea* to *es*

In Figure 19, another Spanish verb was incorrectly changed from the indicative to the subjunctive and immediately marked by the language model, i.e. a problem that is grammatically identical to the one above.



Ya han pasado los días en los que el lexicógrafo –como fue el caso de María Moliner y Manuel Seco– trabaja en solitario en su diccionario durante años para luego buscar a una editorial, o un tipógrafo y un encuadernador, para culminar el proyecto. Por supuesto, esto todavía puede hacerse si el autor _esté_ dispuesto a asumir los costes, pero un diccionario compilado de esta manera dejaría mucho que desear tecnológicamente.

esté → está

**Figure 19:**   Writing assistant with suggestion for correction

In this case, GECToR associates the problem with the following internal code:

— TRANSFORM_INFLEXION_verb_verb-i

This difference in the treatment of two apparently identical problems is due to the fact that Spanish verbs have three different moods: indicative, subjunctive and imperative. In Figure 17, the inflected form _es_ is unique and can only be _indicative, present, singular, third person_. GECToR therefore identifies it unambiguously. By contrast, the inflected form _está_ in Figure 19 is ambiguous, as it can be both _indicative, present, singular, third person_ and _imperative, singular, second person_. Consequently, GECToR can only identify it as _verb-i_ (the first letter in both _indicative_ and _imperative_), as it starts parsing from the left.

In fact, it is the same kind of problem that explains why GECToR only indicates _verb_ and nothing else for the two incorrectly conjugated words underlined in Figures 17 and 19. The reason is that both _sea_ and _esté_ are inflected forms used in both the imperative and the subjunctive of the respective verbs (_ser_ and _estar_). Therefore, when the language model starts parsing from the left, it cannot go further than _verb_ because the two moods begin with different letters.

This problem was not foreseen, but it is logical given the conjugation pattern of Spanish verbs. The challenge is that unambiguity is required to write informative and correct explanations. It remains to be seen how to solve this problem. One possibility is to feed the language model with synthetic data of a syntactic nature, but this implies a redesign of the lexicographical databases used, something already pointed out by Fuertes-Olivera and Tarp (2020) in their discussion of _Write Assistant_. In any case, it will require interdisciplinary collaboration.

As can be seen in Figure 18, the explanation includes the term _indicative_, which in some ways goes against what was previously recommended. The rea-

son for this is that the use of the indicative and the subjunctive is so central to Spanish verbs that it is difficult to say anything meaningful without using these terms. Whether this is a wise idea, or whether another clever solution can be found, will be decided during the tests that make up the next phase of the research project. The aim is to answer the three research questions presented in Section 3, including the one in question here.

## 6.    New role of lexicography

The above discussion shows that lexicographical databases have an important role to play in the new writing assistants, although they need to be adapted to new requirements in some respects. But it also shows that there are a number of new tasks for which lexicographers are well prepared:

— Training of an AI-based language model to support the writing assistant;
— Designing a set of principles for good communication between the writing assistant and its users;
— Preparing a new type of lexicographical database, which is problem-based rather than lemma-based, and contains small texts explaining the alternative suggestions made by the writing assistant.

It could be argued that these three tasks are not lexicography, because they are not what professionals in the discipline usually do. This is true. But it is worth remembering that these *modern glosses* are very similar to the *traditional glosses* that Greek scribes — i.e. the first European lexicographers — inserted into manuscript copies in the Classical Period over two thousand years ago (cf. Hanks 2013). Nor should we forget Rundell's (2012) prediction that the lexicographers of the year 3000 will not be doing exactly the same as their colleagues of today. If the latter do not prepare themselves and move into new areas, they will be left behind by rapid technological development. The design of high-quality writing assistants is a new area that lexicographers will definitely have to cover, although it implies that the centrality of the lemma category in lexicographical work will have to be complemented by the centrality of another data category.

## 7.    Perspectives

Artificial intelligence is here to stay. Lexicographers will have to prepare themselves to interact with it in one way or another, not only in terms of writing assistants, but also in relation to other areas where artificial intelligence is encroaching on their discipline. Exciting challenges lie ahead. There are unprecedented opportunities to provide personalised lexicographical services to users in need. The list is long: writing assistants, reading assistants, learning apps, machine translators, etc. But, as we have seen, there are also major risks that need to be addressed. As written in Spanish in Figure 19, lexicographers should

not hide at home or even in traditional publishing houses. They must proactively seek out the experts who embody the latest technology. They should do this on an equal footing and with confidence, knowing that they have much to contribute to keeping technology in check and creating new techno-lexicographical masterpieces.

This positive, optimistic, and inquisitive look into the future, without fear of making mistakes, is the best way to celebrate Rufus H. Gouws's — to date — 43 years of successful academic work.

## Acknowledgments

## Endnotes

1. After the article was submitted, ChatGPT was introduced into the project as a production tool, resulting in both a modified project plan and new tasks for the lexicographers. See Huete-García and Tarp (2024) and Tarp and Nomdedeu-Rull (2024) for further details.
2. While preparing this article (January–February 2023), *DeepL Write* was launched in a beta version that is discussed here. Like similar tools, it is expected to continuously improve with new features and may therefore be modified in some aspects by the time the article is published.

## References

### A.    Writing assistants

**ColloCaid:** http://www.collocaid.uk/
**DeepL Write** (beta version): https://www.deepl.com/write
**Ginger:** https://www.gingersoftware.com/
**Grammarly:** https://www.grammarly.com/
**HARenEs:** http://harenes.taln.upf.edu/CakeHARenEs/check
**LanguageTool:** https://languagetool.org
**ProWritingAid:** https://prowritingaid.com/
**WhiteSmoke:** https://www.whitesmoke.com/
**Write Assistant:** https://writeassistant.com/
**Writefull:** https://writefullapp.com/

## B.    Other literature

**Alonso-Ramos, M. and M. García-Salido.** 2019. Testing the Use of a Collocation Retrieval Tool Without Prior Training by Learners of Spanish. *International Journal of Lexicography* 32(4): 480-497.

**Bothma, T.J.D. and R.H. Gouws.** 2022. Information Needs and Contextualization in the Consultation Process of Dictionaries that are Linked to e-Texts. *Lexikos* 32(2): 53-81.

**Chomsky, N.** 2023. Noam Chomsky on ChatGPT: It's "Basically High-Tech Plagiarism" and "a Way of Avoiding Learning". *Open Culture*, February 10, 2023

**Frankenberg-García, A., R. Lew, J.C. Roberts, G.P. Rees and N. Sharma.** 2019. Developing a Writing Assistant to Help EAP Writers with Collocations in Real Time. *ReCALL* 31(1): 23-39.

**Fuertes-Olivera, P.A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.

**Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

**Granger, S. and M. Paquot.** 2015. Electronic Lexicography Goes Local: Design and Structures of a Needs-driven Online Academic Writing Aid. *Lexicographica* 31(1): 118-141.

**Grefenstette, G.** 1998**.** The Future of Linguistics and Lexicographers: Will There Be Lexicographers in the Year 3000? Fontenelle, T., P. Hiligsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *Proceedings of the Eighth EURALEX International Congress in Liège, Belgium*: 25-41. Liège: English and Dutch Departments, University of Liège.

**Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Allan, K. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.

**Huete-García, Á. and S. Tarp.** 2024. Training an AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos* 34. (to appear)

**Nomdedeu-Rull, A. and S. Tarp.** 2024. *Introducción a la lexicografía en español: funciones y aplicaciones.* London: Routledge. (to appear)

**Norman, D.** 2013. *The Design of Everyday Things*. New York: Basic Books.

**Omelianchuk, K., V. Atrasevych, A. Chernodub and O. Skurzhanskyi**. 2020. GECToR ― Grammatical Error Correction: Tag, Not Rewrite. Burstein, J., E. Kochmar, C. Leacock, N. Madnani, I. Pilán, H. Yannakoudakis and T. Zesch (Eds.). 2020. *Proceedings of the 15th Workshop on Innovative Use of NLP for Building Educational Applications*: 163-170. Seattle: Association for Computational Linguistics.

**Rundell, M.** 2012**.** The Road to Automated Lexicography: An Editor's Viewpoint. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*, 15-30. Oxford: Oxford University Press.

**Tarp, S.** 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249.

**Tarp, S.** 2020: Integrated Writing Assistants and their Possible Consequences for Foreign-Language Writing and Learning. Bocanegra-Valle, A. (Ed.). 2020. *Applied Linguistics and Knowledge Transfer: Employability, Internationalization and Social Challenges*: 53-76. Bern: Peter Lang.

**Tarp, S.** 2022a. A Lexicographical Perspective to Intentional and Incidental Learning: Approaching an Old Question from a New Angle. *Lexikos* 32(2): 203-222.

**Tarp, S.** 2022b**.** Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology. *Lexikos* 32: 66-87.

**Tarp, S. and R.H. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.

**Tarp, S. and R.H. Gouws.** 2020. Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture. *Lexikos* 30: 470-498.

**Tarp, S. and R.H. Gouws.** 2023. A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and Implications for the Future. *Lexikos* 33: 425-427.

**Tarp, S. and A. Nomdedeu-Rull.** 2024. Who has the Last Word? Lessons from Using ChatGPT to Develop an AI-based Spanish Writing Assistant. *Círculo de lingüística aplicada a la comunicación* 97. (to appear)

**Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.

**Verlinde, S. and G. Peeters.** 2012. Data Access Revisited: The Interactive Language Toolbox. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 147-162. Oxford: Oxford University Press.

# Semantiese betrekkinge en woordeboekstrukture in 'n Afrikaanse tesourus

Michele F. van der Merwe, *Departement Kurrikulumstudie: Afrikaans, Fakulteit Opvoedkunde, Universiteit Stellenbosch, Stellenbosch (michelevdm@sun.ac.za)*

**Opsomming:** In 'n tesourus word die struktuur van die woordeskat en die besondere verhoudings tussen die leksikale items in daardie woordeskat op die voorgrond geplaas. 'n Woordeskat vertoon 'n bepaalde struktuur waarbinne die verskillende leksikale items op grond van veral hulle betekenisse met mekaar saamhang. Die onderwerp van hierdie navorsing is die leksikografiese aanbieding van onderskeie semantiese betrekkinge tussen leksikale items in die verskillende uitgawes van die *Groot Tesourus van Afrikaans*. In 'n vergelykende studie tussen die twee uitgawes word die volgende navorsingsvraag gevra: Watter veranderings en verskille, indien enige, kan daar ten opsigte van die aanbieding van semantiese betrekkinge tussen die twee uitgawes waargeneem word? Semantiese betrekkinge tussen leksikale items (onder andere sinonimie, hiponimie en teenoorgesteldheid) en die aanbieding daarvan in albei tesourusuitgawes word belig. Die leksikografiese hantering van semantiese betrekkinge word bespreek teen die agtergrond van die leksikografieteorie rakende woordeboekstrukture en e-leksikografie. Aanbevelings ten opsigte van 'n meer gebruikersgerigte hantering van semantiese betrekkinge word gemaak, ook met betrekking tot 'n elektroniese tesourus.

**Sleutelwoorde:** TESOURUS, SINONIEM, TEENOORGESTELDHEID, HIPONIEM, MERONIEM, SEMANTIESE VERHOUDING, WOORDEBOEKSTRUKTUUR, KONTEKSTUELE LEIDING, GEBRUIKERSGERIG, ELEKTRONIESE TESOURUS

**Abstract: Semantic Relations and Dictionary Structures in an Afrikaans Thesaurus.** In a thesaurus, the structure of the vocabulary and the particular relationships between the lexical items in that vocabulary are foregrounded. A vocabulary shows a specific structure within which the different lexical items are connected to each other based on especially their meanings. The topic of this research is the lexicographic presentation of different semantic relations between lexical items in the different editions of the *Groot Tesourus van Afrikaans* and the *Pharos Tesourus van Afrikaans*. In a comparative study between the two editions, the following research question is asked: What changes and differences, if any, can be observed regarding the presentation of semantic relations between the two editions? Semantic relations between lexical items (among others synonymy, hyponymy and oppositeness) and their presentation in both thesaurus editions are highlighted. The lexicographic handling of semantic relations is discussed against the background lexicography theory regarding dictionary structures and e-lexicography. Recommendations regarding a more user-oriented handling of semantic relations are made, also in relation to an electronic thesaurus.

**Keywords:** THESAURUS, SYNONYM, OPPOSITE, HYPONYM, MERONYM, SEMANTIC RELATIONS, DICTIONARY STRUCTURE, CONTEXTUAL GUIDANCE, USER DIRECTED, ELECTRONIC THESAURUS

## 1.      Inleiding

Die *Groot Tesourus van Afrikaans* (voortaan GT), verskyn in 1994 en 26 jaar later word 'n uitgebreide hersiene uitgawe, naamlik die *Pharos Tesourus van Afrikaans* (voortaan TA), deur dieselfde samestellers gepubliseer.

Die tesourus is tematies georden en beskik oor drie tematies geordende toegangstrukture, naamlik die indeks, die klassifikasieplan en 'n sentrale teks. Die klassifikasieplan is identies vir beide uitgawes, aangesien dit "'n aftekening van ons konseptuele leefwêreld is" (De Stadler 2022).

In die tesourus is daar 855 konsepte ingedeel onder groter konsepkategorieë. Leksikale items wat op die genommerde konsepte betrekking het, is dan daaronder ingedeel. Sommige konsepbenamings het verander. Daar is oor die 140 000 leksikale items ingedeel in dié raamwerk, met 20 000 nuwe items. 'n Vernuwende rigting is ingeslaan deur die opname van meer variëteite van Afrikaans (De Stadler 2022).

Die 2020-uitgawe is ook elektronies beskikbaar via 'n intekenportaal van die uitgewer, Pharos. Met bestudering van die elektroniese weergawe blyk dit 'n identiese weergawe van die hardekopie-uitgawe te wees, met die enigste ooglopende verskil die voordeel om met behulp van rekenaartegnologie 'n aanlyn soektog na leksikale items uit te voer.

Die fokus van hierdie artikel is op die aanbieding van semantiese betrekkinge deur die twee verskillende uitgawes van die tesourus. 'n Vergelykende studie tussen die GT en TA word uitgevoer ten einde die volgende navorsingsvraag te beantwoord: Watter veranderinge en verskille, indien enige, ten opsigte van die aanbieding van semantiese betrekkinge kan tussen die twee uitgawes waargeneem word? Terselfdertyd word die aanbieding van semantiese betrekkinge teen die agtergrond van die leksikografiese teorie rakende woordeboekstrukture en e-leksikografie bespreek. Dit sluit aan by die leksikografiese teorie van die gebruikersgerigtheid van 'n tesourus. Hoe suksesvol word gebruikersgerigtheid ten opsigte van die aanbieding van semantiese betrekkinge in beide publikasies nagevolg? Die uitgangspunt van die artikel is dat gebruikersgerigtheid ten opsigte van die oordrag van semantiese betrekkinge suksesvol kan figureer alleenlik wanneer voldoende kontekstuele leiding aan die gebruiker verskaf word.

Eerstens word die eiesoortige tematiese aard van 'n tesourus bespreek; tweedens word semantiese betrekkinge tussen leksikale items (onder andere sinonimie, hiponimie en teenoorgesteldheid) en die aanduiding daarvan in beide tesourusse belig; en derdens word woordeboekstrukture en semantiese betrekkinge in 'n elektroniese tesourus met verwysing na toepaslike leksikografiese voorbeelde ontleed. Aanbevelings word gemaak en moontlike wysigings

voorgestel vir meer gepaste leksikografiese hantering van tesourusartikels, ook gesien in die lig van die eiesoortige aard van 'n elektroniese tesourus.

## 2.    Onomasiologiese benadering van 'n tesourus

Tydens 'n literatuursoektog blyk dit dat 'n tesourus nie baie aandag geniet in leksikografiekringe nie en nog minder in die Afrikaanse leksikografie. Bronne wat opgespoor kon word, sluit in 'n artikel van prof. R.H. Gouws (2017), aan wie hierdie bundel opgedra word, asook 'n resensie oor die GT deur Van Schalkwyk (1996). Gouws (2017: 133) verwys na die belangrike bydrae van die GT tot die Afrikaanse leksikografie. Met die strekking van beide artikels in ag geneem, lewer die tesourus 'n beduidende bydrae tot die Afrikaanse leksikografie en is daarom dus navorsingswaardig. (Die skrywer is bewus van *Woordkeusegids: 'n Kerntesourus van Afrikaans* in 1992 deur Southern Boekuitgewers gepubliseer en in 2006 as *Sinonieme en verwante woorde* deur Pharos heruitgegee. Dit is tans aanlyn beskikbaar, maar vorm dit nie deel van die fokus van hierdie artikel nie. Vos resenseer die tesourus in 1994.)

Gouws (2017: 133) wys in die tipologiese klassifikasie van woordeboeke op die tesourus as die een opvallende uitsondering op die alfabetiese ordening van woordeboeke en beskryf 'n tesourus as 'n woordeboeksoort wat met sy onomasiologiese benadering 'n ordening volgens begrippe eerder as 'n alfabetiese ordening van woorde daarstel. 'n Ordening volgens begrippe, met ander woorde 'n tematiese ordening, is volgens Gouws (2017: 134) kenmerkend van vroeëre leksikografiese werke. Leksikograwe het geredeneer dat 'n woordeboek as houer van kennis 'n afspieëling moes wees van die menslike verstand (die bekendste houer van kennis) en dat die menslike verstand nie alfabeties georden is nie, maar dat 'n mens eerder tematies dink (Gouws 2017: 134).

Dit is opvallend dat die onomasiologiese aard van 'n tesourus telkens deur navorsers as 'n wesenlike kenmerk aangedui word wanneer daar oor 'n tesourus navorsing onderneem word. Marello (1989: 1083) beskryf 'n tesourus as 'n woordeboek waarin die leksikale items gerangskik word met inagneming van hulle verhouding tot bepaalde konsepte of konsepkategorieë, maar ook met inagneming van hul onderlinge betekenisbetrekkinge. Hartmann en James (2000: 142) verwys na die sistematiese opsporing van sinoniemnetwerke tussen woorde binne semantiese domeine. Kilgarriff en Yallop (2000: 1373) definieer 'n tesourus as 'n bron waar woorde met soortgelyke betekenisse saamgegroepeer is. Volgens Kay en Alexander (2016: 367) is 'n tesourus 'n leksikografiese naslaanwerk wat leksikale feite binne semantiese domeine weergee. 'n Tesourus vermeng feite rakende taal met feite rakende die wêreld waarin taal gebruik word en as gevolg daarvan verteenwoordig die klassifikasiesisteem 'n sintese van die konseptuele woordeskat van 'n taal en die relatiewe plek van elke konsep in die groter konseptuele sisteem waarvan dit deel vorm (Kay en Alexander 2016: 367).

Volgens Hartmann (2005: 78) is konvensionele tesourusse wesenlik onomasiologies, aangesien dit gewoonlik met 'n gegewe konsep begin waarna die

mees toepaslike woord(e) gevind moet word. Aangesien die onomasiologiese aard van 'n tesourus as wesenskenmerk beskou word en dit as't ware die tesourus se konseptualisering rig, beskik 'n tesourus na verwagting oor 'n relevante teoretiese leksikografiese fundering. 'n Leksikografiese fundering kan spesifieke woordeboekstrukture insluit, gegrond op die teorie van woordeboekstrukture volgens Wiegand[1] (2002; 2003; 2005), asook 'n gebruikersgerigte paradigma. 'n Gebruikersgerigte paradigma voorveronderstel leksikografiese inligtingsoordrag op 'n betekenisvolle wyse met voldoende steun aan die gebruiker om die besondere aard van die tesourus voordelig te vind tydens woordeboekkonsultasie. Inligting oor semantiese betrekkinge en die weergee van byvoorbeeld 'n semantiese konseptualiseringsraamwerk wat tot effektiewe inligtingsontsluiting lei, kan as gebruikersgerigtheid beskou word.

In die volgende afdeling word daar na verskillende semantiese betrekkinge verwys wat in 'n tesourus figureer.

## 3.    Semantiese betrekkinge

Konsepte in 'n tesourus word met inagneming van hul onderlinge betekenisbetrekkinge rangskik (Marello 1989: 1083). In 'n tesourus word die struktuur van die woordeskat en die besondere verhoudings wat tussen die leksikale items in daardie woordeskat bestaan op die voorgrond geplaas. 'n Woordeskat vertoon 'n bepaalde struktuur waarbinne die verskillende leksikale items op grond van veral hulle betekenisse met mekaar saamhang.

Volgens De Stadler en De Stadler (2021: voorwoord) kan verskillende soorte verhoudings binne hierdie struktuur onderskei word, waaronder veral die volgende vier:

a.    Sinonimie, die verhouding tussen twee woorde waarvan die betekenisse min of meer dieselfde is, soos in die geval van *gierigheid*, *behaagsug* en *hebsug*.
b.    Teenoorgesteldheid, die verhouding wat bestaan tussen twee woorde waarvan die betekenisse in opposisie tot mekaar staan, soos in die geval van *liefde* en *haat*.
c.    Hiponimie, die verhouding tussen 'n superordinate term (oorkoepelende term) en daardie terme wat as subordinate (of hiponieme) dien. Dit is die verhouding wat bestaan tussen 'n woord soos *dier* en sy subordinate of hiponieme *bees*, *skaap*, *koedoe*, *springbok*, *leeu*, *renoster*, *olifant*, ens.
d.    Die deel-geheel-verhouding (meronimie), 'n verhouding wat bestaan tussen woorde soos *voertuig* en *stuurwiel*.

'n Tradisionele, kumulatiewe tesourus verskaf slegs 'n geklassifiseerde inventaris van sinonieme of ander semanties verwante items sonder verdere betekenisverklarings of gebruiksverklarings, en leksikale items word onder 'n bepaalde konsep gerangskik sonder om vir hulle betekenisverklarings aan te bied. In die voorwerk van GT word genoem dat dit veral tot sy reg sal kom as die tesourus

saam met 'n goeie verklarende woordeboek gebruik word (De Stadler en De Stadler 2021: voorwoord). Die implikasie is dus dat tesourusgebruikers addisionele ondersteuning met behulp van 'n verklarende woordeboek kan benodig. Addisionele ondersteuning met betrekking tot betekenisverklarings, polisemiese onderskeidings en semantiese verhoudings kan moontlik benodig word. Taalgebruikers benodig inligting rakende die nuanses van woorde sowel as sintaktiese, semantiese en pragmatiese kennis vir doeltreffende taalgebruik. In die volgende afdeling word 'n tesourus se hantering van sinonimie ondersoek.

### 3.1    Sinonimie

Die voorkoms van sinonimie in tesourusse word deur verskeie navorsers bespreek, naamlik Marello (1989), Kilgarriff en Yallop (2000), Peters en Kilgarriff (2000), Hüllen (2006), Knappe (2013), Murphy (2013), asook Gao en Liu (2019).

Volgens Murphy (2016: 442) bereik die studie van semantiese betrekkinge tussen woorde 'n hoogtepunt in die teorie van strukturele semantiek van die Saussure-, Bally- en Wortfeld-bewegings. Lyons en Lehrer ontwikkel verdere teorie vir semantieseveldebenaderings (Murphy 2016: 442). Daarvolgens word 'n semantiese leksikon intern gestruktureer volgens sinonimie, teenstelling en hiponimie. Murphy (2016: 442) wys daarop dat die benadering ten opsigte van semantiese komponent-ontleding voorsiening maak vir die aflei van verhoudinge tussen betekenisonderskeidings van woorde se betekeniskomponente: sinonieme deel sleutelkomponente, antonieme verskil in 'n sleutelkomponent en hiponieme beskik oor dieselfde komponente as hulle subordinate, plus ten minste een addisionele komponent.

Zgusta[2] (1971: 89) beskryf sinonieme as woorde met verskillende vorme, maar met identiese betekenis. Volgens Zgusta (1971: 89) kan die veronderstelde identiese betekenis op twee wyses verstaan word: eerstens as 'n absolute identiteit of tweedens as 'n baie groot ooreenkoms. Die absolute identiteit van leksikale betekenis is uiters ongewoon en kom volgens Zgusta (1971: 89) dus nie baie voor nie. Hy stel dus eerder die gebruik van naby-sinonieme in die leksikografie voor, met die veronderstelling dat een van die drie basiskomponente van betekenis (te wete aanduiding, konnotasie en toepassing) teenwoordig is.

Die leksikografiese hantering van naby-sinonimie noodsaak twee tipes inligting, naamlik sinoniemidentifikasie en sinoniemdifferensiasie (Murphy 2016: 449). Sinoniemdifferensiasie word byvoorbeeld in spesiale ingevoegde tekste in sinoniemwoordeboeke verskaf, maar tesourusse wat sinonieme lys is meer algemeen (Murphy 2016: 449). Tydens sinoniemlysting stel 'n tesourus een sinoniem as die mees neutrale lid van die kategorie voor en dit word die sentrale sinoniem, kernsinoniem of voorkeursinoniem genoem (Van Sterkenburg in Murphy 2016). TA hanteer byvoorbeeld *dood* as die sentrale sinoniem en lys *sterfte, sterfgeval, oorlye* as deftig en verouderd, en *einde, graf, slaap, die ewige slaap* as sinonieme. Die sinoniemlysting bevat naby-sinonieme met wisselende mate

van formaliteit en gebruiksfere. Naby-sinonieme soos *sterfte, sterfgeval* en *oorlye* kan as voorbeelde van meer formele taalgebruik beskou word, *einde* as informele taal en *slaap, die ewige slaap* as eufemisties. Tesourusgebruikers is produktiewe skeppers van tekste en soek dikwels meer formele woorde om die minder formele of kru woorde te vervang waaraan hulle gedink het (Murphy 2016: 449). Tydens die sinoniemlysting in TA word geen kontekstuele leiding aan die gebruiker gebied om tussen die wisselende mate van formaliteit en gebruiksfeer te onderskei nie.

Vervolgens word etlike lukraak gekose artikels se hantering van sinonimie met verwysing na GT en TA bespreek. In sommige gevalle word verskille en/of verbeterings tussen die twee uitgawes van die tesourus waargeneem.

Eerstens word die sinoniemlysting van *kok* onder die artikel 419 Voedselbereiding in GT en TA bespreek. In beide uitgawes word *kok* as sentrale sinoniem gelys. Die inskrywing in GT is soos volg:

**kok**, koksmaat, sjef, bereider, spysenier, huishoudkundige

**Figuur 1:**    kok, uittreksel uit *Groot Tesourus van Afrikaans* (artikel verkort)

Die sinoniemlysting bevat sinonieme met aansienlike verskille in gebruiksfeer en betekenis. Daar is beduidende nuanseverskille tussen 'n <u>sjef</u> en 'n <u>koksmaat</u>, 'n <u>bereider</u> en 'n <u>spysenier</u>, en 'n <u>huishoudkundige</u> en 'n <u>koksmaat</u>.

In die artikel van TA kan 'n verskil waargeneem word, naamlik die byvoeging van die woord *huiskok*. Die inskrywing in TA is soos volg:

*kok*, huiskok, koksmaat, sjef, bereider, spysenier, huishoudkundige

**Figuur 2:**    kok, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

Die woord *huiskok* kan heel moontlik eerder as 'n subordinaat van die woord *kok* in die hiponimiese verhouding tussen die leksikale items beskou word. In beide artikels word nie rekenskap gegee van verskille in die sinoniemlysting nie en onvoldoende gebruiksleiding word verskaf. Deur te verwys na opleidingsverskille of die moontlike plek waar die persone werk, byvoorbeeld 'n <u>sjef</u> in 'n hotel of restaurant en 'n <u>koksmaat</u> op 'n skip, kan meer konteks aan die gebruiker verskaf word.

Tweedens word sinoniemlysting van *predikant* onder die artikel 849 Prediking in GT en TA bespreek. In beide uitgawes word *predikant* as sentrale sinoniem gelys. Die inskrywing in GT is soos volg:

**predikant** 852, dominee, herder, sieleherder, prediker 852, hulpprediker, leke-prediker, straatprediker, leraar, lerares, hulpleraar, hulplerares, wyksleraar, wykslerares, kapelaan, weermagkapelaan, polisiekapelaan, evangeliedienaar, eerwaarde, pastoor, pastor, priester, priesteres, vader, monseigneur, pater, deken, aartsdeken, biskop 852, aartsbiskop, kardinaal, pous, rabbi, ghoeroe, imam, chatieb, geestelike, sendeling, sendingwerker, sendingwerkster, straat-werker, apostel, dissipel, ouderling, ouderlingsvrou, diakonie, diaken, diakones, diakensvrou, altaardienaar, koster, kosterin, orrelis, orreliste, kerkorrelis

**Figuur 3:**    predikant, uittreksel uit *Groot Tesourus van Afrikaans* (artikel verkort)

Daar bestaan nuanseverskille tussen byvoorbeeld 'n <u>sieleherder</u> en 'n <u>prediker</u>, 'n <u>hulpprediker</u> en 'n <u>lekeprediker</u>, 'n <u>straatprediker</u> en 'n <u>leraar</u>, 'n <u>wykslerares</u> en 'n <u>kapelaan</u>, en 'n <u>evangeliedienaar</u> en 'n <u>eerwaarde</u>. Dit is 'n voorbeeld van semanties verwante woorde wat langs mekaar gegroepeer word en met behulp van kommas van mekaar geskei word. Dit laat die gebruiker onder die indruk dat hulle naby-sinonieme of sinonieme is, terwyl dit nie die geval is nie. 'n Gebrek aan konteksleiding veroorsaak dat gebrekkige inligtingsoordrag plaas-vind en dat die gebruiker nie voldoende leiding ontvang nie.

*predikant* 852, dominee, herder, sieleherder, prediker 852, evangelis, hulpprediker, lekeprediker, straatprediker, leraar, lerares, hulpleraar, hulplerares, wyksleraar, wykslerares, kapelaan, weermagkapelaan, polisiekapelaan, evangeliedienaar, eerwaarde, pastoor, pastor, priester, priesteres, vader, monseigneur, pater, deken, aartsdeken, offisiant, biskop 852, aartsbiskop, kardinaal, pous, rabbi, ghoeroe, imam, chatieb, gatiep, geestelike, sendeling, missionaris, sendingwerker, sendingwerkster, straatwerker, apostel, dissipel, ouderling, ouderlingsvrou, diakonie, diaken, diakones, diakensvrou, altaardienaar, koster, kosterin, orrelis, orreliste, kerkorrelis; predikamp

**Figuur 4:**    predikant, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

Met 'n eerste lees word en met behulp van 'n elektroniese woordtelling word gemerk dat die artikel van GT uit 58 leksikale items bestaan en die artikel van TA uit 63 leksikale items. Sekere verskille tussen die twee uitgawes kan waarge-neem word, naamlik die byvoeging van leksikale items in TA.

Die eerste vier naby-sinonieme en hulle volgorde is identies in beide arti-kels. Twee leksikale items word as naby-sinonieme in TA bygevoeg, naamlik *evangelis* (ook eerste genoem) en *offisiant*. Volgens die *Handwoordeboek van die Afrikaanse Taal* (HAT) word die woord *evangelis* as "'n ongeordende prediker" verklaar en 'n *offisiant* as "priester wat die mis opdra". Laasgenoemde is van 'n etiket, naamlik "Rooms-Katoliek" voorsien. Die woord *evangelis* beskik oor drie polisemiese onderskeidings, wat die semantiese intrige verder kompliseer. Die twee woorde kan dus nie as naby-sinonieme gelys word nie. Konteksgebaseerde leiding word nie verskaf nie, maar ter versagting kan daar aangevoer word dat *offisiant* tussen *aartsdeken* en *biskop* staan en as deel van die gebruiksfeer beskou

behoort te word; iets waarvan 'n gesofistikeerde tesourusgebruiker moontlik bewus kan wees.

Die artikel in TA word met drie leksikale items uitgebrei, naamlik *chatieb*, *gatiep* en *predikamp*. Die eersgenoemde twee woorde is geensins naby-sinonieme nie, maar is wisselvorme van mekaar en behoort so aangedui te word.

Die woord *predikamp* word met 'n kommapunt van ander gelyste woorde geskei. Volgens die toeligtende aantekeninge skei kommas leksikale items in 'n artikel en word kommapunte gebruik om fyner onderskeidings aan te dui. 'n Betekenisverskil kan dus tussen die gelyste groep naby-sinonieme en *predikamp* bestaan. Die toeligting in verband met die gebruik van die tipografiese konvensies in hierdie verband is nie voldoende nie, aangesien die gebruiker nie weet of daar 'n sinonimiese betrekking al dan nie tussen die leksikale items bestaan nie. Die gebruiker weet ook nie in watter volgorde, indien enige, die leksikale items verskaf word nie.

Derdens word semantiese betrekkings tussen leksikale items van *atletiekbroekie* onder die artikel 729 Atletiek in GT en TA bespreek.

In die GT-artikel word daar nie sinonieme of naby-sinonieme van die lemma verskaf nie; vergelyk byvoorbeeld **atletiekbroekie** wat 'n subparagraaf onder *atletiek* vorm. Daar kom *atletiekfrokkie, atletiekskoen, hardloopskoen, spykerskoen, wegspringblokke, skyf, werpskyf, diskus, gewig, spies, hamer, hoogspringpale, dwarslat, paalspringtoerusting, paal, springpaal, springstok, hekkie, stophorlosie* voor. Die leksikale items kan nie as sinonieme of naby-sinonieme van mekaar of van die sentrale sinoniem **atletiekbroekie** beskou word nie. Dit wil voorkom asof leksikale items lukraak bymekaar gebondel is in 'n artikel, met geen waarneembare semantiese betrekking tussen mekaar of ten opsigte van hul gebruik of betekenis nie.

In TA word die artikel aansienlik verbeter met die vervanging van **atletiekbroekie** met die leksikale item *atletiektoerusting*. Die leksikale item *atletiektoerusting* dien as hiponiem vir die subordinate *atletiekdrag, atletiekklere, atletiekbroekie, atletiekfrokkie, atletiekskoen, hardloopskoen, spykerskoen, wegspringblokke, skyf, werpskyf, diskus, gewig, spies, hamer, hoogspringpale, dwarslat, paalspringtoerusting, paal, springpaal, springstok, hekkie, stophorlosie* en die korrekte semantiese verhouding tussen leksikale items word verskaf.

Alhoewel leksikale items tot dieselfde semantiese veld behoort, word hulle bymekaar gebondel in groot en losse groepe, en geen leiding word aan die gebruiker verskaf oor hul presiese betekenis en toepassing nie. Daar word nie voldoende tussen die betekenis en die gebruik van die leksikale items onderskei nie. Die volle semantiese prentjie word dus nie aan die gebruiker verskaf nie. Met innovasie kan die saak reggestel word en 'n beter tesouruservaring kan aan die gebruiker gebied word. Daar kan op 'n deurlopende basis konteksleiding aan die gebruiker verskaf word, veral met die gebruik van etikette of ingevoegde tekste. Konteksleiding, waar 'n leksikale item gekwalifiseer word deur verdere semantiese onderskeidinge te verskaf, kan ook tussen hakies na die spesifieke leksikale item voorsien word.

Murphy (2016: 449) wys daarop dat die semantiese betrekkinge in tesourusartikels in *The New Collins Thesaurus* van so 'n aard is dat die sinoniempad tussen woorde en hul sinonieme in die meeste gevalle binne ses stappe na semantiese teenstelling lei. As voorbeeld word *authentic* genoem wat tot die teenstelling *unauthentic* lei:

*authentic-believable-probable-ostensible-pretended-spurious-unauthentic.*

Vergelyk die hantering van *vermoedelik* in die subparagraaf onder die artikel *glo* in TA, wat binne vier stappe tot die teenstellings *agterdogtig, wantrouig, onseker, twyfelagtig* lei. Die semantiese teenstelling volg na 'n kommapunt as struktuurmerker.

**vermoedelik**, hipoteties, konjekturaal, vooropgeset, stellig; agterdogtig, wantrouig, onseker, twyfelagtig

**Figuur 5:**  vermoedelik, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

Vervolgens word 'n verdere semantiese betrekking tussen leksikale items ondersoek, naamlik teenstelling.

### 3.2    Semantiese teenstelling

Binêre opposisie is een van die belangrikste eienskappe van taalstrukture en volgens Lyons (1977: 271, 272) kan die opvallendste manifestasie hiervan in antonimie gevind word. Lyons (1977: 271, 272) definieer 'n aantal terme, naamlik *kontras, teenstelling, antonimie, gradeerbare teenstelling* en *komplementêre teenstelling*. *Kontras* is die algemeenste term en dra geen implikasie nie. *Teenstelling* word beperk tot digotomie of binêre kontraste. *Antonimie* word beperk tot gradeerbare teenstelling, byvoorbeeld *groot-klein* of *hoog-laag*. *Gradeerbare teenstelling* manifesteer die eienskap van polariteit die beste van alle ander teenoorgesteldheid, byvoorbeeld *lank–kort* of *vinnig–stadig*. *Komplementêre teenoorgesteldheid* verwys na ongradeerbare opposisie, naamlik *manlik–vroulik*. Bosman en Pienaar (2014: 261) onderskei ook die voorkoms van *omkeerbare teenoorgesteldes*, waar leksikale items sintakties as teenoorgesteldes beskou word, byvoorbeeld *pasiënt–dokter*.

Die voorkoms van semantiese teenstelling in 'n tesourus kan gebruikers ondersteun om 'n leksikale item dadelik in 'n semantiese veld te plaas en sodoende kan betekenisbetrekkinge duideliker raak vir die tesourusgebruiker. Duideliker betekenisbetrekkinge kan 'n tesourus meer soortgelyk aan die gebruiker se mentale leksikon maak.

Betrekkinge van semantiese teenstelling word in die verskillende strukture van GT en TA aangetref, te wete die inhoudsopgawe, klassifikasieplan en

sentrale teks. In die inhoudsopgawe van beide tesourusse word byvoorbeeld na *skepping, lewe* en *dood, vreugde* en *droefheid, hoop* en *wanhoop, moed* en *vrees, gramskap* en *sagmoedigheid, liefde* en *haat* verwys. Die betrekkinge word nie pertinent aangedui of aan die gebruiker uitgewys met behulp van struktuurmerkers nie.

Die klassifikasieplan bestaan eweneens uit 'n digotomie van teenoorgesteldes wat vindbaar is aangesien hulle in numeriese volgorde is en langs mekaar in die klassifikasieplan voorkom. Vergelyk byvoorbeeld *bestaan* en *nie-bestaan, selfstandigheid* en *onselfstandigheid* onder die hofie *Bestaan* in die klassifikasieplan. Met die lees van die klassifikasieplan kan die oplettende tesourusgebruiker 'n netwerk van semantiese betrekkings ontdek. Dit word egter nie aan die gebruiker verduidelik nie. Die gebruik van struktuurmerkers kan hier van groot waarde wees, aangesien dit gebruik kan word om betrekkinge van teenoorgesteldheid vir die gebruiker aan te dui.

Indien daar 'n betrekking van teenoorgesteldheid tussen die leksikale items bestaan, volg hul op mekaar in die klassifikasieplan en in die sentrale teks (tesourusgedeelte). Die volgende lemmas volg op mekaar in GT en staan in 'n betrekking van teenoorgesteldheid ten opsigte van mekaar: (158) *opstyg* – (159) *neerdaal*, (161) *bedek* – (162) *ontbloot*, (211) *opgaan* – (212) *afgaan*, (249) *lewe* – (250) *dood*, (631) *nodig* – (632) *onnodig* en (633) *nuttig* – (634) *nutteloos*. In TA is die tersaaklike lemmas behou met een verandering, naamlik (158) word vervang met *na bo* en (159) met *na onder*. Alle leksikale items wat op mekaar volg, staan egter nie in 'n betrekking van teenoorgesteldheid ten opsigte van mekaar nie.

Vervolgens word 'n voorbeeld in die sentrale teks van GT bespreek. In die artikel 590 *Bestuur* kom die subartikel *regeer*, wat 'n subparagraaf vorm, voor. Die eerste sewe leksikale items wat as inskrywings verskaf word, kan as naby-sinonieme van mekaar beskou word, naamlik *gesag uitoefen, staatsgesag uitoefen, die land bestuur, die bewind voer, heers, heerskappy voer, bevel voer.* Dan volg 'n hele aantal leksikale items wat nie sinonieme van *regeer* is nie. Geen konteksleiding word verskaf nie en die gebruiker word in die duister gelaat oor hul semantiese betrekking met die lemma *regeer* en met mekaar. Die volgende leksikale items volg na die kommapunt (wat fyner betekenisonderskeidinge aandui): *politiek bedryf, verpolitiseer, institusionaliseer, demokratiseer, nasionaliseer, privatiseer, die gesag sentraliseer, desentraliseer, gesag afwentel, sit (die parlement), in sitting wees, debat voer (parlementêre - - ), 'n spreekbeurt neem, eerste lesing van 'n wetsontwerp hou, promulgeer ('n wet -), afkondig ('n wet -), interpelleer, 'n regering saamstel, 'n portefeulje hê, van portefeulje verander.*

Bevredigende oordrag van semantiese betrekkinge vind nie plaas nie en 'n voorstel in hierdie verband word gemaak: 'n Verdere onderskeid kan tussen die leksikale items getref word met behulp van 'n struktuurmerker. Die simbool ◉ kan byvoorbeeld gebruik word om leksikale items wat semanties nader aan mekaar verwant is, te groepeer. Konteksleiding kan tussen hakies verskaf word. Die res van die artikel kan soos volg daar uitsien:

politiek bedryf ◎ verpolitiseer ('n saak) ◎ institusionaliseer (instel), saamstel (regering), sentraliseer, desentraliseer, (gesag) afwentel ◎ demokratiseer, nasionaliseer, privatiseer ◎ sit (parlement), in sitting wees, (parlementêre) debat voer, spreekbeurt neem, interpelleer, eerste lesing van wetsontwerp hou ◎ (wet) afkondig of promulgeer ◎ 'n portefeulje hê, van portefeulje verander.

Van die leksikale items wat in die artikel saamgebondel is, kan as teenoorgesteldes van mekaar beskou word en nie as sinonieme of naby-sinonieme nie; vergelyk *nasionaliseer* en *privatiseer, die gesag sentraliseer* en *desentraliseer*. Die leksikograaf maak ongelukkig nie melding van hierdie betekenisbetrekkinge nie.

Met 'n vergelyking van die TA-inskrywing blyk dit dat die inhoud van die artikel onveranderd gebly het.

*regeer* 588, gesag uitoefen, staatsgesag uitoefen, die land bestuur, aan bewind wees, die bewind voer, heers, heerskappy voer, bevel voer; politiek bedryf, verpolitiseer, institusionaliseer, demokratiseer, nasionaliseer, privatiseer, die gesag sentraliseer, desentraliseer, gesag afwentel, sit (die parlement 590. Bestuur en regeer), in sitting wees, debat voer (parlementêre 590. Bestuur en regeer 590. Bestuur en regeer), 'n spreekbeurt neem, eerste lesing van 'n wetsontwerp hou, promulgeer ('n wet 590. Bestuur en regeer), afkondig ('n wet 590. Bestuur en regeer), interpelleer, 'n regering saamstel, 'n kabinet saamstel, 'n portefeulje hê, van portefeulje verander

**Figuur 6:**  regeer, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

Dieselfde aanbeveling kan dus ook by TA gemaak word. Kruisverwysings kan op 'n konsekwente wyse gebruik word om semantiese betrekkinge suksesvol aan te dui.

Die derde algemene semantiese betrekking wat deurgaans in tesourusse voorkom, naamlik hiponimie, word in die volgende afdeling bespreek.

### 3.3    Hiponimie

Hiponimie en meronimie word volgens Murphy (2016: 443) as simmetriese of hiërargiese betrekkings beskou, aangesien indien A 'n tipe (deel van) B is, B nie 'n tipe (deel van) A kan wees nie. Die gebrek aan simmetrie maak hierdie betrekkings belangrik in die definiëringsproses, aangesien 'n definiendum beskryf kan word sonder om sirkeldefiniëring te gebruik. Hiponimie en meronimie word direk in nielinguistiese betrekkinge tussen denotata reflekteer (Murphy 2016: 443). Hiponimie kan gedefinieer word as 'n semantiese betrekking van insluiting en Bosman en Pienaar (2014: 264) verskaf die volgende voorbeelde: *affodil–blom*; *hond–dier*. Die betrekking *insluiting* beteken dat indien iets 'n *affodil* is, dit ook

noodwendig 'n *blom* is en indien iets 'n *hond* is, dit ook noodwendig 'n *dier* is. Die woord wat 'n oorkoepelende klas of algemene term (byvoorbeeld *soogdier*) benoem, heet die superordinaat en die woorde wat die ingeslote klas (soos *hond*) benoem, heet hiponieme of subordinate. Die betekenis van die superordinaat is dus meer algemeen en die betekenis van die hiponiem(e) meer spesifiek, aldus Bosman en Pienaar (2014: 264). Murphy (2016: 443) wys egter op ons sinmaak van hiponimie as 'n paradigmatiese "soortvan"-verhouding, byvoorbeeld alle *professore* is *soogdiere*, maar die uiting "alle professore is soogdiere" kan as 'n vreemde uiting ervaar word.

Bosman en Pienaar (2014: 264) meld dat die voorbeelde wat gewoonlik gebruik word om die betrekking van hiponimie aan te dui, selfstandige naamwoorde is, maar in 'n WordNet-analise van werkwoorde is bevind dat veral die betrekking van insluiting en troponimie gebruik word. As voorbeeld word die werkwoorde *snork* en *slaap* genoem, waar daar 'n verhouding van insluiting voorkom: As jy snork, hou dit in dat jy slaap. *Neurie* is ook 'n voorbeeld van *sing* (Bosman en Pienaar 2014: 264).

Hiponimie is een van die leksikale betrekkinge wat in 'n tesourus aangedui word en in die voorwerk van GT en TA word dit pertinent genoem. Spesifieke hiponimiese betrekkinge word nie aangedui nie. In die tesourus word daar geen leiding aan die gebruiker verskaf ten opsigte van hiponimiese betrekkinge tussen leksikale items nie. Voorbeelde uit GT en TA word ter illustrasie bespreek.

In die geval van die artikel 363 *Waterdier* kan *amfibie* in die eerste plek as 'n sinoniem beskou word, maar dit kan ook as subordinaat van die superordinaat *waterdier* in GT dien. Die volgende subafdelings van die lemma *waterdier* kan almal as subordinate van die hiponiem *waterdier* beskou word: *manteldier, skulpdier, seekat, seeanemoon, koraal, seewurm, strandluis, vis, varswatervis, seevis* en *watersoogdiere*. Die subafdelings vertak verder in naby-sinonieme van die verskillende subordinate, byvoorbeeld *vis*, wat soos volg lui:

**vis,** seevis, pelagiese vis, varswatervis, beenvis, kraakbeenvis, roofvis, eetvis, hengelvis, lynvis, sportvis; vin, stekelvin, rugvin, eerste rugvin, dorsale vin, tweede rugvin, stert, stertvin, anaalvin, buikvin, borsvin, sylyn, kieu, kief, brangieë (*meervoud*), kieudeksel, kiefdeksel, kieuspleet, kieuboog, kieudraad, kieukammetjie, kieuhark, halfkieu, asemhalingsgat (walvis), spuitgat, skub, grom (ingewande), graat, visbeentjie

**Figuur 7:**   vis, uittreksel uit *Groot Tesourus van Afrikaans* (artikel verkort)

Die eerste 11 inskrywings kan as naby-sinonieme van die woord *vis* beskou word en die volgende inskrywings, geskei deur 'n kommapunt, eerder as dele van 'n vis se anatomie. Daar word nie duidelike kontekstuele leiding ten opsigte van die hiponimiese betrekking tussen die leksikale items in die artikel, maar

ook in ander artikels, byvoorbeeld *vissery* en *seekos*, aan die gebruiker verskaf nie. Daar is ook nie 'n verwysing na die uitdrukking *nóg vis, nóg vlees*, wat onder die artikel *bestaanswyse* voorkom, nie. Die volledige netwerk van semantiese betrekkinge tussen die leksikale items word dus nie vir die gebruiker duidelik nie.

Die artikel van 363 *Waterdier* in TA sien soos volg daaruit:



**363. Waterdier**
**s.nw.**
*waterdier*, amfibie
*vis*, seevis, soutwatervis, pelagiese vis, varswatervis, soetwatervis, beenvis, kraakbeenvis, roofvis, eetvis, hengelvis, lynvis, sportvis, vingerling; vin, stekelvin, rugvin, eerste rugvin, dorsale vin, tweede rugvin, stert, stertvin, anaalvin, buikvin, borsvin, sylyn, kieu, kief, brangieë (*meervoud*), kieudeksel, kiefdeksel, kieuspleet, kieuboog, kieudraad, kieukammetjie, kieuhark, halfkieu, asemhalingsgat (walvis), spuitgat (walvis), neusgat (walvis), skub, grom (ingewande), graat, visgraat, visbeentjie
*seevis*, blaasop, blaasoppie, baardman, belman, biskop, witbiskop, mosselkraker, sandbloue, beenbek, poenskop, dassie, kolstert, glasvis, heilbot, jakopewer, knorhaan, leervis, marlyn, bottervis, pampelmoes, pompelmoes, selakant, panga, dikbekkie, platvis, spiering, seeperdjie, springer, tienponder, stompneus, strepie, mooinooientjie, skol, varkbek, varkie, wildeperd, bontrok, loodsvis, seepaling, slangaal; eetvis, ansjovis, moggel, baars, briekwabaars, aartappelbaars, elf, galjoen, bandgaljoen, bastergaljoen, kraaibek, knorder, gespikkelde knorder, spiesknorder, grys knorder, geelbek, Kaapse salm, harder, suidelike harder, haring, haringvis, hottentotsvis, jandorie, geelstert, albakoor, halfkoord, tuna, geelvintuna, langvintuna, streeppenstuna, kabeljou, stompneuskabeljou, snapperkabeljou, kob, daga, kleinkabeljou, nondi, klipkabeljou, katonkel, knorhaan, koningvis, koningklipvis, koningklip, dorado, makriel, marsbanker, masbanker, pelser, sardyn, sardien, sardientjie, roman, rooi roman, rooibaardman, rooiharder, salm, baber, seebaber, wit seebaber, silwervis, doppie, skelvis, skotsman, slimjannie, snoek, steenbras, witsteenbras, rooisteenbras, tjortjor, varkie, janbruin, blouoog, tambryn, biskop, daeraad, steentjie, stokvis, stompneus, witstompneus, rooistompneus, tarbot, tong, tongvis; roofvis, kabeljou, marlyn, pylstert, pylstertvis, elektriese pylstert, rog, manta, vleet, vioolhaai, vioolvis, seeduiwel, saagvis, swaardvis, seilvis, monnikvis, barrakuda, kapteinvis, snoek, haai, withaai, grootwithaai, witdoodshaai, walvishaai, zambezihaai, duisterhaai, kortvin-mako, tierhaai, skeurtandhaai, blouvinhaai, vaalhaai, spierhaai, koeihaai, seskiefhaai, sebrahaai, hamerhaai, hamerkophaai, koperhaai, makohaai, tonynhaai, sandhaai, gevlekte sandhaai, sandkruiper, aashaai, seehond, hondshaai, see-engel, engelvis, seepaling, snoekpaling, heuningkoek-bontpaling, sidderaal, lamprei, prikvis; getypoelvis, klipvis, dikkop, band-dikkop, blennie, kamkopblennie, suigvis, klipsuier; jellievis, kwal, drilvis, ribkwal, seekwal, medusa, bloublasie
*varswatervis*, baars, stekelbaars, baber, barber, brasem, forel, salmforel, geelvis, karper, ghieliemientjie, slymvis, paling, rivierpaling, goudvis, goudvissie, pirana
*manteldier*, ongewerwelde seedier, rooi-aas, palprooi-aas
*amfibie*, amfibiese dier, padda, kikvors, paddavis, skurwepadda, brulpadda, reënpadda, groot reënpadda, janblom, donderpadda, blaasop, salamander, salmander, platanna
*skulpdier*, mollusk, weekdier, ongewerwelde, ongewerwelde dier, slak 361, waterslak, seeslak, wulk, porseleinslak; naakslak; kauri, kouri, kourie, kinkhoring, nautilus, papiernautilusskulp, keëlskulp, marginaskulp, ploegskulp, ramshoringskulp, borrelvlotskulp, gesiggieskulp, kiton; alikreukel, tolskulp, tulbandskulp; oester, oesterskulp, katoog, mossel, hardedopmossel, sagtedopmossel, klipmossel, valsklipmossel, witmossel, swartmossel, kammossel, venusmossel, gapermossel, pypmossel, sleutelgatmossel, pantoffelmossel; perlemoen, perlemoer, klipkous, siffie, neriet
*seekat*, agarm, oktopus, inkvis, tjokka
*seeanemoon*, aktinie, seeroos, seekastaiing, seekomkommer, seester, slangster, veerster

**Figuur 8:**  waterdier, uittreksel uit *Tesourus van Afrikaans* (aanlyn weergawe) (artikel verkort)

By nadere ondersoek blyk dit dat die artikel hersien is, met 12 subordinate wat gelys word. Die sinoniem *amfibie* is bygevoeg en die volgorde het verander na *vis, seevis, varswatervis, manteldier, amfibie, skulpdier, seekat, seeanemoon, koraal, seewurm, strandluis* en *watersoogdiere.* Die nuwe volgorde van verskillende kategorieë vissoorte kan as 'n verbetering beskou word deurdat dit semantiese konsepte wat nader aan mekaar is en meer natuurlik op mekaar volg na aan mekaar geplaas is. Ter illustrasie van die verbetering kan *vis*, wat as hiponiem vir die volgende optree, verder ontleed word: *seevis, soutwatervis, pelagiese vis, varswatervis, soetwatervis, beenvis, kraakbeenvis, roofvis, eetvis, hengelvis, lynvis, sportvis, vingerling.* 'n Kommapunt skei die volgende inskrywing wat eerder 'n deel-van-verhouding uitbeeld, naamlik *vin, stekelvin, rugvin, eerste rugvin, dorsale vin, tweede rugvin, stert, stertvin, anaalvin, buikvin, borsvin, sylyn, kieu, kief, brangieë (*meervoud*), kieudeksel, kiefdeksel, kieuspleet, kieuboog, kieudraad, kieukammetjie, kieuhark, halfkieu, asemhalingsgat (walvis), spuitgat (walvis), neusgat (walvis), skub, grom (ingewande), graat, visgraat, visbeentjie.* Hierdie artikel lys dus eerder hiponimiese betrekkinge tussen woorde as sinonimiese betrekkinge. Konteksleiding ontbreek egter om 'n volledige oordrag van semantiese betrekkinge tussen woorde te illustreer.

Verdere voorbeelde van hiponimie-artikels is *soogdier, reptiel, voël, oerdier, insek, groenteverbouing, vrugteverbouing* en hul subordinate. Die leksikograaf kan beter leiding verskaf deur van 'n tipografiese konvensie, wat in die toeligtende aantekeninge aan die gebruiker verduidelik word, gebruik te maak om 'n hiponimiese betrekking aan te dui.

Tot dusver was die onderwerp van bespreking die bestaan van semantiese betrekkings tussen leksikale items, met verwysing na woordeboekstrukture soos die sentrale teks, klassifikasieplan en indeks. Daar is deurgaans na die elektroniese weergawe van TA verwys sonder verdere bespreking daarvan, en in die volgende afdeling is die fokus op semantiese betrekkinge en woordeboekstrukture in die elektroniese weergawe.

## 4.     Woordeboekstrukture en semantiese betrekkinge in die elektroniese tesourus

Leksikografie bly geensins onaangeraak deur verskeie uitvindings en ontwikkeling op tegnologiese gebied nie, soos tereg deur verskillende navorsers opgemerk. Fuertes-Olivera en Tarp (2020: 258) wys onder andere op die invloed van nuwe tegnologie op alle aspekte van die leksikografie, naamlik vanaf empiriese databronne vir die insameling, verwerking en stoor van data tot die aanbieding daarvan vir eindgebruikers. Klosa en Gouws (2015: 143) bespreek 'n teoretiese basis vir die aanbieding van data aan eindgebruikers met verwysing na woordeboekstrukture, spesifiek buitetekste, in die elektroniese woordeboeke. Volgens Gouws (2014; 2014a) is dit noodsaaklik vir die suksesvolle aanpassing van 'n algemene teorie vir leksikografie om medium-onspesifiek te word en aan die behoeftes van die praktyk van aanlyn woordeboeke te voldoen sodat vol-

doende aandag geskenk word aan verskillende woordeboekstrukture relevant vir aanlyn woordeboeke. Klosa en Gouws (2015: 143) ondersoek buitetekste in elektroniese woordeboeke deur dit met buitetekste in gedrukte woordeboeke te vergelyk en die dataverspreidingstruktuur in gedrukte en elektroniese woorde- boeke te vergelyk, en tref 'n onderskeid tussen woordeboek-interne en woorde- boek-eksterne buitetekste. In die navorsing word daar na verskillende buitetekste in elektroniese woordeboeke (gebruikersgidse, aanwysers na woordeboekinhoud en taalspeletjies) verwys (Klosa en Gouws 2015: 143). Teoretiese insigte uit laas- genoemde artikel word vervolgens op strukture in die elektroniese tesourus toegepas.

Aangesien woordeboeke as houers van inligting beskou word en uit verskil- lende tekstipes bestaan, word nie net die sentrale woordelys in die aanbied van leksikografiese data in gedrukte woordeboeke benut nie (Klosa en Gouws 2015: 146). Voorwerk- en agterwerktekste word ook gebruik om die ware doel van 'n woordeboek te verwesenlik en die artikelstruktuur van 'n gedrukte woordeboek bestaan uit verskillende komponente, naamlik tekssegmente, kommentaar en soeksones (Klosa en Gouws 2015: 146).

Die gedrukte TA kan as 'n politoeganklike woordeboek beskou word, aan- gesien dit oor meer as een toegangstruktuur en soekroete beskik. Dit beskik oor drie tematies geordende toegangstrukture, te wete die inhoudsopgawe, die klassifikasieplan en die sentrale teks, asook 'n inisieel-alfabetiese toegangstruk- tuur.

Die inhoudsopgawe lys die verskillende konsepkategorieë waarin die tesourus verdeel is, asook die 10 hoofkategorieë. Naas die hoofkategorieë word die konsepnommers verskaf en by die konsepkategorieë word die spesifieke konsepnommer asook die bladsynommer verskaf. Alhoewel die inhoudsop- gawe as deel van die toegangstruktuur beskou moet word, gaan die gebruiker dit nie as 'n soekroete vir die inwinning van kitsinligting gebruik nie, tensy die gebruiker al die lemmas onder 'n spesifieke konsepkategorie wil naslaan, aan- gesien die verskillende lemmas wat in die makrostruktuur voorkom nie gelys word nie. Die inhoudsopgawe vorm nie deel van die elektroniese weergawe van TA nie en dit sal nie funksioneel wees om dit deel te maak daarvan nie, aangesien die gebruiker leksikale items deur middel van 'n soekblok naslaan.

Die tweede tematies geordende toegangstruktuur van TA is 'n klassifika- sieplan. Daarin word die verskillende hoofkategorieë genommer, die konsep- kategorieë met hoofletters genommer en die verskillende lemmas genommer. Verdere tematiese indelings onder die konsepkategorieë word gemaak en met kleinletters genommer. Die klassifikasieplan dien as 'n uiteensetting van die konseptuele raamwerk van die woordeboek en die gebruiker kry die verskil- lende konsepkategorieë en onderafdelings wat in die woordeboek voorkom. Dit vorm tans nie deel van die elektroniese weergawe van TA nie, maar die skrywer beskou dit wel as 'n funksionele buiteteks. Dit kan as funksioneel beskou word vir die eiesoortige aard van die tesourus, aangesien ondersteu- ning vir die identifisering van 'n netwerk van semantiese betrekkings tussen

leksikale items en konsepte aan die gebruiker verleen kan word. Hierdie toe-gangstruktuur kan as 'n soekroete deur die gebruiker gebruik word om 'n arti-kel na te slaan, aangesien die nommer van die artikel verskaf word en 'n elek-troniese soektog so moontlik kan wees.

Die tweede tipe toegangstruktuur van TA is 'n inisieel-alfabetiese een. Aangesien tesourusse nie 'n tradisionele alfabetiese toegangstruktuur het wat as 'n maklike soekroete vir inligting dien nie, het TA 'n gebruikersgerigte hulp-middel, naamlik 'n alfabeties geordende indeks. Die indeks vorm nie deel van die elektroniese weergawe van TA nie en dit sal nie funksioneel wees om dit deel daarvan te maak nie, aangesien die gebruiker leksikale items deur middel van 'n soekblok naslaan. Met behulp van 'n soekveld bereik die gebruiker 'n artikel in die sentrale teks van die elektroniese weergawe. Vergelyk figuur 9 hier onder.
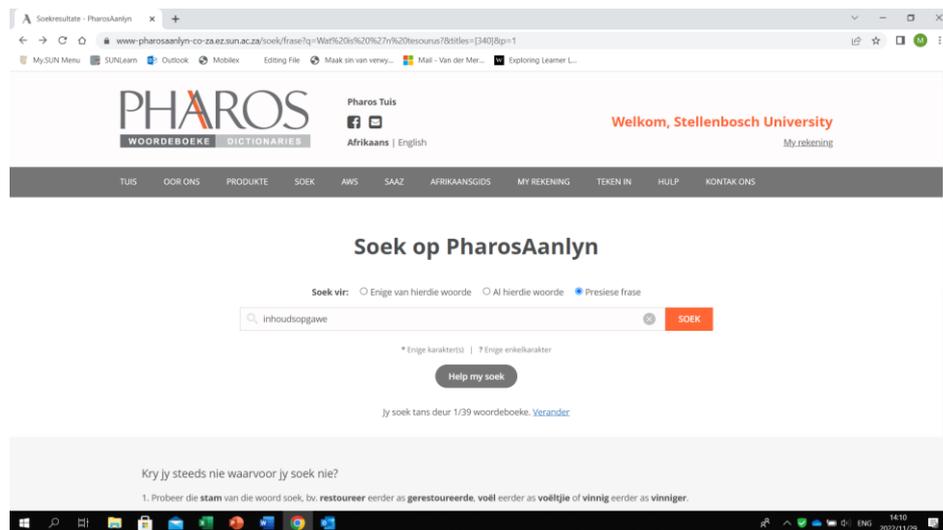


**Figuur 9:**  Soekblok van PharosAanlyn wat toegang tot *Tesourus van Afrikaans* bied

Volgens die soekblok in die elektroniese weergawe is daar drie opsies waarmee jou soektog aangepas kan word, naamlik *Soek vir: Enige van hierdie woorde, Al hierdie woorde* of *Presiese frase.* Die soekblok beklee 'n prominente posisie in die middel van die bladsy en die gebruiker weet presies waar om 'n leksikale item na te slaan. Volgens Klosa en Gouws (2015: 163) blyk inleidende aantekeninge oor die gebruik van 'n woordeboek onnodig te wees, maar 'n prominente soek-veld is belangrik. Dieselfde geld navigasie binne die sentrale teks, waar gebrui-kers in die meeste gevalle weet hoe om deur lang artikels te navigeer en hiper-skakels te gebruik. Die voorwoord en inleidende artikel oor die aard en gebruik

van die tesourus is aanwesig in die elektroniese weergawe, maar slegs indien daar pertinent 'n soektog daarna uitgevoer word. Dit is dus een van die buitetekste wat as oortollig beskou kan word in 'n elektroniese weergawe.

Met ondersoek blyk dit dat die sentrale teks van TA identies is aan die elektroniese teks. 'n Sentrale teks in elektroniese formaat bied baie leksikografiese moontlikhede wat tot voordeel van die gebruiker aangepas kan word. Met die fokus op meer kontekstuele leiding rakende die verskaffing van 'n semantiese netwerk van konsepte en betekenisbetrekkinge tussen leksikale items, kan leksikografiese vernuwing op twee wyses plaasvind. Eerstens kan onderskeid tussen verskillende semantiese betrekkinge (sinonimie, hiponimie) getref word deur hiperskakels na ander verwante artikels in te voeg. Aangesien die tesourus deel vorm van die aanlynwoordeboekportefeulje van Pharos (en dus ook 'n verklarende woordeboek), kan hiperskakels na betekenisverklarings ingevoeg word. Sodoende kan kontekstuele leiding verskaf word vir die tans slegs gelyste items in 'n tesourus. Die moontlikhede met hiperskakels is eindeloos en die uitgewer kan skakels na die Afrikaanse WordNet of ander nuttige webwerwe oorweeg.

Tweedens word 'n soektog wat meer gerig is op die semantiese betrekkinge tussen leksikale items in die elektroniese tesourus voorgestel. Soektogte om die samehang in die leksikon te kontekstualiseer, kan uitgebrei word. Die skrywer stel dus vrae soos volg in die soekveld voor:

*Wat is 'n sinoniem van X?*
*Wat is 'n hiponiem van X?*
*Wat is subordinate van X?*
*Wat is die teenoorgestelde van X?*

Met die gebruik van bogenoemde vrae kan die sentrale teks se data gebruik word vir spesifieke soektogte wat semantiese betrekkings insluit en meer eenduidige antwoorde op 'n gebruiker se moontlike vrae verskaf. Die manier waarop leksikale items tans in die sentrale teks gelys word, maak dit nie vir die gebruiker moontlik om te onderskei watter leksikale items nie onderling vervangbaar is nie, en sodoende kan die samehang in die leksikon vir die gebruiker gekontekstualiseer word.

## 5.    Gevolgtrekking

Kay en Alexander (2016: 380) skets 'n rooskleurige prentjie van die toekomstige waarde van 'n elektroniese tesourus teen die agtergrond van die tegnologiese integrasie van 'n databasis. Tegnologiese ontwikkeling op leksikografiese gebied sluit onder andere die gebruik van 'n tesourus as databasis in, waar 'n gebruiker nie meer van 'n alfabetiese toegangstruktuur gebruik maak om inligting te bekom nie. Die voordele van 'n tesourus, waar leksikale items volgens semantiese betrekkinge gerangskik word, is aansienlik meer aantreklik as alfabetiese

ordening in 'n tradisionele gedrukte woordeboek (Kay en Alexander 2016: 380). Kay en Alexander (2016: 380) beskou 'n tesourusformaat as 'n voorbeeld van naatlose integrasie in die digitale era vir gebruikersgerigtheid en voorspel dat die oudste vorm van leksikografie en die dominate vorm voor die koms van alfabetisering weereens dominant in die postdruk- digitale toekoms sal wees.

Fuertes-Olivera en Tarp (2020: 257) se waarneming is dat leksikografiese produkte nie meer woordeboeke par excellence is nie, maar wel leksikografiese data wat óf in woordeboekformaat óf deur 'n verskeidenheid van hulpmiddels, platforms en dienste aan 'n gebruiker verskaf word. Hulle beskryf die funksionaliteite van 'n digitale skryfassistent as voorbeeld van so 'n hulpmiddel en hoe 'n databasis daarvoor saamgestel kan word (Fuertes-Olivera en Tarp 2020: 257). TA beskik reeds oor 'n elektroniese databasis en kan moontlik vir 'n digitale skryfassistent in Afrikaans aangepas word. TA vorm reeds deel van die aanlyn platform vir Pharos se leksikografiese produkte en is dus goed geposisioneer om 'n sekondêre produk van die tesourus te skep, wat aansluit by die idee van volhoubare leksikografie (Colman 2016: 141).

Aansluitend by die verskaffing van 'n diens aan gebruikers in die digitale omgewing, is die voorkoms van leksikovermaak as buiteteks, soos deur Klosa en Gouws (2015: 164) beskryf. Leksikovermaak verwys onder andere na 'n verskeidenheid taalspeletjies en onderrigmateriaal wat in gedrukte formaat of op 'n interaktiewe wyse gebruik kan word (Klosa en Gouws 2015: 164). Tans is 'n "Woord van die week", 'n "Spreekwoord" en "Het jy geweet?" as leksikovermaak op die Pharos-platform beskikbaar. Daar bestaan egter 'n leemte in die leksikovermaak rakende tesourusgebruik en die skrywer stel 'n fokus op inligting oor semantiese betrekkings tussen leksikale items vir leksikovermaak voor. Inligting rakende betrekkinge van sinonimie, hiponimie en teenoorgesteldheid kan op 'n opvoedkundige, vermaaklike en interaktiewe wyse aan die gebruiker gebied word ten einde die gebruik van 'n tesourus op die voorgrond te plaas.

## Eindnote

1. Doelbewuste verwysing, op grond van die hegte (ook akademiese) vriendskap met die persoon aan wie die bundel opgedra word.
2. Doelbewuste verwysing, op grond van die persoonlike en akademiese vriendskap met die persoon aan wie die bundel opgedra word.

## Bibliografie

### Woordeboeke

**De Stadler, L.G. en A. de Stadler.** 1994. *Groot Tesourus van Afrikaans*. Halfweghuis: Southern Boekuitgewers.

**De Stadler, L.G. en A. de Stadler.** 2021. *Tesourus van Afrikaans*. Kaapstad: Pharos (aanlyn).

**Luther, J., F. Pheiffer en R.H. Gouws.** 2015. *Handwoordeboek van die Afrikaanse Taal* (HAT). Sesde uitgawe. Kaapstad: Pearson Education.


### Ander bronne

**Bosman, N. en M. Pienaar.** 2014. Afrikaanse semantiek. Carstens, W.A.M. en N. Bosman (Reds.). 2014: 245-273.

**Carstens, W.A.M. en N. Bosman (Reds.).** 2014. *Kontemporêre Afrikaanse taalkunde.* Pretoria: Van Schaik.

**Colman, L.** 2016. Sustainable Lexicography: Where to Go from Here With the ANW (*Algemeen Nederlands Woordenboek*, an Online General Language Dictionary of Contemporary Dutch)? *International Journal of Lexicography* 29(2): 139-155.

**De Stadler, L.G.** 2022. Navraag oor tesourus. E-poskommunikasie, 10 November.

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

**Fuertes-Olivera, P.A. en S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36(2020): 257-286.
    Beskikbaar: https://doi.org/10.1515/lex-2020-0014 [2 Desember 2022].

**Gao, J. en H. Liu.** 2019. Valency and English Learners' Thesauri. *International Journal of Lexicography* 32(3): 326-361.
    Beskikbaar: https://doi: 10.1093/ijl/ecy025/ [2 Desember 2022].

**Gouws, R.H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.

**Gouws, R.H.** 2014a. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3). 481-504.

**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.
    Beskikbaar: https://doi:10.5842/53-0-737 [2 Desember 2022].

**Hartmann, R.R.K.** 2005. *Teaching and Researching Lexicography*. Beijing: Foreign Language Teaching and Research Press.

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

**Hausmann, F.J. et al. (Reds.).** 1989–1991. *Worterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography.* Berlyn: Walter de Gruyter.

**Hüllen, W.** 2006. Roget's Thesaurus, 1852 to 2002. *Lexicographica* 21(2005, 2006): 71-82.
    Beskikbaar: https://doi.org/10.1515/9783484604742.71 [2 Desember 2022].

**Kay, C. en M. Alexander.** 2016. Diachronic and Synchronic Thesauruses. Durkin, P. (Red.). 2016: 367-392.

**Kilgarriff, A. en C. Yallop.** 2000. What's in a Thesaurus? *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00), May 2000, Athens, Greece.* European Language Resources Association (ELRA): 1371-1379.
    Beskikbaar: https://aclanthology.org/L00-1134/ [26 April 2022].

**Klosa, A. en R.H. Gouws.** 2015. Outer Features in e-Dictionaries / Außentexte in Online-Wörterbüchern / Caractéristiques extérieures dans les dictionnaires en ligne. *Lexicographica* 31(1): 142-172.
    Beskikbaar: https://doi.org/10.1515/lexi-2015-0008 [2 Desember 2022].

**Knappe, G.** 2013. Phrasal and Phraseological Synonymy in the *Historical Thesaurus of the Oxford English Dictionary. International Journal of Lexicography* 26(3): 305-328.

**Lyons**, **J.** 1977. *Semantics: Volume 2*. Cambridge: Cambridge University Press.

**Marello, C.** 1989. The Thesaurus. Hausmann, F.J. et al. (Reds.). 1989**–**1991: 1083-1094.

**Murphy, M.L.** 2013. What We Talk about When We Talk about Synonyms. *International Journal of Lexicography* 26(3): 279-304.

**Murphy, M.L.** 2016. Meaning Relations in Dictionaries: Hyponymy, Meronymy, Synonymy, Antonymy, and Contrast. Durkin, P. (Red.). 2016: 439-456.

**Peters, W. en A. Kilgarriff.** 2000. Discovering Semantic Regularity in Lexical Resources. *International Journal of Lexicography* 13(4): 287-312.

**Van Schalkwyk, C.** 1996. Resensie: De Stadler, L.G. en Amanda de Stadler. *Groot Tesourus van Afrikaans. Lexikos* 6: 283-293.

**Vos, I.L.** 1994. Resensie: Harteveld, P. (Red.), L.G. de Stadler en D.C. Hauptfleisch. *Woordkeusegids: 'n Kerntesourus van Afrikaans. Lexikos* 4: 305-307.

**Wiegand, H.E.** 2002. Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Zugleich ein Betrag zur Weiterentwicklung einer Theorie der Wörterbuchform. Wiegand, H.E. (Red.). 2002a: 497-595.

**Wiegand, H.E. (Red.).** 2002a. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des* de Gruyter Wörterbuchs Deutsch als Fremdsprache. Lexicographica. Series Maior 110. Tübingen: Max Niemeyer.

**Wiegand, H.E.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Betrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.

**Wiegand, H.E.** 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Betrag zur Theorie der Wörterbuchform. *Lexicographica* 21: 202-379.

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

# VOORSKRIFTE AAN SKRYWERS

(*Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe
van hierdie instruksies of besoek ons webblad:* http://lexikos.journals.ac.za/)

## A. REDAKSIONELE BELEID

### 1. Aard en inhoud van artikels
Artikels kan handel oor die suiwer leksikografie of oor impli-kasies wat aanverwante terreine, bv. linguistiek, algemene taalwetenskap, terminologie, rekenaarwetenskap en bestuurs-kunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navor-sing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evalu-erend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde weten-skaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onder-werp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepu-bliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betref-fende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid arti-kels, aankondigings en nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksi-ko-opname, Leksikobibliografie, Leksikonuus, Lexikofokus, Leksiko-eerbewys, Leksikohuldeblyk, Verslae van konferen-sies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van aka-demiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

### 2. Wetenskaplike standaard en keuringsprosedure
*Lexikos* is deur die Departement van Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. in-komstegenererende navorsingstydskrif goedgekeur. Dit ver-skyn ook op die *Institute of Science Index* (ISI).

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; pro-bleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

Manuskripte word vir publikasie oorweeg met dien ver-stande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

### 3. Taal van bydraes
Afrikaans, Duits, Engels, Frans of Nederlands.

### 4. Kopiereg
Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Direksie van die Woordeboek van die Afri-kaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

### 5. Oorspronklikheid
Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oor-spronklikheid en feitelike inhoud van hulle publikasies. Indien van toepassing, moet besonderhede van die oor-sprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

### 6. Gratis oordrukke en eksemplare
*Lexikos* is sedert volume 28 slegs elektronies beskikbaar op http://lexikos.journals.ac.za. Geen oordrukke of eksemplare is dus beskikbaar nie.

### 7. Uitnodiging en redaksionele adres
Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en verkieslik in elektroniese formaat aan die volgende adres te stuur: lexikos@sun.ac.za, of Die Redakteur: LEXIKOS, Buro van die WAT, Pos-bus 245, 7599 STELLENBOSCH, Republiek van Suid-Afrika.

## B. VOORBEREIDING VAN MANUSKRIP
Die manuskrip van artikels moet aan die volgende redaksio-nele vereistes voldoen:

### 1. Lengte en formaat van artikels
Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat ver-soenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie **8 000 woorde** oorskry nie.

Elke artikel moet voorsien wees van 'n **opsomming** van ongeveer 200 woorde en ongeveer 10 **sleutelwoorde** in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutel-woorde **in Engels**. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigeen van die aan-geduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n **vertaling van die oor-spronklike titel** bevat.

### 2. Grafika
Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aange-dui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

### 3. Bibliografiese gegewens en verwysings binne die teks
Kyk na onlangse nommers van *Lexikos* vir meer inligting. Buiten in spesiale gevalle moet verwysings na *Lexikos*-artikels tot twee of drie per artikel beperk word. Uitsonde-rings moet met die redakteur van *Lexikos* uitgeklaar word. Dít word gedoen om die status van *Lexikos* in verskeie internasionale indekse te behou.

### 4. Aantekeninge/voetnote/eindnote
Aantekeninge moet deurlopend in die vorm van boskrifte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

# INSTRUCTIONS TO AUTHORS

*(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za)*
*or refer to our website:* http://lexikos.journals.ac.za/)

## A. EDITORIAL POLICY

### 1. Type and content of articles

Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.

(4) **Projects:** Discussions of lexicographical projects.

(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.

(6) **Lexicovaria:** Any of a large variety of articles containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.

(7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

### 2. Academic standard and evaluation procedure

The Department of Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index* (ISI).

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

### 3. Language of contributions

Afrikaans, Dutch, English, French or German.

### 4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free, however, to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

### 5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

### 6. Free offprints and copies

*Lexikos* is only available electronically on http://lexikos. journals.ac.za from volume 28 onward. No offprints or copies are available.

### 7. Invitation and editorial address

All interested authors are invited to submit contributions, preferably in electronic format, for publication in *Lexikos* to: lexikos@sun.ac.za, or

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH, Republic of South Africa

## B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

### 1. Format and length of articles

Manuscript should preferably be submitted in electronic format by email or on a disk, in software compatible with MS Word. The typeface used should preferably be 10-point Palatino or Times Roman. Contributions should not exceed **8 000 words**.

Each article must be accompanied by **abstracts** of approximately 200 words and approximately 10 **keywords** in the language in which it is written, as well as **in English**. English articles of South African origin should carry an abstract and keywords in Afrikaans, whilst English articles of foreign origin should carry a second abstract and keywords in any of the other languages mentioned. In cases where this is not done, the editors will provide an Afrikaans version. Ensure that the abstract in the second language also contains a **translation of the original title**.

### 2. Graphics

Figures such as tables, graphs, diagrams and illustrations should be in an appropriate size to be well accommodated within the page size of *Lexikos,* namely 18 cm high by 12 cm wide. The locations of figures within the text must be clearly indicated. If orthographic marks or graphics used in the text prove problematic, a printout of the manuscript or an email in .pdf format may be requested.

### 3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details. Self-references to *Lexikos* should be limited to two or three per article, except in exceptional circumstances. Exceptions should be cleared with the editor of *Lexikos.* This is done to preserve the status of *Lexikos* in various international indices.

### 4. Notes/footnotes/endnotes

Notes must be numbered consecutively by superscript numbers and grouped together at the end of the manuscript under the heading **Endnotes**.